MOLECULES, MORPHOLOGY AND MONOPHYLY: RESOLVING PLEURONECTIFORM PHYLOGENY AND INVESTIGATING WHY IT HAS BEEN SO DIFFICULT TO DO

A Dissertation
submitted to the Faculty of
The Richard Gilder Graduate School
at the
American Museum of Natural History
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy

By

Dawn M. Roje, M.S.

Richard Gilder Graduate School
at the
American Museum of Natural History
New York, NY 10024
23 September 2014

MOLECULES, MORPHOLOGY AND MONOPHYLY: RESOLVING PLEURONECTIFORM PHYLOGENY AND INVESTIGATING WHY IT HAS BEEN SO DIFFICULT TO DO

Dawn M. Roje, M.S.

Chair: John S. Sparks, Ph.D.

ABSTRACT

Although adult flatfishes (order Pleuronectiformes) start out in life as bilaterally symmetrical larvae, they undergo a remarkable metamorphosis, where one eye of the symmetrical larva migrates to the opposite side of the cranium, resulting in highly asymmetrical juvenile and adult forms. Because all flatfishes exhibit this bizarre morphology and variation, both the degree of asymmetry and handedness (direction of eye migration) exists within the order, this group provides multiple tests of hypotheses regarding the evolution of bilateral asymmetry and underlying mechanisms. Unfortunately, undertaking such studies has been elusive because of three major issues confounding pleuronectiform phylogenetics: 1) relationships of the major groups within the order remain mostly unresolved, 2) the sister group of flatfishes is unknown, and 3) monophyly of the assemblage is weakly supported.

To resolve these issues in pleuronectiform phylogenetics, my dissertation research has focused on: 1) evaluating the effects non-neutral markers on phylogeny estimation, principally rhodopsin1 (*rho*), 2) rigorously testing both flatfish monophyly and sister-group hypotheses and 3) resolving relationships within the order, re-examining characters of adult morphology and comparing them to often overlooked larval characters in light of new phylogenetic hypotheses.

In the first study, Chapter II, I use previously published sequence data from 78

acanthomorph (including flatfishes) taxa for *rho*, *rnf213*, *irbp* and *mll*, perform tests for neutrality, and compare neutral versus non-neutral markers for congruence using tree distance metrics and topology testing. I find that while the signal provided by *rho* may be discordant with the others, neutrality alone does not predict congruence and therefore should not be used as a justification to omit data.

In the second study, Chapter III, I optimize new molecular markers and sequence them along with *rho* and *rnf213* for 58 flatfishes and 90 putative outgroups to test monophyly, intraordinal relationships and sister-group hypotheses. Those sequences along with data from a previous study are analyzed to determine possible causes for gene tree incongruence or phylogenetic error. I discover that the new markers are variable, providing large amounts of data, while being conserved so that alignment is unambiguous. When combined with the others and analyzed simultaneously, these markers provide overwhelming support for a monophyletic Pleuronectiformes. Additionally, I demonstrate that abundant missing data is likely the cause of low resolution, validate the importance of investigating substitution saturation as a cause of error and discuss asymmetrical taxonomic distribution as a cause of low resolution at the base of Carangimorpha.

Finally, in the third study (Chapter IV) I infer an ultrametric tree, recode a previously published matrix of characters of adult morphology, combine those with new larval characters and test whether life history is correlated with phylogenic signal. Further, I investigate the accuracy of ML ancestral character state estimation (ACE) to determine if these morphological characters provide additional support for hypotheses of relationships among major pleuronectiform groups. My results suggest that larval

characters should not be treated as a source of independent data, but do provide resolution and additional support for novel relationships within Pleuronectiformes, although they may be in violation of the condition of low rates on ML ACE. Lastly, I show that because larval characters are mostly pleisiomorphic for the order, and that larval morphology is similar to that of putative sister groups, these characters are a potential source of evidence needed to resolve the placement of this lineage within Acanthomorpha.

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor and dissertation committee chair John Sparks for his unwavering support and valuable counsel. He also deserves a huge "thank you" for inviting me to join him on the Explore21 Expedition to the Solomon Islands in 2013. John is the best lead scientist I have ever had the pleasure to go to sea with and his work ethic and ability to push both scientists and crew while still managing to have fun is seriously impressive and something that I aspire to. I am also sincerely grateful to John for sharing his enormous talent as an editor with me. I am certain that it is because of his guidance that I was successful in getting proposals funded and papers published during my four years at the Richard Gilder Graduate School. For that I am eternally grateful.

For sharing her extensive knowledge and experience with me I am forever thankful to Melanie Stiassny, my co-advisor and dissertation committee member. The advice Melanie provided as I attempted to make sense of confusing flatfish anatomy was desperately needed and is still deeply appreciated. I would also like to acknowledge the critical role Melanie has played in my growth as a scientist and woman. The graceful and effective way she is able to navigate through an endlessly frustrating patriarchy is something that I admire in her. It is women like Melanie that have made this challenging endeavor in academia possible for a woman like me and for that she deserves all the thanks I can give her.

I also thank Rob DeSalle, the final member of my dissertation committee, for sharing his always-impressive knowledge and understanding of all things molecular and theoretical with me. His ability to distill complex concepts into digestible and

meaningful ideas gave me the courage I needed to open the “black box” of phylogenetic analysis. Rob challenged me while still allowing me to build my confidence as a researcher and for that I will always be thankful.

Similarly, I am extremely grateful to Ward Wheeler, whose *Systematics* course while being demanding was also illuminating and gratifying. Simply, Ward provided me with the skills I needed to execute this dissertation and for that he deserves my sincere gratitude. I also acknowledge Mark Sidall, Susan Perkins and Eunsoo Kim. Although you were not official committee members, you taught me so much and you too deserve all the gratitude I can offer.

To my fellow graduate students John Denton, Jairo Arroyave, Zach Baldwin, Phil Barden, Ansel Payne, Pedro Peloso, Sebastian Kvist, Shaena Montanari, Ed Stanley, and so many others: Thank you for all of your help and support. I am especially appreciative of the emotional support and understanding that only you could provide.

I am deeply grateful to the RGGS administration. Particularly John Flynn, Anna Manuel, Maria Rios and Taylor Johnson. I depended on you and you never let me down. Thanks for always answering emails, keeping your door open and for providing the guidance necessary to fulfill the requirements of this degree. I would also like to acknowledge the best lab manager I have ever worked with, Mohammed Faiz. He too always dropped what he was doing to help me and am greatful for hisgereosity.

Finally, I would like to acknowledge Christine Thacker, Curator of Ichthyology at the Natural History Museum of Los Angeles County. Although she has never been an official advisor of mine she has had the greatest impact on the way I approach research and navigate through the challenges academia puts forth. Furthermore, she has been and

continues to be my greatest champion and one of my best friends. I am truly thankful for all that she has done for me and remain thankful, as that list continues to grow and grow.

TABLE OF CONTENTS

FIGURES AND TABLES

FIGURES

TABLES

CHAPTER I

INTRODUCTION

Flatfishes (Pleuronectiformes), the commercially important soles, flounders and halibuts, exhibit one of the most remarkable vertebrate metamorphoses known. In this transformation, one eye of a bilaterally symmetrical larva migrates to the opposite side of the cranium, resulting in highly asymmetrical juvenile and adult forms (Fig. 1-1). Post-metamorphosis, adults are easily recognizable, with both eyes on one side of their head and with the other, eyeless, side having either greatly reduced pigmentation, or lacking pigmentation entirely.

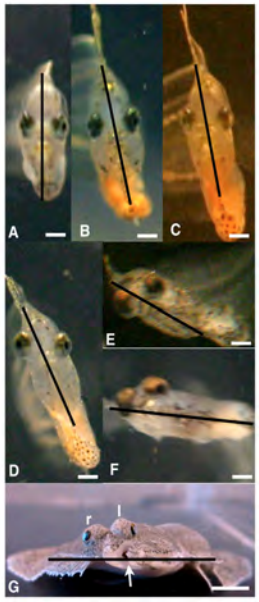


Figure 1-1. Flatfish metamorphic stages (A-G). Viewed head-on. Black bar indicates midline, with right eye migrating dorsally to cross it. Arrow in G denotes terminal mouth position, with right (r) and left (l) eyes above. Bars in A-F represent 0.2 mm and 0.5 mm in G; from Schreiber (2006).

These fishes are the only vertebrates to deviate so dramatically from a bilaterally symmetrical body plan, and all known species (approximately 700 classified into 14 families) exhibit this cranial asymmetry. There is a strong phylogenetic signal for degree of asymmetry within Pleuronectiformes, with soleids and cynoglossids being extremely asymmetrical, whereas psettodids, pleuronectids and paralichthids have symmetrical jaws and well-developed blind-side pectoral fins.

There is, however, no strong phylogenetic pattern predicting which eye will migrate (Chapleau, 1993; Berendzen and Dimmick, 2002). Some groups consist of almost entirely dextral (right-eyed; left side is blind) or sinistral (left-eyed) taxa, but even within those clades there exists variation in “handedness”. This variation is not necessarily fixed

among species; there are many examples where geographic location dictates whether dextrality or sinistrality is exhibited post-metamorphosis.

Because all flatfishes exhibit this bizarre morphology and variation in both the degree of asymmetry and handedness exists within the order, this group provides multiple tests of hypotheses regarding the evolution of bilateral asymmetry and the genetic mechanism underlying this phenomenon. Unfortunately, undertaking such studies has been difficult because of three major issues confounding pleuronectiform phylogenetics: 1) monophyly of the assemblage is weakly supported, 2) the sister group of flatfishes is unknown, and 3) relationships of the major groups within the order remain mostly unresolved.

Chapleau (1993) was the first to evaluate pleuronectiform relationships within a cladistic framework, but was unable to test outgroup hypotheses since all putative characters of adult flatfish morphology he identified were either autapomorphic or pleisiomorphic and therefore uninformative. Although he did not include multiple outgroups for comparison he designated the order as monophyletic based on the following three synapomorphies: 1) metamorphosis with eye migration leading to an asymmetrical cranium, 2) a dorsal fin that overlaps the neurocranium and 3) the presence of the recessus orbitalis (RO; an accessory organ of the eye). Presence of the latter character has yet to be confirmed in all major flatfish lineages. Although presence of the RO as a pleuronectiform synapomorphy is tentative, flatfish monophyly had not been challenged until molecular phylogenetic comparisons between flatfishes and multiple outgroups became feasible and yielded some controversial results.

The controversy usually centered on the placement of the spiny turbot, genus

Psettodes. Often these three species, all of which exhibit incomplete transformation (the migrating orbit does not fully cross the dorsal margin of the head during metamorphosis) would be placed outside the remainder of flatfishes, sometimes with high statistical support, but most often with low support (Chen, 2003; Smith and Craig, 2007; Dettai and Lecointre, 2006; Li et al., 2009, Betancur-R et al., 2013a,b; Near et al., 2013). In most cases this could easily be explained as resulting from low taxon sampling within the order. However, even when a large number of flatfishes were included (Betancur-R et al., 2013a,b) resolution was lacking.

Despite the ambiguity regarding the position of *Psettodes*, these phylogenies showed enough resolution in the placement of Pleuronectiformes among acanthomorphs that the appropriate outgroups for comparison could be narrowed down considerably. Chapleau (1993) recognized that because no derived morphological characters shared between flatfishes and any other clade had been identified, testing sister-group hypotheses in a parsimony framework was impossible. The large-scale molecular studies provided the structure necessary to carry out an analysis specifically designed to test not only the monophyly of Pleuronectiformes, but also hypotheses regarding the second major problem in flatfish phylogenetics: the identity of the sister group. Betancur-R et al. (2013b) were the first to carry out such a study using DNA sequence data analyzed in a probabilistic framework, but they were still unable to provide high support for pleuronectiform monophyly or any sister-group hypotheses. As with the earlier studies, it was the position of *Psettodes* that caused the ambiguity. They identified gene tree discordance and investigated base compositional bias as a cause, but found that those factors could only explain some of the incongruence. Another potential source of

systematic error, the habitually overlooked substitution saturation, has never been investigated as a source of bias potentially causing lack of resolution at the base of Carangimorpha (*sensu* Li et al., 2009). Neither has the use of rhodopsin 1 (*rho*) in fish phylogenetics, which is both ubiquitous and controversial. Because the signal provided by *rho* (a gene that codes for a visual pigment) is consistently found to be incongruent with others (Niemiller et al., 2013; Tornabene et al., 2013), and this is presumably due to strong selection pressures leading to convergence in species inhabiting similar photic environments, it has been deemed an inappropriate phylogenetic marker for not just fishes, but all aquatic taxa (Larmuseau, 2010). If this is true, then the results of all the molecular studies described thus far, especially those where the position of *Psettodes* was ambiguous, need to be re-examined. If it is not true, that needs to be established as well, since identifying selection and incongruent signal are most likely not unique to *rho* and could therefore present a more widespread problem in molecular phylogenetics.

Clearly resolving the first two issues in flatfish phylogenetics has been difficult using both morphological and DNA sequence data and appears to be contingent on the placement of *Psettodes*. This lack of resolution, is not limited to hypotheses regarding monophyly and sister groups, but extends to phylogenetic structure within the order as well. There exists controversy regarding the position(s) of the highly asymmetrical families, Soleidae and Cynoglossidae as well as the monophyly and placement of the "garbage bin taxon", Paralichthyidae (Hensley and Ahlstrom, 1984; Chapleau, 1993; Berendzen and Dimmick, 2002; Betancur-R et al., 2013b). These issues are of particular interest, not only because solving them allows for a robust phylogeny that can be used to test hypotheses regarding the evolution of bilateral asymmetry, but because they may

highlight the necessity of an often ignored suite of morphological characters in systematic ichthyology: the characteristics of larval fishes.

Flatfishes, like most other marine teleosts, have an early life history that is ecologically, physiologically, behaviorally and anatomically distinct from the adults and the use of those characters in phylogenetics has been heralded at least in part for their assumed independence from adult characters (Cohen, 1984; Roje, 2010 and others). The larvae of most larval flatfishes have been described and characters such as elongate dorsal fin rays, head spines, and a trailing gut, have been identified as potentially informative, yet they have never been included in a phylogenetic analysis of the order. Obtaining convergence on a single robust hypothesis regarding pleuronectiform phylogeny when using characters of adult morphology and sequence data remains elusive, and investigating the signal provided by these characters may prove to be fruitful, providing some much needed resolution.

To resolve the three issues in pleuronectiform phylogenetics described above, my dissertation research is focused on: 1) evaluating the effects non-neutral markers on phylogeny estimation, principally rhodopsin1 (*rho*), 2) rigorously testing both flatfish monophyly and sister-group hypotheses and 3) resolving relationships within the order and re-examining characters of adult morphology and comparing them to often overlooked larval characters in light of new phylogenetic hypotheses.

The first study, Chapter II, is comprised solely of a published work (Roje, 2014) where I used an empirical dataset to determine if gene tree congruence can be predicted by the ability to detect selection. I used previously published sequence data from 78 acanthomorph (including flatfishes) taxa for *rho*, *rnf213*, *irbp* and *mll*, performed Z-tests

for neutrality (H_0: dN = dS), and compared neutral versus non-neutral markers for congruence using tree distance metrics and topology testing.

In the second study, Chapter III, I optimized new molecular markers and sequenced them along with *rho* and *rnf213* for 58 flatfishes and 90 putative outgroups to test monophyly, intraordinal relationships and sister-group hypotheses. I then analyzed those sequences along with data from a previous study to determine possible causes for gene tree incongruence or phylogenetic error.

To assess gene tree discordance, I carried out various topology tests comparing individual genes and their gene trees to the concatenated dataset and its ML. I then considered base compositional bias and substitution saturation as potential sources of any incongruence and/or low support for clades. I evaluated the former by treating GC content as a continuous character and mapping it on individual gene trees, as well as the phylogeny inferred using all data. Then, I assessed nucleotide substitution saturation using change-point analysis on plots of third codon position transitions versus corrected genetic distance; the relationship of taxa separated by a distance below the level 1 change-point (where the curve begins to plateau) can be treated as inferred from unsaturated data.

Finally, in the third study (Chapter IV) I examined the signal provided by Chapleau's (1993) characters of adult morphology and larval characters by analyzing them using MP and ML methods and comparing the results to new and previous hypotheses of within-order relationships. Further, I generated an ultrametric tree and used the adult and larval characters to test whether life history is correlated with phylogenic signal. Then I compared likelihoods and estimated rates of state change (inferred using

the Mk 1 equal rates model) when branch length information is included and excluded to determine if these morphological characters provide additional support for hypotheses of relationships among major pleuronectiform groups, even though they may be in violation of the condition of low rates on ML ancestral character estimation (ACE). Finally, based on the results of the ACEs, I investigated characters of larval morphology as potential sources of support for the placement of Pleuronectiformes within Acanthomorpha.

Taken together, these studies constitute a comprehensive, detailed phylogenetic analysis of Pleuronectiformes. I resolve the question of flatfish monophyly, reconstruct relationships within the order, recover its sister taxon, and resolve its placement among Acanthomorpha. I also perform detailed analyses to evaluate the suitability and phylogenetic informativeness of molecular, adult morphological, and larval morphological characters. I consider the potential effects of selection, base composition, and saturation for molecular sequence data in phylogenetics, issues which pertain to all molecular phylogenetic studies, not just those of pleuronectiform or even acanthomorph fishes.

REFERENCES

Berendzen, Peter B., and Walter Wheaton Dimmick. 2002. “Phylogenetic Relationships of Pleuronectiformes Based on Molecular Evidence.” *Journal Information* 2002 (3).

Betancur-R, Ricardo, Richard E. Broughton, Edward O. Wiley, Kent Carpenter, J. Andres Lopez, Chenhong Li, Nancy I. Holcroft, et al. 2013. “The Tree of Life and a New Classification of Bony Fishes.” *PLoS Currents* 5 (April). doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, Ricardo, Chenhong Li, Thomas A. Munroe, Jesus A. Ballesteros, and Guillermo Ortí. 2013. “Addressing Gene Tree Discordance and Non-Stationarity to Resolve a Multi-Locus Phylogeny of the Flatfishes (Teleostei: Pleuronectiformes).” *Systematic Biology*, syt039.

Chapleau, François. 1993. “Pleuronectiform Relationships: A Cladistic Reassessment.” *Bulletin of Marine Science* 52 (1): 516–40.

Chen, Wei-Jen, Céline Bonillo, and Guillaume Lecointre. 2003. “Repeatability of Clades as a Criterion of Reliability: A Case Study for Molecular Phylogeny of Acanthomorpha (Teleostei) with Larger Number of Taxa.” *Molecular Phylogenetics and Evolution* 26 (2): 262–88. doi:10.1016/S1055-7903(02)00371-8.

Cohen, Daniel M. 1984. “Ontogeny, Systematics, and Phylogeny.” *Ontogeny and Systematics of Fishes. Amer. Soc. Ichthyol. Herpetol., Spec. Publ* 1: 7–11.

Dettai, Agnès, and Guillaume Lecointre. 2005. “Further Support for the Clades Obtained by Multiple Molecular Phylogenies in the Acanthomorph Bush.” *Comptes Rendus Biologies* 328 (7): 674–89. doi:10.1016/j.crvi.2005.04.002.

Hensley, D. A., and E. H. Ahlstrom. 1984. “Pleuronectiformes: Relationships.” *Ontogeny and Systematics of Fishes*, 670–87.

Larmuseau, Maarten HD, Tine Huyse, Kim Vancampenhout, Jeroen KJ Van Houdt, and Filip AM Volckaert. 2010. “High Molecular Diversity in the Rhodopsin Gene in Closely Related Goby Fishes: A Role for Visual Pigments in Adaptive Speciation?” *Molecular Phylogenetics and Evolution* 55 (2): 689–98.

Li, B., A. Dettaī, C. Cruaud, A. Couloux, M. Desoutter-Meniger, and G. Lecointre. 2009.

"RNF213, a New Nuclear Marker for Acanthomorph Phylogeny." *Molecular Phylogenetics and Evolution* 50 (2): 345–63.

Niemiller, Matthew L., Benjamin M. Fitzpatrick, Premal Shah, Lars Schmitz, and Thomas J. Near. 2013. "Evidence for Repeated Loss of Selective Constraint in Rhodopsin of Amblyopsid Cavefishes (Teleostei: Amblyopsidae)." *Evolution* 67 (3): 732–48.

Roje, Dawn M. 2014. "Evaluating the Effects of Non-Neutral Molecular Markers on Phylogeny Inference." *PloS One* 9 (2): e87428.

Smith, W.L., and M.T. Craig. 2007. "Casting the Percomorph Net Widely: The Importance of Broad Taxonomic Sampling in the Search for the Placement of Serranid and Percid Fishes." *Journal Information* 2007 (1).

Tornabene, Luke, Yongjiu Chen, and Frank Pezold. 2013. "Gobies Are Deeply Divided: Phylogenetic Evidence from Nuclear DNA (Teleostei: Gobioidei: Gobiidae)." *Systematics and Biodiversity* 11 (3): 345–61.

CHAPTER II

EVALUATING THE EFFECTS OF NON-NEUTRAL MARKERS OF PHYLOGENY ESTIMATION

ABSTRACT

Nucleotide substitution models used in molecular phylogenetics do not account for nucleotide sequences evolving under selection, yet selection is rarely tested for. If non-neutral markers violate these models (i.e. non-independence of sites), it is expected that their reconstructed topologies be incongruent with those inferred from neutral ones and conclusions made from those phylogenies should be reexamined. Using rhodopsin as a phylogenetic marker has recently been called into question for exactly this reason. Rhodopsin is assumed to have evolved under strong positive selection for organisms that inhabit similar aquatic environments, making it unsuitable for the phylogenetics of aquatic organisms, but it is unclear what the effects of non-neutrality on phylogeny estimation are. To evaluate potential incongruence of neutral versus non-neutral markers, and the notion that rhodopsin should not be used in the molecular phylogenetics of fishes, a molecular dataset of 78 acanthomorph taxa and sequences from four nuclear, protein coding loci (including rhodopsin), were examined. Only one marker was found to be neutral while the remaining tests, for all other loci, rejected the null hypothesis of neutrality. To evaluate the possible effect(s) of positively versus negatively selected sites, the three non-neutral markers were analyzed to determine the presence of positively and negatively selected codons. To determine congruence in topology among ML trees inferred by individual neutral and non-neutral markers, as well as the combined (concatenated) dataset, tree, comparisons of distances among trees and hypothesis

(topology) testing were carried out. Results of the tree distance metrics and topology testing support the notion that neutrality alone does not determine congruence in topology, and those data that are inferred to have evolved under selection should not necessarily be excluded. In addition, the number of sites inferred to have evolved under positive selection does not predict congruence with other markers or the topology inferred with the concatenated dataset.

Roje, Dawn M. 2014. "Evaluating the Effects of Non-Neutral Molecular Markers on Phylogeny Inference." *PloS One* 9 (2): e87428.

OPEN ACCESS Freely available online

PLOS ONE

Evaluating the Effects of Non-Neutral Molecular Markers on Phylogeny Inference

Dawn M. Roje*

Department of Ichthyology, American Museum of Natural History, New York, New York, United States of America

Abstract

Nucleotide substitution models used in molecular phylogenetics do not account for nucleotide sequences evolving under selection, yet selection is rarely tested for. If non-neutral markers violate these models (i.e. non-independence of sites), it is expected that their reconstructed topologies be incongruent with those inferred from neutral ones and conclusions made from those phylogenies should be reexamined. Using rhodopsin as a phylogenetic marker has recently been called into question for exactly this reason. Rhodopsin is assumed to have evolved under strong positive selection for organisms that inhabit similar aquatic environments, making it unsuitable for the phylogenetics of aquatic organisms, but it is unclear what the effects of non-neutrality on phylogeny estimation are. To evaluate potential incongruence of neutral versus non-neutral markers, and the notion that rhodopsin should not be used in the molecular phylogenetics of fishes, a molecular dataset of 78 acanthomorph taxa and sequences from four nuclear, protein coding loci (including rhodopsin), were examined. Only one marker was found to be neutral while the remaining tests, for all other loci, rejected the null hypothesis of neutrality. To evaluate the possible effect(s) of positively versus negatively selected sites, the three non-neutral markers were analyzed to determine the presence of positively and negatively selected codons. To determine congruence in topology among ML trees inferred by individual neutral and non-neutral markers, as well as the combined (concatenated) dataset, tree, comparisons of distances among trees and hypothesis (topology) testing were carried out. Results of the tree distance metrics and topology testing support the notion that neutrality alone does not determine congruence in topology, and those data that are inferred to have evolved under selection should not necessarily be excluded. In addition, the number of sites inferred to have evolved under positive selection does not predict congruence with other markers or the topology inferred with the concatenated dataset.

Citation: Roje DM (2014) Evaluating the Effects of Non-Neutral Molecular Markers on Phylogeny Inference. PLoS ONE 9(2): e87428. doi:10.1371/journal.pone.0087428

Editor: Giovanni G. Vendramin, CNR, Italy

Received August 30, 2013; **Accepted** December 21, 2013; **Published** February 18, 2014

Funding: This author has no support or funding to report.

Competing Interests: This author has declared that no competing interests exist.

* E-mail: droje@amnh.org

Introduction

There exist a large number of genes used to infer phylogeny of which most substitution models describing their evolution are independent, finite sites models implicitly assuming neutrality. Although violations of these assumptions should lead to unsound phylogeny inference, this has not been tested on a real dataset. If the evolution of sites is correlated, as they would be for markers that have evolved under strong pressures of natural selection, convergent molecular evolution at those sites can lead to erroneous inference of phylogeny and should be inconsistent with the phylogenetic signal inferred from neutral ones. In fact, non-neutral convergent molecular evolution has been detected in mitochondrial genes and has been shown to cause problems in phylogenetic inference [1], namely that the mtDNA phylogeny is incongruent with the more robust phylogeny inferred from nuclear DNA. The practice of testing for selection regardless of species gene tree incongruence, however, is almost nonexistent. Is it highly unlikely that all of the numerous molecular markers sequenced for various taxonomic groups are all neutral. If they are not, and selection does negatively affect phylogenetic inference (either by violating assumptions of the models or representing convergent evolution) then all of the conclusions made from those studies would be untrustworthy. Many would argue that those studies should be trusted because sequences from multiple genes were (usually) analyzed, and that when multiple independent lines of evidence are analyzed they are anticipated to converge on one hypothesis.

One way of testing this notion is to compare inferred phylogenies with the correct, true tree. The only way a topology can be considered correct is if it is derived through simulation. Hagstrom et al. [2] and Hang et al. [3] carried out such simulation studies and determined that if selection occurred between nodes, parsimony and neighbor-joining methods were not only able to recover the true tree but their performance was improved. These studies, however, only examined the reconstruction of four taxon statements—a trivial number of taxa when compared to those used in most phylogenetic studies.

Recently, Larmuseau et al. [4] investigated the effects of a marker, rhodopsin1 (referred to in their study as *rh1* and in others as well as this study as *rho*), under selection on the phylogeny estimation of a real group of organisms: the sand gobies of the teleost fish family, Gobiidae. They determined that similarities in *rho* sequences were due to convergence of species inhabiting similar light environments by linking functional regions of the protein to particular photic habitat and by constructing and comparing phylogenies based on *rho* and the 'neutral' genes, 12S and 16S. They determined that "The discordant patterns between

the 'neutral' and the RH1 phylogeny indicate that rhodopsin is not an optimal marker for phylogeny reconstruction of aquatic taxa. Therefore, it is recommended to avoid the use of rhodopsin as a 'neutral' marker in phylogenetic studies," (p. 695). If the assertions of Larmuseau et al. [4] are correct, conclusions made from the many fish phylogenies in which this marker has been used as part of a "total-evidence" (concatenated dataset) approach [5–10] should be considered suspect. However, tests for neutrality were only carried out on the *rho* sequences, ignoring the possibility that selection could also be detected for the supposed neutral *12s* and *16s* sequences. Additionally, although the 'neutral' and *rho* phylogenies appeared to be incongruent, no tree distance metrics were generated and no hypothesis testing was carried out.

Larmuseau et al.'s [4] claim that phylogenetic signal provided by rhosopsin is incongruent with different markers has been corroborated by others working on goby phylogeny [5,6]. Tornabene et al. [6] noted that the tree inferred from *rho* data was different from the others, but did not test any of the markers for selection, nor did they test whether the difference in topologies were statistically significant. Because of this they could not prove or disprove Larmuseau's et al.'s [4] assertion that unrelated goby species inhabiting similar photic enviormnments will converge on similar *rho* sequences. Niemiller et al. [5] did test for selection in *rho* and rejected the null hypothesis of neutral evolution in certain goby lineages. These tests, however, were only carried out for *rho* and they did not evaluate any of the other individual markers for topology congruence. It is the goal of this study to test for topological incongruence between neutral and non-neutral nuclear markers, as well as to test the specific notion that *rho* is an inappropriate marker for the phylogeny reconstruction in fishes.

Materials and Methods

Tests of selection are based on assessing the ratio of synonymous to non-synonymous nucleotide substitutions, therefore, only protein coding loci were chosen for this study. To test the notion that rhodopsin may not be an appropriate marker for phylogenetic analysis of aquatic organisms [4], the acanthomorph dataset of Li et al. [11] was selected. Those data include ring finger protein 213 (*rnf213*), mixed-lineage leukemia (*mll*), inverted repeat binding protein (*irbp*) and rhodopsin (*rho*) sequences; the former was sequenced for Li et al.'s (2009) study, the *rho* sequences were generated by Chen et al. [7], and the *mll* and *irbp* sequences (with the exception of a few sequenced by Li et al. [11]) were generated by Dettai and Lecointre [8]. Only taxa with accessioned sequence data for all four markers were included. Of those, *Epinephalus aeneus* was excluded because the published *irbp* sequence (AY362227) BLASTed [12] against the NCBI nucleotide database (with default parameters) as *mll*. All sequences used in this study were retrieved from GenBank.

All four markers were aligned using MUSCLE [13] implemented in Geneious (Biomatters Ltd., Auckland, New Zealand) with full penalty for terminal gaps, a gap open score of −1, and a maximum of eight refinement iterations. Model selection, utilizing the AICc (corrected Akaike Information Criterion), and phylogeny inference, using the ML criterion, were carried out in Treefinder [14]. Five alignments were analyzed, comprising each of the four individual markers and a concatenated dataset.

To test for selection, the Nei-Gojobori method (Proportion), implemented in MEGA 4.0 [15] was used to estimate the number of non-synonymous substitutions per non-synonymous sites (dN) and the number of synonymous substitutions per synonymous site (dS) for each of the four markers. Variances were generated by bootstrapping (5000 replicates) the data and then the null hypothesis of neutral evolution (H_0: dN = dS) was tested using a Z-test covering the overall average (per marker) of dN and dS.

To detect the type of selection (positive or negative) per site the fixed effects likelihood (FEL) and the random effects likelihood (REL) methods were implemented in HyPhy [16,17]. Both methods are based on ML estimates for the parameters of a nucleotide substitution and codon model, testing whether dN/dS >1 per site. Unlike REL, FEL estimates are conditional on a specific phylogeny; for these FEL analyses the TE phylogeny was used. The REL method assumes distributions for the synonymous and nonsynonymous rates and identifies positively selected sites using empirical Bayes factors. To determine significance, a cutoff p-value of 0.05 was used for FEL and an acceptance ratio of 0.05 for REL.

Tree distance metrics were generated and statistical tests were carried out on all five (four individual markers plus the concatenated dataset) topologies to assess their pairwise similarity or difference. Two tree distance metrics were used to evaluate topology congruence: the symmetric distance of Robinsons and Foulds [18], carried out in Phylip [19], and the SPR (subtree pruning and regrafting) heuristic distance [20], implemented in TNT [21]. The Symmetric Difference measures how many partitions are on one tree and not the other, whereas SPR distance measures the minimum number of SPR moves required to transform one tree into another. For both tree metrics, trees were treated as unrooted and for SPR distances, 2000 replications per comparison were carried out.

Because a tree metric is not a statistical hypothesis test, three paired sites tests, the KH [22], AU [23] and SH [24] tests were carried out. All tests compared likelihood differences among tree topologies (the five generated for this study) to the empirical variation in log likelihoods for a given dataset, with the AU and SH tests approximately correcting for multiple trees. The null hypothesis (H_0: all topologies for comparison are equally good explanations of the data) was rejected when $P<0.05$. The hypothesis tests were implemented in Treefinder [14] using the RELL (resample estimated log-Likelihood) nonparametric bootstrap method for the KH and SH tests with 50,000 replicates to generate the null distributions. A multiscale bootstrap technique was used for the AU test, which is considered to exhibit the least amount of bias and is less conservative than the SH test [14]. The models used to calculate the likelihoods for each topology were the same as those chosen using the AICc in Treefinder [16].

To determine if incongruence was caused only by the presence of positively selected codons, any marker that was found to be incongruent had its positively selected sites (all those sites detected using FEL and REL) removed and another round of KH, AU and SH tests was carried out in Treefinder [14]. If sites under positive selection are the only cause of incongruence, their removal should result in failure to reject the null hypothesis of the paired sites tests.

Results

The aligned matrices consisted of 78 OTUs with the following lengths per marker: 991 bp (603 informative sites) of *rnf213*, 542 bp (364 informative sites) of *mll*, 716 bp (488 informative sites) of *irbp*, and 759 bp (404 informative sites) of *rho*. The final concatenated, TE, dataset consisted of 3008 bp (1859 informative sites). All four alignments were straightforward in that no internal stop codons were detected and all indels corresponded to one or more codons.

Using the AICc, Treefinder [14] indicated that the best-fit nucleotide substitution model (given a parameter penalty and correction for sample size) for the combined dataset (all four genes)

is GTR+G. Individually, using the same approach, the following models were chosen: HKY+G for *rho*, J2+G for *irbp* and GTR+G for *mll*; and *rnf213*; those models were used to generate the ML trees. The maximum (negative) log-likelihoods, determined by Treefinder [12], for each of the five trees were as follows: −18045.83 for *rnf213*, −11770.489 for *mll*, −17056.246 for *irbp*, −14702.74 for *rho* and −62719.57 for the tree inferred from the concatenated dataset (from now on referred to as the "Con" tree or topology). The ML trees inferred from the concatenated dataset, as well as each individual marker, with 1000 bootstrap replicates is available in File S1.

The overall test of selection resulted in the following p-values: 0.639 for *rnf213* and 0.000 for the remaining three markers. *Rnf213* was the only marker that failed to reject ($P<0.05$) the null hypothesis (H_0: dN = dS) of neutral evolution; all others can be treated as having evolved under some pressure(s) of selection.

The results of the tree distance comparisons are summarized in Table 1 (symmetric differences) and Table 2 (SPR distances). The shortest distances of all pairwise comparisons were that of *rnf213* (the neutral marker) to the Con tree (symmetric distance of 70 and SPR distance of 19). The *irbp* tree was furthest from the Con tree with a symmetric distance of 110 and an SPR distance of 46 or 47. Comparisons among the individual (inferred from one marker) topologies resulted in symmetric distances ranging from 98 (*rnf213*/*mll*) to 122 (*irbp*/*rho*) and SPR distances ranging from 36, 37 (*rho*/*mll*) to 61, 63 (*irbp*/*rho*). The was no clear pattern indicating that the non-neutral markers (*mll*, *irbp* and *rho*) were further from the neutral one (*rnf213*) than either was to each other. For example, the symmetric distance from *rnf213* to *rho* was 112, but the distance from *mll* to *irbp* was the same. *Rho* was the furthest from the neutral marker, but those distances (symmetric and SPR distances) were not greater than the distances from *rho* to *irbp*.

The symmetric and SPR distances were not entirely consistent in rank; they differed in the tree distances from *mll* to *rho* and vice versa. The symmetric distance between *mll* and *rho* was 120, the second largest distance among all pairwise symmetric distances comparisons. The SPR distances, however, had *mll* and *rho* 36 or 37 steps away from each other, the second smallest SPR distance among those comparisons.

Tree distance metrics cannot be used to reject the hypothesis that two trees are not significantly different. Because of this, KH, AU and SH tests were conducted on differences in log-likelihoods of the five topologies and datasets; the results are summarized in Table 3.

All three paired sites tests were consistent in their rejection ($P<0.05$) or acceptance ($P>0.05$) of the null hypothesis, with the SH test being the most conservative. In all pairwise comparisons, except for the *rnf213*/Con, *mll*/Con and *irbp*/Con comparisons, the hypothesis that the topologies are explained equally well by the data was rejected. Of the three that failed to reject, the largest p-values were associated with comparison of *rnf213* to the Con tree. All tests indicate that the *rho* tree is significantly different from not only all other individual marker trees, but also the Con tree.

Of the three non-neutral markers, *rho* had the most sites where positive selection was detected (see Table 4). Two sites (codon positions 121 and 165) were detected by FEL and four sites (codon positions 118, 121, 165 and 169) were detected by REL. Because the two sets of sites overlap, all four codons detected by REL were removed from the Rho alignment that was then used in three

Table 1. Pairwise Symmetric Distances between topologies; non-neutral markers are in boldface.

	rnf213	*mll*	*irbp*	*rho*	Concatenated
rnf213	—	98	110	112	70
mll	98	—	112	120	88
irbp	110	112	—	122	110
rho	112	120	122	—	102
Concatenated	70	88	110	102	—

doi:10.1371/journal.pone.0087428.t001

Table 2. Pairwise SPR distances.

	rnf213	*mll*	*irbp*	*rho*	Concatenated
rnf213	—	41	48	48	19
mll	40	—	49	36	41
irbp	50	43	—	63	47
rho	51	37	61	—	42
Concatenated	19	41	46	45	—

First number in cell represents the distance from the tree in the left column to the tree in the top row; the second number is the distance computed from the reverse direction; non-neutral markers are in boldface.

doi:10.1371/journal.pone.0087428.t002

Table 3. Results of the KH, AU and SH tests.

Data	Tree	KH Test (P-value)	AU Test (P-value)	SH Test (P-value)
rnf213	*rnf213*	—	—	—
	mll	0.000	0.000	0.000
	irbp	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	TE	**0.220**	**0.222**	**0.661**
mll	*rnf213*	0.000	0.000	0.003
	mll	—	—	—
	irbp	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	TE	**0.058**	**0.057**	**0.198**
irbp	*rnf213*	0.001	0.000	0.003
	mll	0.000	0.000	0.000
	irbp	—	—	—
	rho	0.000	0.000	0.000
	TE	**0.111**	**0.098**	**0.345**
rho	*rnf213*	0.000	0.000	0.000
	mll	0.000	0.000	0.000
	irbp	0.000	0.000	0.000
	rho	—	—	—
	TE	0.000	0.000	0.000
Concatenated	*rnf213*	0.000	0.000	0.000
	mll	0.000	0.000	0.000
	irbp	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	TE	—	—	—

Boldface rows represent those tree topologies that are not significantly different from the phylogeny with the best likelihood score.

doi:10.1371/journal.pone.0087428.t003

Table 4. Summary of results from the FEL and REL analyses of the three non-neutral markers.

	FEL		REL	
	Positive	Negative	Positive	Negative
irbp	1	179	0	185
mll	1	150	1	158
rho	2	168	4	193

doi:10.1371/journal.pone.0087428.t004

paired sites test to determine congruence with the *rnf213*, *irbp*, *mll* and Con topologies. All four tests resulted in p-values of 0.000, rejecting ($P<0.50$) the null hypothesis that the topologies for comparison are equally good explanations of the modified *rho* data.

Discussion

It is important to remember that the Con topology is not (necessarily) the "true tree", but the sum of the contributions of the four individual datasets. For this reason, the closeness in tree distance of *rnf213*, the neutral marker, to the TE phylogeny can be attributed to the proportion of data it contributed to the concatenated dataset (603 of 1859 informative sites); that amount being the greatest contribution of all four markers. Therefore, comparisons of topology congruence of the trees inferred by the individual markers will be the focus of this discussion, in that they are better indications of the behavior of non-neutral versus neutral markers.

Both tree distance metrics show that the amount of shared partitions and number of SPR moves to edit one topology into the other is no shorter or longer among the non-neutral markers than they are to *rnf213*. This is easily seen upon examination of the distances of the *irbp* topologies to all others. Both metrics have the *irbp* topology farthest from the *rho* tree and both markers failed the Z-test of selection (H_0: dN = dS). Also, the closest distance between individual marker topologies was that between *rnf213* (neutral) and *mll* (non-neutral), with a symmetric difference of 88. *Rho* and *irbp* are closer to *rnf213* than they are to each other. This lack of clustering of topologies inferred from non-neutral markers supports the notion that non-neutrality alone does not warrant the omission of these data from phylogenetic analyses.

The hypothesis tests described above provide additional support for the inclusion of non-neutral molecular data in phylogeny inference. The p-values for the *rnf213*, *mll* and *irbp* data given the Con topology were all above 0.05, indicating that the Con topology is an equally good explanation of neutral as well as non-neutral data. It is worth noting, however, that *rnf213* had the greatest p-values, but, just as with the tree distance metrics, this can be attributed to its large contribution to the concatenated dataset.

The *rho* data had the third largest contribution of informative sites to the combined concatenated dataset but was significantly different from all other topologies. This, combined with the incongruent behavior of the *rho* topology in the tree distance comparisons, supports the assertion of Larmuseau et al. [4] that phylogenetic signal provided by Rho is disparate from that of other markers. Although, it is important to remember that this is only in one case. To make these conclusions more generalizable it would be necessary investigate the effects of *rho* in many different groups of aquatic organisms and compare those results with that of terrestrial taxa. The fact that the Rho data are incongruent with both neutral *and* non-neutral topologies, however, allows for the rejection of Larmuseau et al.'s [4] claim that this difference is due to natural selection since selection alone was not able to predict congruence. In addition, the presence of positively selected sites in the Rho alignment does not predict phylogenetic incongruence. If transformations at those sites are homoplasies then the signal they provide is misleading, but their removal does not make Rho congruent with the other makers or the Con topology—something else has contributed to the unique evolutionary history of this marker. Of the four markers evaluated here, *rho* is the only G-protein coupled receptor (GPCR); it would be valuable to compare the signal provided by GPCR markers to those from other protein families.

Incongruent signal provided by different loci is precisely why the practice of molecular phylogenetics requires the use of multiple unlinked markers. It is expected that that any misleading information, such as convergence, will be ameliorated by the remaining data, which is exactly what occurred here. Although the phylogenetic signal of Rho was inconsistent with the other markers, it was also inconsistent for the Con topology, which is usually regarded as the most favorable phylogeny. This shows that the mere state of being neutral or non-neutral does not warrant exclusion of data. This case study, while highlighting the potential problems with the use of a particular marker on phylogeny estimation also shows that the common practice of using multiple markers can mitigate its potentially deleterious effects.

Supporting Information

File S1 Nexus tree files with bootstrap support for all five trees presented; Con, *rnf213*, *mll*, *irbp* and *rho*.
(ZIP)

Acknowledgments

I am extremely grateful to Ward Wheeler, John Sparks and Melanie Stiassny for their support and guidance throughout these analyses and during preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: DR. Performed the experiments: DR. Analyzed the data: DR. Contributed reagents/materials/analysis tools: DR. Wrote the paper: DR. Sole contributor of all work from inception to submission: DR.

References

1. Castoe TA, de Koning APJ, Kim HM, Gu W, Noonan BP, et al. (2009) Evidence for an Ancient Adaptive Episode of Convergent Molecular Evolution. PNAS 106: 8986–8991.
2. Hagstrom GI, Hang DH, Ofria C, Torng E (2004) Using Avida to test the effects of natural selection on phylogenetic reconstruction methods. Artif Life 10: 157–166.
3. Hang D, Ofria C, Schmidt TC, Torng E (2003) The effect of natural selection on phylogeny algorithms. In: Genetic and Evolutionary Computation Conference, Chicago, Il, pp. 13–24.
4. Larmuseau MHD, Huyse T, Vancampenhout K, Van Houdt, JKJ, Volckaert PAM (2010) High molecular diversity in the rhodopsin gene in closely related goby fishes: A role for visual pigments in adaptive speciation? Mol Phyl Evol 55: 689–698.

5. Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near T (2012) Evolution 67–3: 732–748.
6. Tornabene L, Chen Y, Pezold F (2013) Gobies are deeply divided: phylogenetic evidence from nuclear DNA (Teleostei: Gobioidei: Gobiidae). Syst Biodivers 11: 345–361.
7. Chen WJ, Bonillo C, Lecointre G (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol Phyl Evol 26: 262–288.
8. Dettai A, Lecointre G (2005) Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. Comptes Rendus Biologies 328: 674–689.
9. Taylor MS, Hellberg ME (2005) Marine radiations at small geographic scales: speciation in neotropical reef gobies (*Elacatinus*). Evolution 59: 374–385.
10. Mayden RL, Tang KL, Conway KW, Freyhof J, Chamberlain S, et al. (2007) Phylogenetic relationships of *Danio* within the order Cypriniformes: a framework for comparative and evolutionary studies of a model species. J Exp Zool B Mol Dev Evol 308B: 642–654.
11. Li B, Dettaï A, Cruaud C, Couloux A, Desoutter-Meniger M, et al. (2009) RNF213, a new nuclear marker for acanthomorph phylogeny. Mol Phyl Evol 50: 345–363.
12. Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
13. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc Acid Res 32: 1792–1797.
14. Jobb G (2008) TREEFINDER version of October 2008. Munich, Germany. Available: www.treefinder.de.
15. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.
16. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21(5): 676–679.
17. Kosavosky Pond SL, Frost SDW (2005) Not so different after all: comparisons of methods for detecting amino acid sites under selection. Mol Biol and Evol 22: 1208–1222.
18. Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Math. Biosci 53: 131–147.
19. Felsenstein J (1993) PHYLIP: Phylogeny Inference Package. Version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
20. Goloboff P (2008) Calculating SPR distances between trees. Cladistics 24: 591–597.
21. Goloboff P, Farris J, Nixon K (2003) T.N.T. Tree Analysis using New Technology. Program and documentation. Available: http://www.zmuc.dk/public/phylogeny/tnt.
22. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence, data and the branching order in Hominoidea. J Mol Evol 29: 170–179.
23. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol 51: 492–508.
24. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16: 1114–1116.

CHAPTER III

MONOPHYLY AND PHYLOGENETIC PLACEMENT OF PLEURONECTIFORMES

ABSTRACT

Flatfishes (order Pleuronectiformes) are a group of approximately 700 species in 14 families that all undergo a remarkable metamorphosis in which one eye of a symmetrical larva migrates across the top of the cranium resulting in highly asymmetrical juvenile and adult forms. Despite this distinctive morphology, pleuronectiform monophyly is poorly supported and the sister group(s) are unknown. Here, to test monophyly, intraordinal relationships and sister-group hypotheses, I optimize new molecular markers, single-exon genes (SEGs), and sequence them along with *rho* and *rnf213* for 58 flatfishes and 90 putative outgroups. I analyze those sequences along with data from a previous study to determine possible causes for gene tree incongruence or phylogenetic error. I discover that one purported SEG is most likely made up of multiple exons, but two others, *tmem22* and *chst2*, are comprised of a single exon that is variable, providing large amounts of data, while being conserved so that alignment is unambiguous. When combined with the others and analyzed simultaneously, these markers provide overwhelming support for a monophyletic Pleuronectiformes. Additionally, I demonstrate that abundant missing data is likely the cause of low resolution, validate the importance of investigating substitution saturation as a cause of error and discuss asymmetrical taxonomic distribution as a cause of low resolution at the base of Carangimorpha.

INTRODUCTION

Pleuronectiform fishes are the only vertebrates to deviate so dramatically from a bilaterally symmetrical body plan, and all known species (approximately 700 species representing 14 families) exhibit eye migration and cranial asymmetry. Because all flatfishes undergo this metamorphosis, whereas their closest relatives do not, they provide a wonderful opportunity to test hypotheses regarding the evolution and developmental mechanisms of bilateral asymmetry. Unfortunately, undertaking such a study has been elusive because of two major issues confounding pleuronectiform phylogenetics: 1) the sister group of flatfishes is unknown and 2) support for monophyly of the order has been lacking and/or underwhelming.

Alternative hypotheses, within which asymmetry has evolved multiple times (Chabanaud, 1949; Amaoka, 1969) have been suggested in the past, but were never evaluated in a phylogenetic context. Chapleau (1993) was the first to do so, but noted that because all putative characters of adult flatfish morphology were either autapomorphic or pleisiomorphic they were uninformative when testing outgroup hypotheses. Although he did not include multiple outgroups for comparison he designated the order as monophyletic based on the following three synapomorphies: 1) metamorphosis with eye migration leading to an asymmetrical cranium, 2) a dorsal fin that overlaps the neurocranium and 3) the presence of the recessus orbitalis (an accessory organ of the eye).

Uncertainty about flatfish monophyly largely stems from variation in the placement of the spiny turbots, Psettodidae (one genus, *Psettodes*), a taxon that is variably included or excluded at the base of Pleuronectiformes. Chapleau (1993)

recognized that the exact distribution of the recessus orbitalis was unknown (especially its presence or absence in *Psettodes*) and that because *Psettodes* maintained so many plesiomorphic character states, the exclusion of that genus from the order would allow for many more synapomorphies.

Resolving familial level relationships within Percomorpha, including the placement of Psettodidae, has proven to be a difficult task (Chen et al., 2003; Dettai and Lecointre, 2006; Smith and Craig, 2007; Li et al., 2009; Betancur-R et al., 2013a; Near et al., 2013). Although great advances have been made, there are still many unstable nodes, including those resolving flatfishes and their putative. These studies always recovered a para- or polyphyletic Pleuronectiformes or were unable to resolve that node at all (Chen et al., 2003; Dettai and Lecointre, 2005; Smith and Wheeler, 2006; Smith and Craig, 2007;Li et al, 2009; Betancur-R et al., 2013a; Near et al., 2013). The placement of *Psettodes* was uncertain in these hypotheses, often resolving outside flatfish, sometimes with high statistical support, but most often with low support. In most cases, however, this could easily be attributable to extremely low taxon sampling within the order and as the diversity of sampled flatfish increased in subsequent studies, the support for a Pleuronectiformes that did not include *Psettodes* decreased, but monophyly was never recovered.

The lack of resolution provided by the Euteleost Tree of Life (EToL) study (Betancur-R et al., 2013a) was particularly notable since they sampled the most flatfish (covering most family level diversity) and yet again, statistical support at the base of that part of the tree was low. Inclusion of extensive outgroup candidates allows for a more robust test of monophyly, but the inclusion of many distantly related taxa for comparison

(as in the EToL study) can lead to error in phylogeny estimation (Milinkovitch and Lyons-Weiler 1998; Tarrío et al., 2000). However while creating ambiguity as to monophyly, these studies provided the framework necessary to carry out a phylogenetic study specifically designed to test not only monophyly, but hypotheses regarding the second problem in flatfish phylogenetics: the identity of the sister group.

Betancur-R et al. (2013b) was the first to publish the results of such a study. They sampled a large number of flatfish taxa as well as including many recognized outgroups and included 20 molecular markers; three were developed for their study, but the remaining 17 had also been used in the studies mentioned above. Unfortunately, Betancur-R et al. (2013b) were only able to recover monophyly upon the removal of some of the data and were not able to add any resolution to hypotheses regarding putative sister taxa. They noted a high degree of gene-tree discordance that was probably due to systematic error, as opposed to independent lineage sorting (ILS), and indentified significant compositional bias, the vast majority of which was in third codon positions. They investigated base compositional bias as the cause of some, if not all, of the error, but only found strong associations between GC composition and gene tree discordance in some of their gene trees. Other possible sources of systematic error, however, particularly substitution saturation, were not evaluated.

Given that there is still very little resolution regarding monophyly of Pleuronectiformes and the sister group remains a mystery, I sought to identify, optimize, and sequence new molecular markers along with previously established protein coding loci for a large number of flatfish taxa and putative outgroups. Those data were then thoroughly analyzed to investigate their contributions to the phylogeny inferred from the

concatenation of all the data. Since it is so often overlooked, even when inferring deep level phylogeny, the presence and effect of substitution saturation as well as base compositional bias was investigated as a possible source of systematic error and gene tree incongruence. Similarly, because under- and over- parameterization can also cause poor estimation of phylogeny the concatenated dataset was analyzed separately using a four-partition and twelve-partition scheme and then compared for congruence using topology tests. Finally, because so many markers and taxa were included and hypotheses regarding monophyly and placement of flatfishes remained unresolved, the "EToL" dataset was reanalyzed and compared to the data generated here to investigate whether the presence of abundant missing data was a possible cause for the lack of resolution.

MATERIALS AND METHODS

2.1 TAXON SAMPLING

The sampling scheme was designed to test monophyly of Pleuronectiformes, their placement within Carangimorpha, and also intraordinal relationships. Therefore, the number of pleuronectiform taxa included was maximized as well as non-flatfish carangimorphs and other putative percomorph allies. All non-flatfish families that are part of the "L" clade of Chen et al. (2003), also referred to as Carangimorpha (Dettai and Lecointre, 2006; Li et al., 2009; Betancur-R et al., 2013a; Near et al., 2013), were included, as this was determined by Li et al. (2009) to be a "reliable" clade based on supertree analysis of their results and the results of others (Chen et al., 2003; Dettai and Lecointre, 2006). Pleuronectiformes is a member of the "L" clade, although their exact placement within it is uncertain. Johnson (1984; 1993) defined the Carangoidei as Carangidae, Echeneidae, Coryphaenidae, Rachycentridae plus the monotypic

Nematistiidae and although the latter two families were not included in Li et al.'s supertree, they are considered members of the "L" clade, or Carangimorpha (Dettai and Lecointre, 2006; Li et al., 2009; Near, 2013) and are included here. Within Carangimorpha, sampling was maximized for the species-rich Carangidae, with 27 species representing 18 genera. A breakdown of the remaining non-flatfish carangimorphs sampled for this study is as follows: two coryphaenids, two echeneids, two rachycentrids, one menid, the only nematistiid species, two centropomids, two polynemids, two, sphyraenids, one toxotid, the only xiphiid species and three istiophorids. Beyond Carangimorpha, two anabantids, one bathyclupeid, one caristiid, two channids, two emmelichthyids, six gempylids, one inermiid, two latids (two individuals identified as *Lates* sp.), one lobotid, one monodactylid, one nandid, four osphronemids, one polycentrid, two pomatomids, 21 scombrids, and five trichiurids were all included in this study because of their close affinity to the "L" clade (Chen et al., 2003; Dettai and Lecointre, 2006; Li et al., 2009; Betancur-R et al., 2013a).

Sampling within Pleuronectiformes was maximized to test monophyly of the order and phylogenetic structure within it. Both suborders, Psettoidei (consisting only of *Psettodes* spp.) and Pleuronectoidei (all non-psettoid flatfishes), were included with representatives from the following eleven families: Achiridae (four species, four genera), Bothidae (11 species, nine genera), Citharidae (two species, two genera), Cynoglossidae (two species, two genera), Paralichthyidae (12 species, nine genera), Pleuronectidae (14 species, 13 genera), Poecilopsettidae (one species), Psettodidae (one *Psettodes* sp. and one *P. erumei*), Samaridae (one species), Scopthalmidae (one species) and Soleidae (seven species, six genera). All of these families are thought to be monophyletic

(Chapleau, 1993; Berendzen and Dimmick, 2002; Betancur-R et al., 2013b) with the exception of Paralichthyidae. Chapleau (1993) recognized three distinct paralichthyid groups: 1) a *Cyclopsetta* group (including *Cyclopsettta*, *Syacium*, *Citharichthys* and *Etropus*), 2) a *Pseudorhombus* group (including *Pseudorhombus*, *Tarhops* and *Cephalopsetta*) and 3) "the remaining Paralichthyidae" (p. 528)—all of which have unknown affinity to the rest of the Pleuronectoidei. Because of this, sampling within Paralichthyidae was designed to represent all three groups and can be broken down as follows: two species of *Syacium*, one *Citharichthys*, one *Pseudorhombus*, one *Tarhops*, two *Etropus*, one *Hippoglossina*, one *Ancyclopsetta*, one *Paralichthys* and one *Xystreurys*. The taxonomic diversity sampled across Pleuronectiformes, therefore, includes 57 species (58 individuals) from suborders, 11 families and 49 genera.

For the following species data were generated from two individuals: *Coryphaena hippurus*, *Lates* sp., *Betta ibanorum*, *Pomatomus saltatrix*, *Rachycentron canadum*, and *Scopthalmus aquosus*. For all other species, one individual was included in the dataset. Most tissues were obtained from existing collections; a few were newly acquired for the purpose of this study. Tissues were collected from skeletal muscle or fin clips at the site of capture and stored in 95% ethanol. Voucher specimens are available for most samples and tissues have been retained for all. A list of all individuals included in this study along with their corresponding catalog and/or tissue numbers (corresponding to whole voucher specimens) is given in Table 3-1.

2.2 Selection of Molecular Markers

For phylogenetic analyses involving intrafamilial relationships, choice of appropriate molecular markers is crucial. Some molecular markers, such as mtDNA,

Table 3-1. Tissue and/or voucher number with isolate numbers representing all sequence data per taxon (rho, rnf213, tmem22, chst2 and the EToL data that contains 20 loci). A gray cell means data was included for that partition. Numbers in the grey EToL cells represent vouchers and in the few cases where the EToL data was generated from a different species (same genus) the name is in that cell next to its voucher number.

Order	Family	Species	Tissue Number (Voucher)	Isolate	*rho*	*rnf213*	*tmem22*	*chst2*	EToL
Pleuronectiformes	Achiridae	*Achirus lineatus*	KU 5115	67					KU 5115
		Gymnachirus melas	KU 5123	68					KU 5123
		Parachirus xenicus	KU 5719	69					
		Trinectes maculata	KU 11	66					KU 01501
	Bothidae	*Tosarhombus octoculatus*		204					
		Asterorhombus cocosensis	KU 7102	78					KU 07102
		Bothus pantherinus	KU 5642	77					E0038/*B. robinsi*
		Bothus lunatus	KU 154	70					KU 00154
		Chascanopsetta lugubris	AMNH T-010-929	79					E1181
		Chascanopsetta prognahus	AMNH T-017-1675	80					
		Crossorhombus kobensis	KU 2487	71					
		Engyophrys senta	KU 5128	76					
		Engyprosopon cf. multisquama	KU 2499	72					
		Laeops kitaharae	KU 2505	73					KU 02506
		Psettina tosana	KU 2507	74					KU 02511
	Citharidae	*Citharoides macrolepis*	KU 2466	82					KU 02468
		Samariscus japonicus	KU 2469	83					KU 02469
	Cynoglossidae	*Cynoglossus interruptus*	KU 2477	85					KU 02478
		Symphurus atricauda	KU 473	84					E0023
	Paralichthyidae	*Ancylopsetta dilecta*	KU 5129	95					
		Ancylopsetta ommata	KU 10	86					KU 00010
		Citharichthys xanthostigma	KU 450	87					E0446/*C. sordidus*
		Etropus crossotus	KU 5244	98					KU 05244
		Etropus microstomus	KU 1505	91					KU 01506
		Hippoglossina oblonga	KU 1018	89					
		Paralichthys dentatus	KU 1239	90					E0020/*P. californicus*
		Pseudorhombus pentophthalmus	KU 2479	92					KU 02481
		Syacium micrurum	KU 5200	96					KU 05200
		Syacium papillosum	KU 5095	94					
		Tarphops oligolepis	KU 2496	93					
		Xystreurys liolepis	KU 465	88					KU 00465
	Pleuronectidae	*Atheresthes evermanni*	KU 2075	1					KU 02075
		Atheresthes stomias	UW 110485	2					
		Cleisthenes herzensteini	UW 118098	62					
		Eopsetta jordani	UW 048805	12					E0444
		Glyptocephalus zachirus	UW 047669	14					E0416
		Hippoglossoides elassodon	UW 047315	31					E0424
		Hippoglossus stenolepis	UW 048793	6					E0689/*H. hippoglossus*
		Lepidopsetta bilineata	UW 048833	20					E0438
		Parophrys vetulus	UW 047297	30					E0445
		Platichthys stellatus	UW 047679	32					E0026
		Pleuronectes pinnifasciatus	UW 044943	44					
		Pleuronichthys coenosus	UW 047634	35					
		Psettichthys melanostictus	UW 047683	38					KU 00583
		Verasper moseri	UW 118096	56					
	Poecilopsettidae	*Poecilopsetta plinthus*	KU 2472	99					KU 02473
	Psettodidae	*Psettodes erumei*	AMNH T-036-3504	100					E1165
		Psettodes sp.	AMNH T-009-0819	101					E1180/*P. belcheri*
	Samaridae	*Samariscus xenicus*	KU 2482	102					
	Scopthalmidae	*Scopthalmus aquosus*	KU 1252	205					
		Scopthalmus aquosus	KU 1253	206					KU 1253
	Soleidae	*Aesopia cornuta*	AMNH T-017-1676	109					
		Aseraggodes heemstrai	KU 4996	107					
		Aseraggodes kobensis	KU 2476	104					KU 04996
		Heteromycteris japonicus	KU 2491	105					KU 2493
		Pardachirus pavoninus		203					
		Soleichthys heterorhinos	KU 7229	108					KU 07229
		Zebrias japonica	KU 2504	106					
Perciformes	Anabantidae	*Microctenopoma milleri*	AMNH T-020-1905	111					
		Microctenopoma sp.	AMNH T-035-3442	112					G1373/*M. nanum*
	Bathyclupeidae	*Bathyclupea argentea*	KU 8189	113					
	Carangidae	*Alectis ciliaris*	KU 1198	118					KU 03523
		Alectis indicus	KU 8966	136					
		Atropus atropus	KU 8967	137					
		Carangoides ferdau	KU 6972	130					KU 06972
		Carangoides plagiotaenia	KU 7077	131					KU 07154
		Caranx caninus	KU 8480	134					E0510/*C. crysos ruber*
		Caranx heberi	KU 6826	129					
		Chloroscombrus chrysurus	KU 5100	122					KU 06796
		Decapterus macarellus	KU 191	115					KU 00191
		Decapterus punctatus	KU 1170	117					KU 05322
		Elagatis bipinnulata	KU 6819	127					KU 06820
		Gnathanodon speciosus	KU 7213	132					KU 07213
		Hemicaranx amblyrhynchus	KU 5142	123					KU 05142
		Oligoplites saurus	KU 40	114					KU 00040
		Parastromateus niger	AMNH T-036-3505	138					

Order	Family	Species	Tissue Number (Voucher)	Isolate	*rho*	*rnf213*	*tmem22*	*chst2*	EToL
Perciformes (cont.)	Carangidae (cont.)	*Selar crumenophthalmus*	KU 5144	124					KU 06795
		Selene brevoortii	KU 8505	135					
		Selene browni	KU 5855	125					KU 05857
		Seriola fasciata	KU 3260	119					E0623/*S. dumerili*
		Seriola rivoliana	KU 3520	120					KU 03520
		Scomberoides commersonnianus	KU 8988	189					
		Scomberoides lysan	KU 5653	186					KU 05653
		Trachinotus baillonii	KU 6824	128					
		Trachinotus blochii	KU 6793	126					G1504/*T. carolinus*
		Trachurus symmetricus	KU 486	116					E0598/*T. lathami*
		Trachurus trachurus	KU 8074	133					
		Uraspis secunda	KU 3988	121					
	Caristiidae	*Caristius sp.*	KU 3627	139					KU 6500
	Centropomidae	*Centropomus medius*	KU 8498	141					E1158
		Centropomus undecimalis	KU 37	140					KU 00037
	Channidae	*Parachanna insignis*	AMNH T-053-5201	142					
		Parachanna sp.	AMNH T-019-1875	143					
	Coryphaenidae	*Coryphaena hippurus*	KU 5434	144					
		Coryphaena hippurus	KU 7212	145					KU 7212
	Echeneidae	*Echeneis naucrates*	KU 5140	147					KU 05140
		Remora osteochir	KU 3854	146					KU 03854
	Emmelichthyidae	*Erythrocles monodi*	KU 8371	149					
		Erythrocles schlegelii	KU 7270	148					KU 07270
	Gempylidae	*Gempylus serpens*	KU 5420	154					KU 05421
		Nealotus tripes	AMNH T-009-0830	155					KU 03112
		Neoepinnula americana	KU 3264	152					
		Nesiarchus nasutus	KU 3558	153					
		Paradiplospinus antarcticus	KU 925	151					
		Ruvettus pretiosus	KU 443	150					KU 00443
	Inermiidae	*Inermia vittata*	KU 329	156					
	Istiophoridae	*Istiophorus platypterus*	KU 5428	158					KU 05428
		Makaira nigricans	KU 5430	159					KU 05430
		Tetrapturus albidus	KU 5391	157					
	Latidae	*Lates* sp.	AMNH T-002-0176	160					E1135/*L. calcarifer*
		Lates sp.	AMNH T-011-1017	161					E1149/*L. microlepis*
	Lobotidae	*Lobotes surinamensis*	KU 8544	162					G1359
	Menidae	*Mene maculatus*	AMNH T-035-3480	163					E1131
	Monodactylidae	*Monodactylus argenteus*	KU 6758	164					KU 6758
	Nandidae	*Polycentropsis abbreviata*		202					
	Nemastiidae	*Nemastius pectoralis*	SIO 09-155	165					E1146
	Osphronemidae	*Belontia hasselti*	AMNH T-8138	200					
		Betta ibanorum	AMNH T-018-1741	166					G1226/*B. splendens*
		Betta ibanorum	AMNH T-018-1796	167					
		Trichogaster trichopterus	AMNH T-040-3927	168					
	Polycentridae	*Polycentrus schomburgkii*	AMNH T-8139	201					G1444
	Polynemidae	*Polydactylus octonemus*	KU 5116	170					KU 05116
		Polynemus sp.	KU 807	169					
	Pomatomidae	*Pomatomus saltatrix*	KU 6	171					
		Pomatomus saltatrix	KU 3992	172					KU 03992
	Rachycentridae	*Rachycentron canadum*	KU 1236	173					
		Rachycentron canadum	KU 3521	174					KU 03521
	Scombridae	*Acanthocybium solandri*	KU 5425	184					KU 07184
		Auxis rochei rochei	KU 5343	180					KU 05343
		Euthynnus affinis	KU 6772	187					KU 06772
		Euthynnus alletteratus	KU 5429	185					KU 05429
		Gymnosarda unicolor	KU 6781	188					KU 06781
		Katsuwonus pelamis	KU 5386	181					KU 05698
		Sarda sarda	KU 1121	175					KU 01121
		Scomber japonicus	KU 1567	178					KU 01568
		Scomber scombrus	KU 1482	177					KU 05183
		Scomberoides commersonnianus	KU 8988	189					
		Scomberoides lysan	KU 5653	186					
		Scomberomorus cavalla	KU 1222	176					
		Scomberomorus maculatus	KU 3930	179					KU 05193
		Thunnus atlanticus	KU 5423	183					
	Sphyraenidae	*Sphyraena argentea*	KU 493	190					KU 00493
		Sphyraena barracuda	KU 6800	191					KU 06800
	Toxotidae	*Toxotes jaculatrix*	AMNH T-8137	199					E1155
	Trichiuridae	***Aphanopus carbo***	**KU 2300**	**193**					KU 02392
		Benthodesmus simonyi	**KU 3542**	**195**					KU 03561
		Benthodesmus tenuis	KU 2952	194					
		Lepidopus altifrons	KU 3546	196					
		Trichiurus lepturus	**KU 1206**	**192**					KU 05079
	Xiphiidae	*Xiphias gladius (Makaira* sp.*)*	KU 5420	198					KU 5420

rDNA or introns, may not be suitable for deep-level fish phylogenetics, as they may have acquired too many mutations over time, overwriting phylogenetic signal and yielding an ambiguous alignment. In addition, widely divergent markers may be too difficult to amplify for all sampled taxa, leading to large amounts of missing data, a problem which most phylogenetic algorithms and models are unable to fully repair for large datasets (Wiens, 2003; Wiens, 2006; Roure et al., 2013). Because of this, only nuclear protein-coding loci were used for this study and new markers were developed to provide novel data. Single-exon genes (SEGs) described and analyzed by Tine et al. (2010) were the focus of the search for new markers. Two established nDNA protein-coding markers, *rho* (Chen, et al., 2003) and *rnf213* (Li et al., 2009) were also used for phylogenetic inference.

The utility of *rho* as a phylogenetic marker used to infer the relationships of aquatic organisms, and fishes in particular, has been questioned (Larmuseau et al, 2010; Tornabene et al., 2013; Roje, 2014), it was included because it has been shown that even markers exhibiting selection, violating assumptions of nucleotide substitutions models and possibly resulting in convergence for fishes may provide useful signal, and should not be excluded simply for that reason. In addition, whether or not *rho* exhibits a history that is incongruent with other markers is a) unknown until it is sequenced and analyzed for a given set of taxa and b) irrelevant, as the "true tree" is always unknown for a real set of data and genes are not expected to be perfectly congruent, which is why multiple loci are used to infer phylogeny (Pamilo and Nei 1988; Brower et al., 1996; Maddison, 1997; Nichols, 2001; Degnan and Rosenberg; 2006; and many others).

Target SEGs were chosen for PCR experiments if the following criteria were met: 1) they were present in both *Dicentrarchus labrax* and *Gasterosteus aculeatus* (the genomes compared in great detail by Tine et al., 2010), 2) paralogous genes were not detected in the percomorph genomes available on the Ensemble Genome Browser (http://www.ensembl.org/index.html), 3) the genes were between ~800 bp and 1.5 kb in length, which made them long enough to amplify and sequence in one fragment, maximizing the amount of data per sequencing plate and 4) when SEGs were identified in the genome assembly of other percomorphs (*Gadus morhua, Tetraodon nigroviridis, Fugu rubripes, Oryzias latipes, Xiphophorus maculatus* and *Oreochromis niloticus)* and then aligned with each other, sequences were sufficiently conserved and provided regions for suitable primers (either in the up- and downstream regions outside the genes or as close as possible to the ends).

Three genes met these criteria: *tmem22* (transmembrane protein 22, a synonym of *slc35g2b*), *chst2* (carbohydrate (N-acetylglucosamine-6-O sulfotransferase 2, a synonym of *chst2b*), and *socs6* (suppressor of cytokine signaling 6, a synonym of *chst2b*), but not all of the genomic alignments used to design primers contained the same set of reference taxa. *O. latipes chst2* is comprised of two exons (although it appears similar to a SEG in that the gene is overwhelmingly dominated by one very large exon that is separated from its small downstream exon by an very short intron) and *Fugu chst2* that contains four exons. Because of their increased length these sequences were not used to design the *chst2* primers.

2.3 DNA Extractions, Primer Design, PCR Amplification, and Sequencing

Genomic DNA was extracted from approximately 25 mg of skeletal muscle or fin tissue, using the animal tissue protocol from the Qiagen DNeasy tissue extraction kit. Aliquots of genomic DNA were used as template in all polymerase chain reactions (PCR) to amplify double-stranded DNA product from five nuclear genes. Primers were designed using Primer 3 (Untergasser et al., 2012) implemented in Geneious v5.1 (Drummond et al., 2010); a list and description of all primers used for amplification and sequencing is listed and described in Table 2. Regarding the three SEGs that met the criteria described above many primer combinations were used, but the following pairs were by far the most successful for the majority of taxa: TMEM22_F2/TMEM22_R2, CHST2_F2/CHST2_R1 and SOCS6_F1/SOCS6_R1.

Table 2. Primers used for the amplification of *rho, rnf213* and the three new SEGs (single-exon genes): *tmem22*, *chst2* and *socs6*.

Marker	Direction	Primer	Sequence (5' → 3')	Reference
rho	Forward	Rh193	CNTATGAATAYCCTCAGTACTACC	Chen et al. (2003)
	Reverse	Rh1039r	TGCTTGTTCATGCAGATGTAGA	Chen et al. (2003)
rnf213	Forward	C17 F3111	GCTGACTGGATTYAAAACCTT	Li et al. (2009)
	Reverse	C17 R4096	CCANACCAGAGGGATCATRCT	Li et al. (2009)
tmem22	Forward	TMEM22_F1	GATGKAKCTTCATCCTCMCTGC	Designed for this study
		TMEM22_F2	CTGCAGTAYGGATCMAAAAAGCG	Designed for this study
		TMEM22_F3	GCCACYCACTTCCCCCAGCCTGG	Designed for this study
	Reverse	TMEM22_R1	CATTTGAYGGGYGAGTCTAGGA	Designed for this study
		TMEM22_R2	GYGAGTCTAGGATCTCTTGATA	Designed for this study
		TMEM22_R3	CACACTGATCATGATGACCAG	Designed for this study
chst2	Forward	CHST2_F1	TGCTCAWGTCCTTGGCCT	Designed for this study
		CHST2_F2	GTGATTTGCTCAAGTCCTTGGCCTC	Designed for this study
	Reverse	CHST2_R1	GTCCTCCGCAGGAARAGGATTG	Designed for this study
socs6	Forward	SOCS6_F1	GAARATAAGCCTTAARACCATCCG	Designed for this study
		SOCS6_F2	GCCTTAARACCATCCGCAAGWC	Designed for this study
		SOCS6_F3	CCTCARCATAAAGGGCAARG	Designed for this study
	Reverse	SOCS6_R1	TCAGTAGTGYTTCTCYTGCAGAT	Designed for this study
		SOCS6_R2	GYTTCTCYTGCAGATARTCTTTC	Designed for this study
		SOCS6_R3	GRGAGCGCACTTGCATAAACC	Designed for this study

GE Healthcare illustra™ PuReTaq Ready-To-Go™ PCR Beads were used for all reactions with the addition of 1 µl each of 10 mM primer, 1-3 µl of DNA template and enough PCR grade water to reach a final volume of 25 µl. All genes were amplified in one fragment, and with the exception of *rnf213*, the following PCR profile was used: an initial denaturation step at 94°C for 2 min followed by 30-35 cycles of PCR. The cycles included denaturation at 94°C for 30 s, annealing at 48-60°C for 45 s and extension at 73°C for 60 s. A final extension step at 73°C for 10 min was carried out and then all samples were held at 4°C. Because preliminary PCR experiments (using a generic thermocycling protocol) led to the amplification of multiple fragments of varying concentration and (in many cases) similar length of the target marker, Palumbi's (1996) Touchdown Procedure was used for the amplification of all *rnf213* fragments.

PCR products were visualized under UV light on 1.5% agarose gels stained with SYBR® Safe DNA gel stain (Invitrogen Corporation, Carlsbad, CA, USA). With the exception of *rnf213*, products were cleaned with AMPure XP beads following the Agencourt protocol (Beckman Coulter Co.). Because faint bands of non-target length were also amplified using the *rnf213* primers, even after using Palumbi's (1996) Touchdown Procedure, the entire reaction volume was run on an agarose gel and the target band was excised. The agarose containing the *rnf213* fragments was then cleaned using the QIAquick Gel Extraction kit from Qiagen.

Both the forward and reverse DNA strands were sequenced separately using Big Dye 3.0 dye terminator ready reaction kits (Applied Biosystems, Foster City, CA) with the following volumes of reagents per reaction: 0.75 µl of Big Dye, 2 µl of sequencing buffer, 2 µl of 3.2 mM primer, 2 µl of cleaned PCR product, and 3.25 µL of deionized

water, for a total volume of 10 µl. Some of these concentrations are a slight modification of the STeP recipe from Platt et al. (2007). The stepped elongation time protocol (STeP) developed and described by Platt et al. (2007) was used for all sequencing reactions, which were subsequently cleaned with the CleanSEQ purification system (Agencourt Biosciences). The cleaned reactions were then run on an ABI 3730 DNA analyzer (Applied Biosystems, Rockville, MD).

2.4 Sequence Editing and Alignment

All forward and reverse chromatographs were aligned in Geneious v5.1 (Drummond et al., 2010) and checked against each other for base pair calls. Instances of heterozygosity (overlapping peaks of equal strength) were coded according to the IUPAC ambiguity code. The edited sequences were aligned using MAFFT (Katoh, 2009) implemented in Geneious v5.1 (Drummond et al., 2010) with the FT-NS-I x1000 algorithm, a gap open penalty of 2 and an offset value of 0.125.

All primer sequences were trimmed from the alignments, and because all markers were protein coding, there were a few instances where one or two ambiguous bases ("Ns") were added to the end(s) of the sequence(s) to obtain the appropriate reading frame. The reading frame was double checked against reference protein sequences from the Ensemble Genome Browser (http://www.ensembl.org/index.html) of the percomorph genomes listed above. A few instances of editing by the addition or removal of one or two internal gaps occurred only if the downstream sequence was not in frame—no gaps were added or removed subjectively and no bases were removed, added or alternatively coded. Because no internal stop codons were detected, and the sequence (and total alignment) lengths corresponded to their expected values (based on the reference SEGs),

tmem22 and *chst2* appear to be SEGs for these taxa and possibly many other percomorphs.

2.5 Phylogenetic Inference and Data Partitioning

The main goal of this study was to generate DNA sequence data and phylogenic trees that would then be compared with each other and previously published phylogenies and datasets to test flatfish monophyly and to diagnose the reasons of potential systematic error. These phylogenies and data include: individual gene trees, all unique two-gene trees and all unique three-gene trees inferred from the markers sequenced for this study (*rho*, *rnf213*, *tmem22* and *chst2*). In addition, I evaluate a tree inferred from the concatenated alignment of each gene (from here on referred to as "Con"), a tree that will be referred to as "EToL" that was inferred from Betancur-R et al.'s (2013a) matrix (data from: Dryad Repository doi:10.5061/dryad.c4d3j) that was trimmed to match the taxa mentioned above and included only 20 of their 21 markers (because of redundancy their *rho* sequences were excluded) and a phylogeny inferred from the concatenation of the all data mentioned above. From here on the trimmed matrix from Betancur-R, et al. (2013a), as well as the resulting phylogeny, will be referred to as "EToL" (referencing the Euteleost Tree of Life Project) and the concatenation of the Con and EToL alignments and the tree inferred from that matrix will be referred to as "Con-EToL".

To accommodate heterogeneity among sequence data, especially when analyzing datasets of concatenated loci, the data may be partitioned and subsets analyzed with different models. Typically, the minimum number of partitions chosen is one per gene because it is assumed that each gene has a unique evolutionary rate and pattern. Alternatively, a scheme in which partitions correspond to each codon position in protein-

coding genes may be used, because it not only accounts for differences between genes, but also for the theoretical differences in mutation rates at each codon position. Increasing the number of partitions will increase the number of models used, and parameters estimated. This in turn will increase the fit to the data and consequently the likelihood score, but will not necessarily yield the best estimates of branch length and toplogy (Sullivan and Joyce, 2005). Unfortunately, the estimation of too many parameters may lead to overparameterization, yielding unidentifiable parameters that have effects on the data, but cannot be estimated accurately (Rannala, 2002). Li et al., 2008 described a system of optimal data partitioning utilizing the AIC (Akaike Information Criterion), but this method allows data to be removed from the analysis, which defeats the purpose of using multiple unlinked loci to overcome phylogenetic error.

To address the effect of increased partitioning of data, two partitioning schemes including “by gene” and “codon position per gene” were used to infer two Con trees for comparison. Hypothesis testing (described below) was then carried out on those trees to determine if they were equally good explanations of those data given the two partitioning schemes, models chosen and number of parameters estimated. If the trees were deemed congruent, then the one inferred with fewer parameters estimated would be considered best. Otherwise, the tree with the highest (closest to zero) likelihood score (lnL) would be chosen.

The individual gene trees (and the two- and three-gene trees) were analyzed according to the partitioning scheme chosen as best for the Con tree. They were not subjected to an extensive search for optimal partitioning scheme because these trees were

only used to investigate congruence with the Con tree (and others) as well as to check whether they exhibited any bias that could lead to phylogenetic error. The Con dataset, as opposed to the EToL or Con-EToL datasets, was considered the best possible to infer branch length and topology as it included all four markers sequenced and with superior coverage across taxa.

Because the EToL alignment contained 20 markers, computational time limited the ability to select models by gene or codon position (if applicable). Therefore, the EToL matrix was treated as one partition and a model was chosen using the AICc (corrected Akaike Information Criterion) in Treefinder (Jobb, 2008). Nucleotide substitution models were also selected for each gene partition (*rho, rnf213, tmem22, chst2*) or codon position per gene (for the four sequenced genes, not the EToL data) using the AICc carried out in Treefinder (Jobb, 2008).

All phylogenies were inferred utilizing the ML criterion with 1000 bootstrap replicates implemented in Treefinder (Jobb, 2008). Because outgroup choice should include taxa closely related to the ingroup (Milinkovitch and Lyons-Weiler 1998; Tarrío et al., 2000), and it is still unclear which taxon meets that criterion given the 148 taxa sampled here, all trees were left unrooted. This choice, to use unrooted trees, should not affect the ability to resolve a sister group to flatfishes or test monophyly of the order, as the inclusion of the 90 non-flatfish taxa was designed to provide proper outgroup comparison without forcing a root on to the tree.

2.6 Hypothesis Testing

To determine if topology differences were statistically significant, the following three paired sites tests were carried out on all individual genes as well as the Con and

EToL data and their corresponding ML topologies: the KH (Kishina and Hasegawa, 1989), the AU (Shimodaira, 2002) and the SH (Shimodaira and Hasegawa, 1999) tests. These tests evaluate the difference in likelihood between topologies to the empirical variation in log likelihoods for a given dataset. The hypothesis tests were implemented in Treefinder (Jobb, 2008) using the RELL (resample estimated log-Likelihood) nonparametric bootstrap method with 50,000 replicates used to generate the null distributions for the KH and SH tests. For the AU test, a multiscale bootstrap technique, used to incur the least amount of bias, was used and is less conservative than the SH and KH tests (Shimodaira, 2002; Jobb, 2008). The models used to calculate the likelihoods for each topology were the same as those chosen using the AICc in Treefinder (Jobb, 2008). The null hypothesis (H_0: all topologies for comparison are equally good explanations of the data) was rejected when $P < 0.05$.

2.7 Potential Sources of Bias: Base Composition and Substitution Saturation

Betancur-R et al. (2013b) proposed that base compositional bias was likely the cause of at least some of their gene tree discordance and could explain the inability of their concatenated dataset (and possibly others) to resolve a monophyletic Pleuronectiformes with high statistical support. To see if their results could be replicated with different data, compositional bias of each OTU was calculated over its sequence and then treated as a continuous character (the proportion of GC). This character was then mapped on the corresponding gene tree and the Con tree, allowing for reconstruction of the evolution of base composition for those taxa. These analyses were performed in Mesquite 2.75 using the ACGT Compositional Bias module (Maddison and Maddison, 2011).

Another potential source of bias and systematic error that may lead to gene tree discordance is nucleotide substitution saturation, which is identified when the amount of measureable mutation plateaus as pairwise distance increases, presumably due to multiple hits at sites, and implying the loss of phylogenetic information. Even though substitution models correct for this they are not always able to do so adequately, especially for sequences that are known to have exceptionally high mutation rates or in instances when short sequences are used to infer the relationships of many OTUs (Xia et al., 2003).

To examine the degree of nucleotide substitution saturation at third codon positions of each of the four genes, transitions and transversions were plotted against corrected genetic distances in DAMBE (Xia and Xie, 2001) for all pairwise comparisons among taxa. Only the third position was examined because changes at that position almost always correspond to synonymous mutations that, theoretically, occur at much faster rates than nonsynonymous ones (saturation is expected for third positions and not for first and second). The model F84 was used to calculate corrected genetic distances because it is most similar to HKY, the model chosen for all third positions using the AICc (see in Results section below), and HKY is not available in DAMBE.

Because transitions are expected to reach saturation before transversions, they were used to obtain the most conservative threshold pairwise genetic distance (TPWD), or that point on the curve where the data begins to act saturated (non-linear). Those nodes where the pairwise distance between two OTUs is greater than the TPWD (where the curve begins to plateau) can be thought of as being inferred from saturated data (Roje, 2010). This is opposed to those taxa that have a low genetic distance (below the TPWD) separating them—that node can be considered as being inferred by unsaturated data.

The method used here for determining this point on the curve is identical to that described by Roje (2010) and can be summarized as follows: the TPWD is the level 1 change-point, with confidence greater than 95%, in the variation of transitions over genetic distance. Since saturation occurs when the curve begins to plateau, the change in variation of transitions was chosen over the change in values. In addition, the level 1 change-point was used as the TPWD because it is the first change detected and the most visibly apparent (Taylor, 2000a). All change-point analyses were carried out using change-point analyzer 2.3 (Taylor, 2000b).

RESULTS

3.1 SINGLE EXON GENES AS PHYLOGENIC MARKERS

Three of the SEG's described by Tine et al. (2010) met the criteria listed above (*tmem22*, *chst2* and *socs6*). In the case of *socs6*, however, although target fragments were amplified and sequenced for 89% of the taxa, the alignment was not straightforward and resulted in a large number of internal gaps that did not correspond to codons. The primers used were internal, so even with the gaps the alignment length was similar in length to the reference sequences. All 131 *socs6* sequences were submitted to BLAST and the scores for *Dicentrarchus labrax* were high, but the best scores were with mRNA transcripts of more closely related taxa. Based on this and a qualitative assessment of the "bird's eye view" of the entire alignment it seems there is at least one intron in the center of the gene separating what appears to be two exons (the more conserved regions); the fasta file of the *socs6* alignment is available with the others, but in this case individual sequences are only identifiable by their isolate number.

The conclusion that *socs6* for these taxa is not a SEG is further corroborated by

the recent results of Chen et al., (2014) who published the genome of *Cynoglossus semilaevis,* a flatfish. Their assembly shows *socs6* contains multiple exons and introns (NCBI Reference Sequence: NC_024309.1), with the gene being approximately 3.6 times longer than the SEGs identified in the genomes of *Dicentrarchus labrax, Gasterosteus aculeatus, Gadus morhua, Tetraodon nigroviridis, Fugu rubripes, Oryxias latipes, Xiphophorus maculatus* and *Oreochromis niloticus*. The *socs6* sequences for the taxa sampled here are most likely not SEG's and because their true structure is unknown the assumption that sites in that alignment are homologous is a poor one. Therefore, *socs6* was excluded from all downstream analyses and the final dataset consisted of the other two SEGs (*tmem22* and *chst2*) plus *rho* and *rnf213*; the number of sequences (per taxon and per marker) generated are summarized in Table 3-1.

The *tmem22* alignment was 1.062 kb in length, had 489 phylogenetically informative sites (46.0% of total sites), a pairwise percent identity of 92.0%, GC content of 55.0%, and a total of 297 gaps (all of which were sequential, corresponding to at least one codon) in the alignment. The *chst2* alignment was similar in most *tmem22* statistics in that it was 1.039 kb long, had 556 phylogenetically informative sites (53.5% of the total), a pairwise percent identity of 91.7% and a GC content of 52.8%. It differed most from the *tmem22* alignment in the number of gaps: 6 versus 297. Like *tmem22*, however, all the gaps in the *chst2* alignments corresponded to codons. Since no indels were present that lead to internal stop codons, all sequences were approximately the same length and the translated amino acid sequences seem conserved; both *tmem22* and *chst2* appear to be SEGs for all the taxa sequenced.

The total length of the exonal *rho* fragment alignment was 850 bp with 427 informative sites (41.1%) and no gaps. It had a pairwise percent identity of 88.7% and the GC content was 53.3%. The alignment of the exonal *rnf213* fragment was similar in length (830 bp) to *rho*, but differed considerably in all other statistics: 547 sites were phylogenetically informative (65.9% of the total), the pairwise percent identity was 87.6%, there were 40 gaps (all representing codons) and the GC content was only 45%.

3.2 The ETOL Alignment

The original matrix generated by Betancur-R et al. (2013a) was trimmed to maximize overlap, and minimize the amount of chimeric sequences. Once trimmed, the final dimensions of the EToL alignment were 94 sequences by 20001 sites. Of the 94 sequences in the EToL alignment, the vast majority was generated from the same individual sampled for this study; either they shared the same KU tissue number or whole specimen voucher (different KU tissue, but from the same individual). In 13 cases, however, another species from the same genus was used in lieu of the species sampled here therefore, the Con-EToL phylogeny was generated using chimeric sequences; this is summarized in Table 1.

Betancur-R et al. (2013a) noted the large amount of missing data in their concatenated alignment of 1414 taxa approximately 21 kb in length, and this was also apparent in the trimmed alignment, but was exaggerated. The number of total gaps (including free-end) in the alignment was 1.141344 million or 60.7% of the total number of sites. For the most part an EToL marker had less than 50% coverage in the trimmed alignment and in one case that number was as low as 11.7%. This is in contrast to the Con alignment that consists of 148 sequences and 3781 sites, much less than EToL, but

with superior coverage per gene: 96% for *rho*, 85% for *rnf213*, 75% for *tmem22* and 78% for *chst2*. All alignments, including *socs6*, are available as fasta files in the American Museum of Natural History Dissertation Database.

3.2 Phylogenetic analyses and Data Partitioning

Eighteen unrooted phylogenies were generated from the data matrices described above: four gene trees (*rho*, *rnf213*, *tmem22* and *chst2*), six two-gene trees, four three-gene trees, two Con trees, the EToL tree and the concatenation of the all data, the Con-EToL tree. The same model (in Treefinder notation: GTR[Optimum,Empirical]:G[Optimum]:5) was chosen as the best fit for all single gene partitions except *rho*, for which HKY[{3,1,1,1,1,3},Empirical]:G[Optimum]:5 was chosen. For the 12 (codon position per gene) partitions HKY[{3,1,1,1,1,3},Empirical]:G[Optimum]:5 was chosen for 10 of them with the two exceptions being HKY[Optimum,Empirical]:G[Optimum]:5 that was chosen for the second codon positions of both *tmem22* and *chst2*.

The Con tree inferred using the four-partition, by gene, analysis had an lnL of -72426.6 with 327 parameters estimated. As expected, the phylogeny obtained using 12 partitions had a greater likelihood score and a larger number of parameters estimated: lnL = -71142.69 and 354, respectively. Both ML phylogenies, however, recovered a monophyletic Pleuronectiformes with Psettoidei sister to Pleuronectoidei and both had high bootstrap proportions at that node (92 for the four-partition tree and 81 for the 12-partition tree). Only the 12-partition ML phylogeny was able to recover a sister group to the flatfishes, Nandidae, but this was with very low support (bootstrap of 47). Both topologies, with branch lengths, are shown in Figures 3-1 and 3-2.

Figure 3-1. Comparison of alternative data partitioning on the Con phylogeny: Four partitions. Note a monophyletic Pleuronectiformes in blue with bootstrap support at each node. The circled taxon is *Cynoglossus interruptus* and the arrow is pointing to *Poecilopsetta plinthus*, both recovered with long branches.

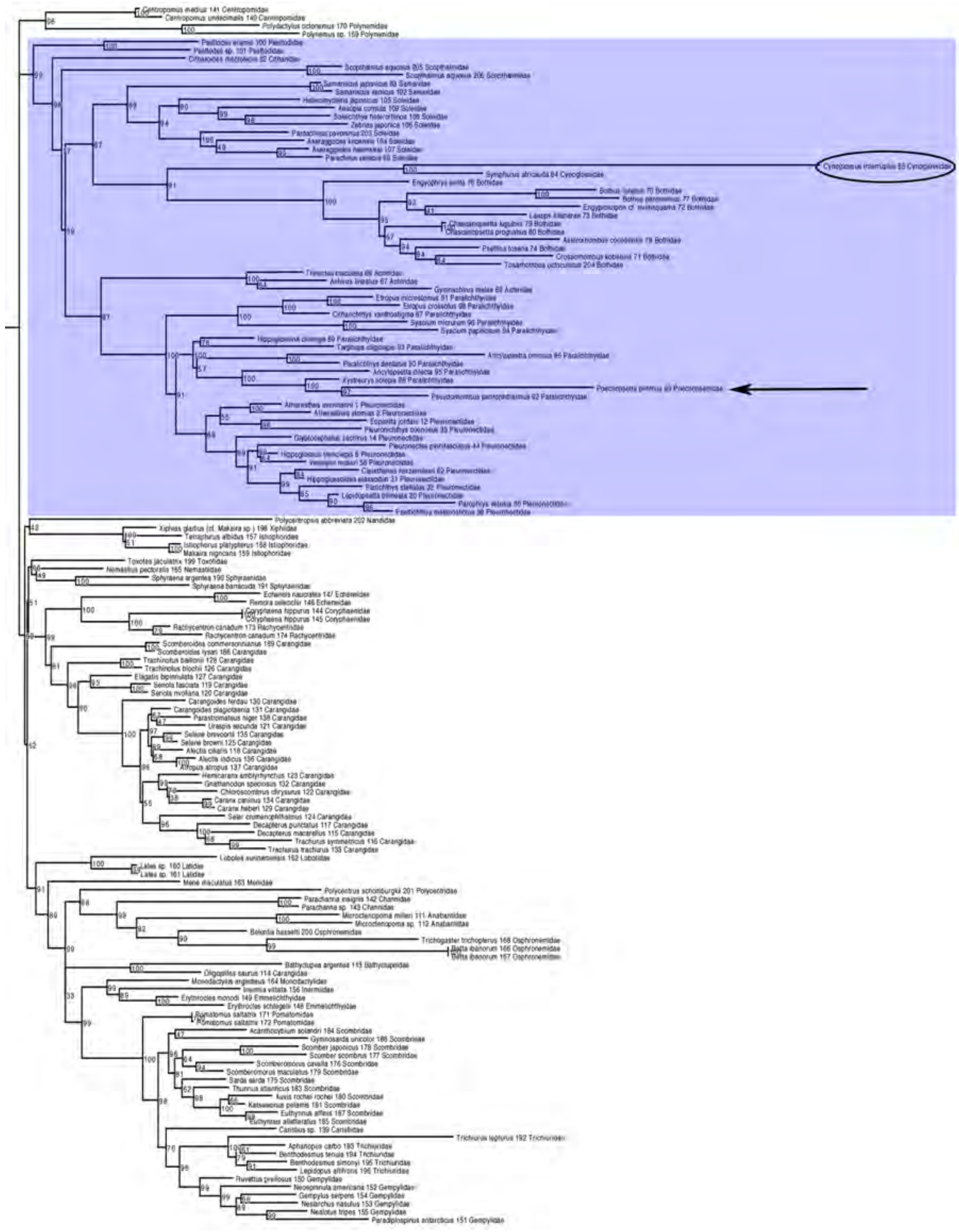
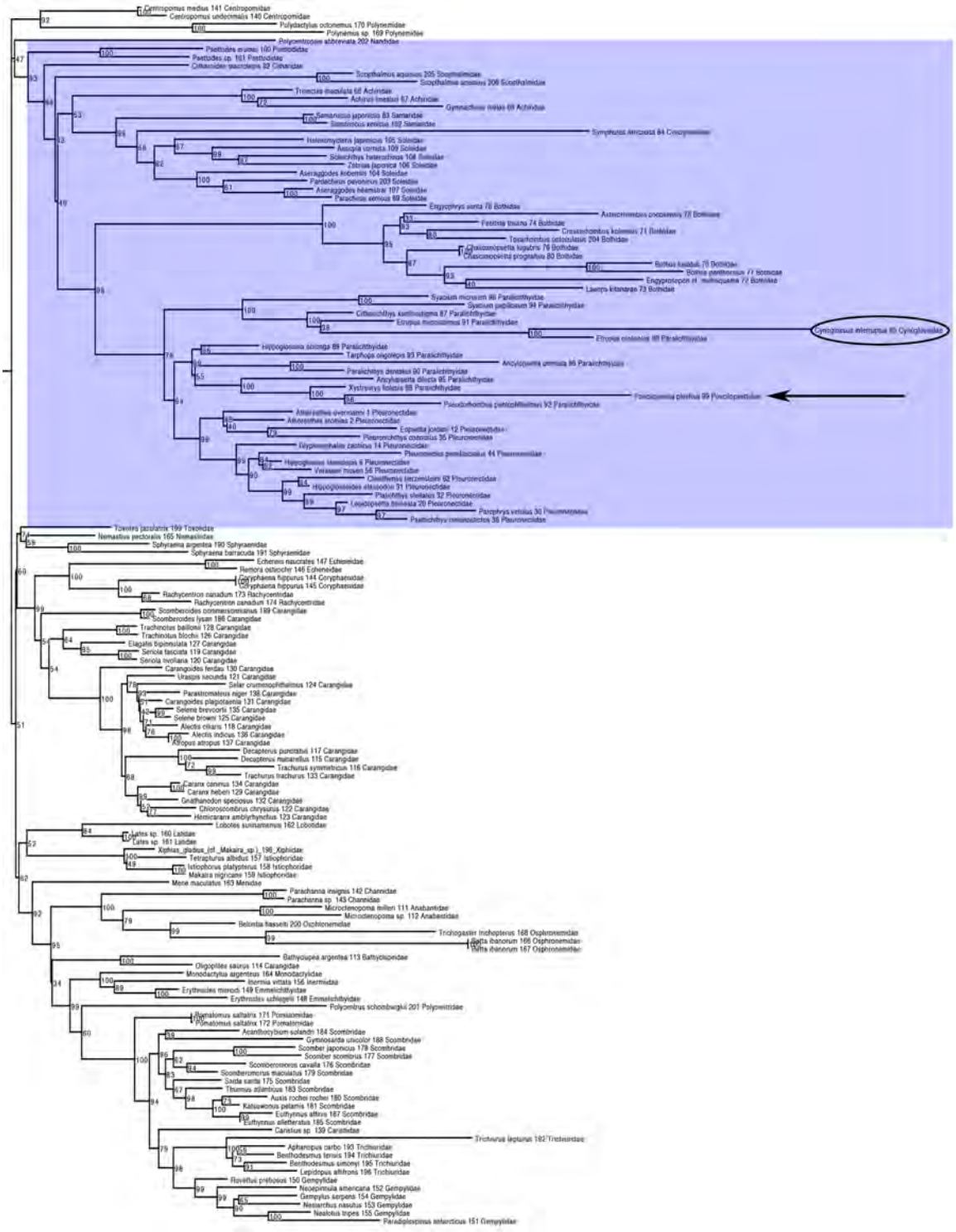

Centropomus medius 141 Centropomidae
Centropomus undecimalis 140 Centropomidae
Polydactylus octonemus 170 Polynemidae
Polynemus sp. 159 Polynemidae
Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Citharoides macrolepis 82 Citharidae
Scophthalmus aquosus 205 Scophthalmidae
Scophthalmus aquosus 206 Scophthalmidae
Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes haemskai 107 Soleidae
Parachirus xenicus 69 Soleidae
Cynoglossus interruptus 85 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Bothus lunatus 70 Bothidae
Bothus pantherinus 77 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta lugubris 79 Bothidae
Chascanopsetta prognathus 80 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Trinectes maculatus 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 68 Achiridae
Etropus microstomus 91 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 8 Pleuronectidae
Verasper moseri 58 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Platichthys stellatus 22 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 36 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
Polycentropsis abbreviata 202 Nandidae
Xiphias gladius (cf. Makaira sp.) 198 Xiphiidae
Tetrapturus albidus 157 Istiophoridae
Istiophorus platypterus 158 Istiophoridae
Makaira nigricans 159 Istiophoridae
Toxotes jaculatrix 199 Toxotidae
Nematistius pectoralis 165 Nematistiidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Elagatis bipinnulata 127 Carangidae
Seriola fasciata 119 Carangidae
Seriola rivoliana 120 Carangidae
Carangoides ferdau 130 Carangidae
Carangoides plagiotaenia 131 Carangidae
Parastromateus niger 138 Carangidae
Uraspis secunda 121 Carangidae
Selene brevoortii 135 Carangidae
Selene browni 125 Carangidae
Alectis ciliaris 118 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Gnathanodon speciosus 132 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Caranx caninus 134 Carangidae
Caranx heberi 129 Carangidae
Selar crumenophthalmus 124 Carangidae
Decapterus punctatus 117 Carangidae
Decapterus macarellus 115 Carangidae
Trachurus symmetricus 116 Carangidae
Trachurus trachurus 133 Carangidae
Lobotes surinamensis 162 Lobotidae
Lates sp. 160 Latidae
Lates sp. 161 Latidae
Mene maculatus 163 Menidae
Polycentrus schomburgkii 201 Polycentridae
Parachanna insignis 142 Channidae
Parachanna sp. 143 Channidae
Microctenopoma milleri 111 Anabantidae
Microctenopoma sp. 112 Anabantidae
Belontia hasselti 200 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Betta ibanorum 166 Osphronemidae
Betta ibanorum 167 Osphronemidae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Monodactylus argenteus 164 Monodactylidae
Inermia vittata 156 Inermiidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Acanthocybium solandri 184 Scombridae
Gymnosarda unicolor 186 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Caristius sp. 139 Caristiidae
Trichiurus lepturus 192 Trichiuridae
Aphanopus carbo 193 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Ruvettus pretiosus 150 Gempylidae
Neoepinnula americana 152 Gempylidae
Gempylus serpens 154 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae

Figure 3-2. Comparison of alternative data partitioning on the Con phylogeny: Twelve partitions. Note a monophyletic Pleuronectiformes in blue with bootstrap support at each node. The circled taxon is *Cynoglossus interruptus* and the arrow is pointing to *Poecilopsetta plinthus*, both recovered with long branches.


Centropomus medius 141 Centropomidae
Centropomus undecimalis 140 Centropomidae
Polydactylus octonemus 170 Polynemidae
Polynemus sp. 169 Polynemidae
Polycentrosoma abbreviata 202 Nandidae
Psettodes erumei 160 Psettodidae
Psettodes sp. 161 Psettodidae
Citharoides macrolepis 92 Citharidae
Scophthalmus aquosus 205 Scophthalmidae
Scophthalmus aquosus 206 Scophthalmidae
Trinectes maculata 68 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 69 Achiridae
Samaris cristatus 83 Samaridae
Samaris cristatus 102 Samaridae
Symphurus tricauda 84 Cynoglossidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Solea ovata 106 Soleidae
Solea senegalensis 104 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 107 Soleidae
Aseraggodes klemsdali 108 Soleidae
Paraichirus elimus 89 Soleidae
Engyophrys senta 78 Bothidae
Asterorhombus intermedius 77 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octosulcatus 204 Bothidae
Chascanopsetta lugubris 76 Bothidae
Chascanopsetta prognatha 80 Bothidae
Bothus lunatus 78 Bothidae
Bothus pantherinus 77 Bothidae
Engyprosopon m. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Syacium micrurum 86 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Cynoglossus interruptus 85 Cynoglossidae
Etropus crossotus 88 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 96 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 98 Paralichthyidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Pleuronichthys ocellatus 35 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Clidoderma asperrimum 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Psammichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 26 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 36 Pleuronectidae
Toxotes jaculatrix 199 Toxotidae
Nematistius pectoralis 165 Nematistiidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Scomberoides commersonianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Elagatis bipinnulata 127 Carangidae
Seriola fasciata 119 Carangidae
Seriola lalandi 120 Carangidae
Carangoides ferdau 130 Carangidae
Uraspis secunda 121 Carangidae
Selar crumenophthalmus 124 Carangidae
Parastromateus niger 136 Carangidae
Carangoides plagiotaenia 131 Carangidae
Selene brevoorti 135 Carangidae
Selene brownii 125 Carangidae
Alectis ciliaris 118 Carangidae
Alectis indicus 138 Carangidae
Atropus atropus 137 Carangidae
Decapterus punctatus 117 Carangidae
Decapterus macarellus 115 Carangidae
Trachurus symmetricus 116 Carangidae
Trachurus trachurus 133 Carangidae
Caranx caninus 134 Carangidae
Caranx heberi 129 Carangidae
Gnathanodon speciosus 132 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Lobotes surinamensis 162 Lobotidae
Lates sp. 160 Latidae
Lates sp. 161 Latidae
Xiphias_gladius_(of_Makaira_sp.)_196_Xiphiidae
Tetrapturus albidus 157 Istiophoridae
Istiophorus platypterus 158 Istiophoridae
Makaira nigricans 159 Istiophoridae
Mene maculata 163 Menidae
Parachanna insignis 142 Channidae
Parachanna sp. 143 Channidae
Microctenopoma belleri 111 Anabantidae
Microctenopoma sp. 112 Anabantidae
Belonidae fraseri 200 Ophichthonemidae
Trichogaster trichopterus 168 Osphronemidae
Betta ibanorum 166 Osphronemidae
Betta ibanorum 167 Osphronemidae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Monodactylus argenteus 164 Monodactylidae
Ibernia vittata 156 Iberniidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Polycentrus schomburgkii 201 Polycentridae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Acanthocybium solandri 184 Scombridae
Gymnosarda unicolor 188 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Caristius sp. 139 Caristiidae
Trichiurus lepturus 182 Trichiuridae
Aphanopus carbo 193 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Ruvettus pretiosus 150 Gempylidae
Neoepinnula americana 152 Gempylidae
Gempylus serpens 154 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae

Although both partitioning schemes recovered *Cynoglossus interruptus* (circled in Figs. 3-1 and 3-2) with an extremely long branch relative to the overall length of the trees, the possible phylogenetic error this can cause (Felsenstein, 1978; Bergstrom, 2005) has been annulled by the four-partitioned scheme, as *C. interruptus* was recovered with the other cynoglossid, *Symphurus atricauda* with a bootsrap value of 100. This is opposed to the phylogeny inferred from the12-partitioned dataset where a paralichthyid, *Etropus crossotus,* is recovered as sister to the *C. interruptus* with 100% support instead of with its congener (also, note that *S. atricauda* exhibits a very long branch in the latter case).

In addition, all three of the hypothesis tests carried out on the two Con trees and their respective partitioning schemes failed to reject ($P > 0.05$) the null hypothesis that all topologies for comparison are equally good explanations of the data. When comparing the four-partition data to the 12-partition topology, P-values of 0.317 for the KH and SH and 0.313 for the AU test were generated and when comparing the 12-partition data to the four-partition topology the following P-values were obtained: 0.403 for the KH and SH tests and 0.407 for the AU test. The four-partition tree is nearly identical to the other Con tree, cannot be rejected as an equally good explanation of the more highly parameterized data, seems to be less sensitive to long-branch attraction (possibly caused by potential overparameterization), and so was deemed the best estimate of phylogeny and was used for all downstream analyses and discussions.

3.3 Monophyly and Phylogenetic Placement of Pleuronectiformes

The ML phylogeny of the partitioned by gene Con dataset recovered a monophyletic Pleuronectiformes with high bootstrap proportion of 92 (a fully annotated

version is given in Figs. 3-3 and 3-4). This is a novel result, as it is the first time monophyly has been recovered without the removal of data (Betancur et al., 2013b) and with such high statistical support at that node. Whether alternative hypotheses are significantly different than the ML Con (or any other) tree will be addressed below.

The deeper nodes of this Con tree were recovered with low support (Fig. 3-5), not surprising considering the very short branches at the base (Fig. 3-1). Because of this, it is not possible to say these data clearly support any one sister group to flatfishes over the other. Regardless, the unrooted Con tree is comprised of three main clades: flatfishes, a highly supported Centropomidae + Polynemidae clade and the remaining taxa with a bootstrap proportion of 69. The high support for the Centropomidae + Polynemidae clade is significant, as their placement within Carangimorphariae (*sensu* Betancur-R et al., 2013a) has been unresolved and sensitive to alternative datasets and analyses. Within Carangiomorphariae, support for a monophyletic Istiophoriformes was high and it was recovered as sister to the only nandid sampled, *Polycentropis abbreviata*, but with only 42% bootstrap support.

Polycentropis abbreviata is acting like a "rogue" taxon in these analyses, displaying an unstable position in the different trees. This is not surprising considering the low taxon sampling within Nandidae, however, extant nandid diversity includes only 10 species in 3 genera. Low extant diversities are also characteristic of Psettodidae (3 species, one genus), Toxotidae (10 species, one genus), and Nemastiidae (one species) all taxa whose phylogenetic position has been difficult to pinpoint in all the recent molecular phylogenies aimed at testing acanthomorph relationships (Chen et al., 2003; Smith and Wheeler, 2006; Li et al., 2009; Betancur-R et al., 2013a; Near et al., 2013) and, except for

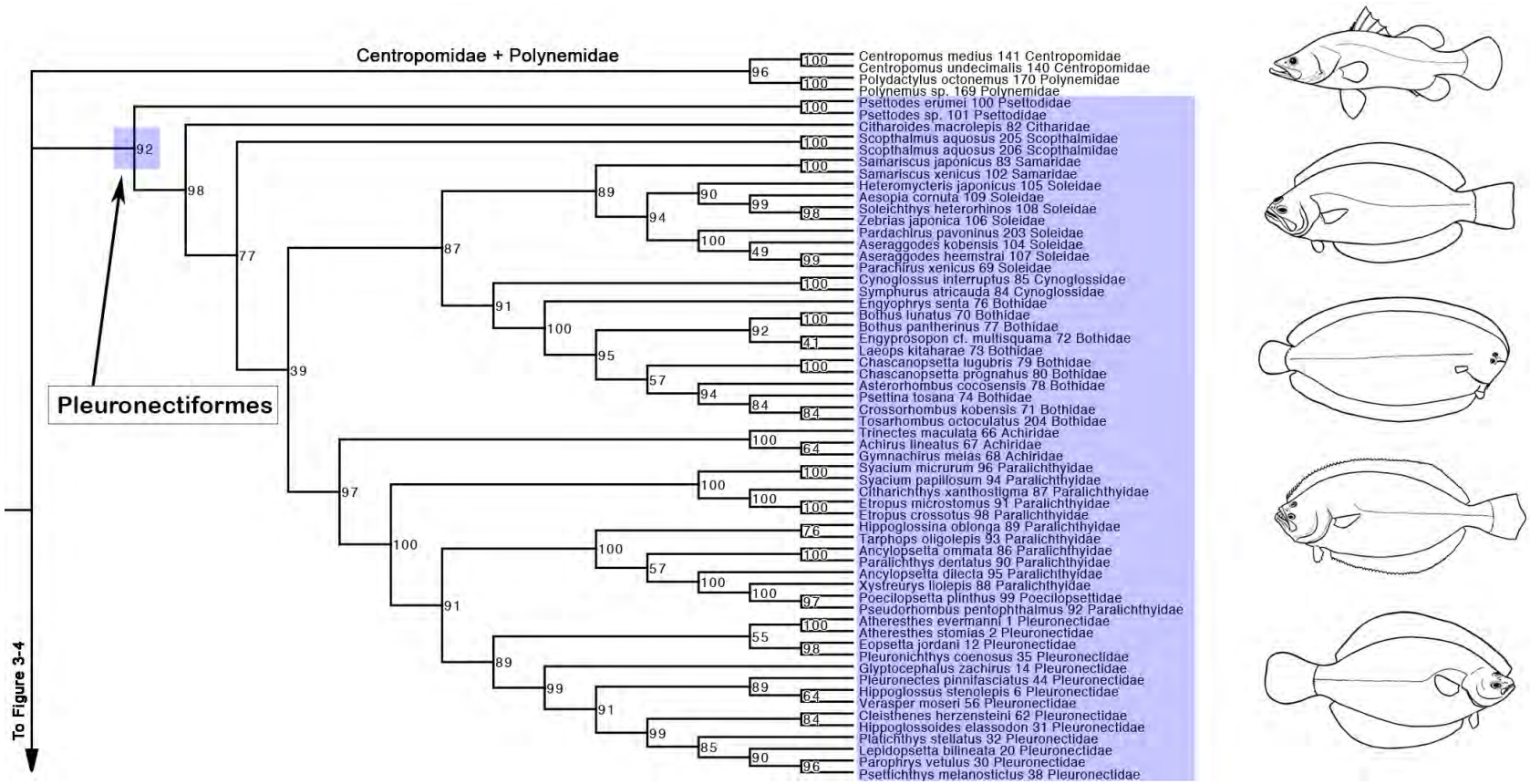
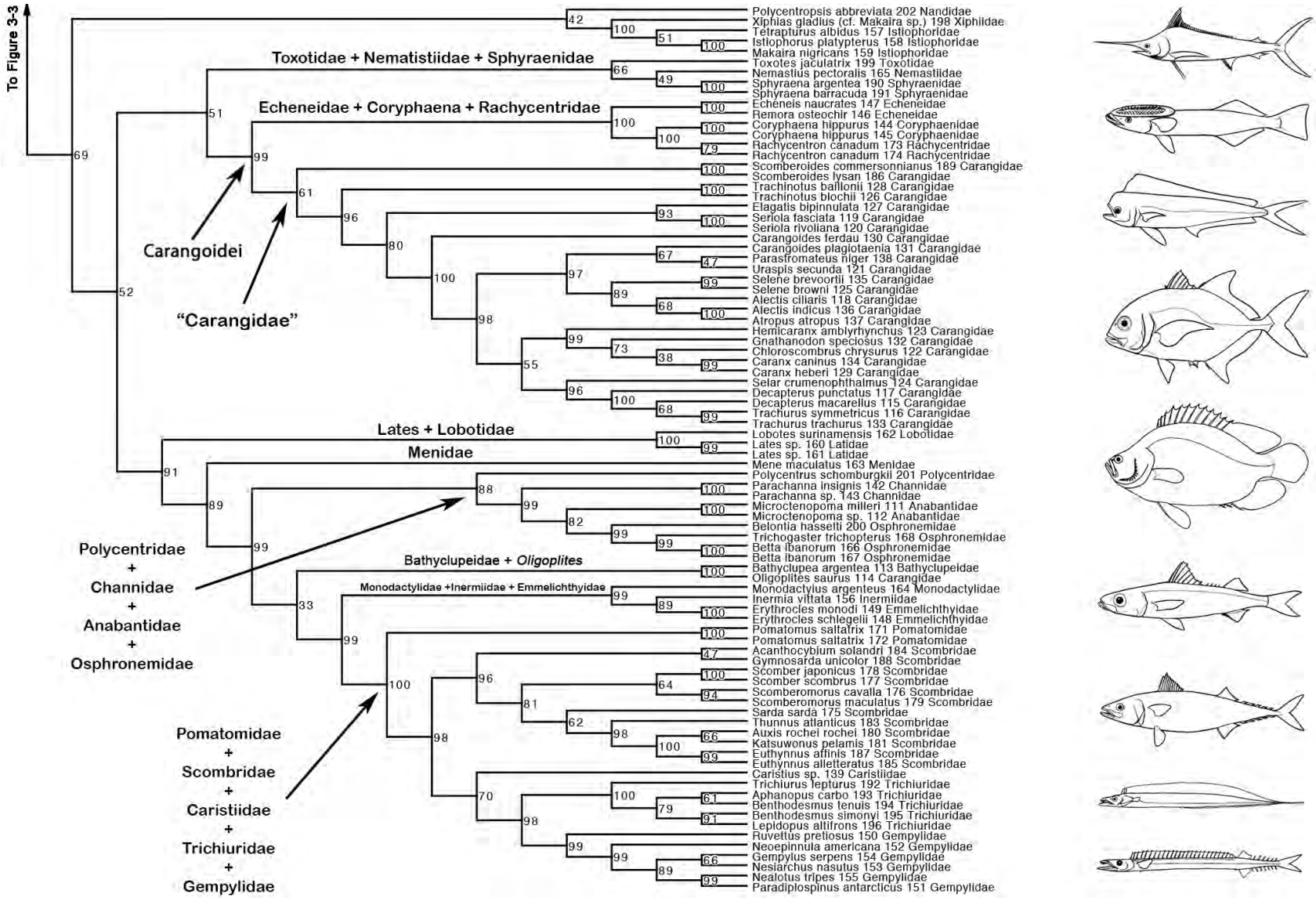


Figure 3-3. Part 1 of Maximum likelihood tree, with bootstrap support at nodes, estimated with Treefinder of the Con (rho, rnf213, tmem22 and chst2) dataset that was partitioned by gene. Blue represents a monophyletic Pleuronectiformes. All illustrations are reproduced from Nelson (2006).

Figure 3-4. Part 2 of Maximum likelihood tree, with bootstrap support at nodes, estimated with Treefinder of the Con (*rho*, *rnf213*, *tmem22* and *chst2*) dataset that was partitioned by gene. All illustrations are reproduced from Nelson (2006).


To Figure 3-3
Toxotidae + Nematistiidae + Sphyraenidae
Echeneidae + Coryphaena + Rachycentridae
Carangoidei
"Carangidae"
Lates + Lobotidae
Menidae
Polycentridae + Channidae + Anabantidae + Osphronemidae
Bathyclupeidae + Oligoplites
Monodactylidae +Inermiidae + Emmelichthyidae
Pomatomidae + Scombridae + Caristiidae + Trichiuridae + Gempylidae
Polycentropsis abbreviata 202 Nandidae
Xiphias gladius (cf. Makaira sp.) 198 Xiphiidae
Tetrapturus albidus 157 Istiophoridae
Istiophorus platypterus 158 Istiophoridae
Makaira nigricans 159 Istiophoridae
Toxotes jaculatrix 199 Toxotidae
Nemastius pectoralis 165 Nemastiidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Elagatis bipinnulata 127 Carangidae
Seriola fasciata 119 Carangidae
Seriola rivoliana 120 Carangidae
Carangoides ferdau 130 Carangidae
Carangoides plagiotaenia 131 Carangidae
Parastromateus niger 138 Carangidae
Uraspis secunda 121 Carangidae
Selene brevoortii 135 Carangidae
Selene browni 125 Carangidae
Alectis ciliaris 118 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Gnathanodon speciosus 132 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Caranx caninus 134 Carangidae
Caranx heberi 129 Carangidae
Selar crumenophthalmus 124 Carangidae
Decapterus punctatus 117 Carangidae
Decapterus macarellus 115 Carangidae
Trachurus symmetricus 116 Carangidae
Trachurus trachurus 133 Carangidae
Lobotes surinamensis 162 Lobotidae
Lates sp. 160 Latidae
Lates sp. 161 Latidae
Mene maculatus 163 Menidae
Polycentrus schomburgkii 201 Polycentridae
Parachanna insignis 142 Channidae
Parachanna sp. 143 Channidae
Microctenopoma milleri 111 Anabantidae
Microctenopoma sp. 112 Anabantidae
Belontia hasselti 200 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Betta ibanorum 166 Osphronemidae
Betta ibanorum 167 Osphronemidae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Monodactylus argenteus 164 Monodactylidae
Inermia vittata 156 Inermiidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Acanthocybium solandri 184 Scombridae
Gymnosarda unicolor 188 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Caristius sp. 139 Caristiidae
Trichiurus lepturus 192 Trichiuridae
Aphanopus carbo 193 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Ruvettus pretiosus 150 Gempylidae
Neoepinnula americana 152 Gempylidae
Gempylus serpens 154 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae

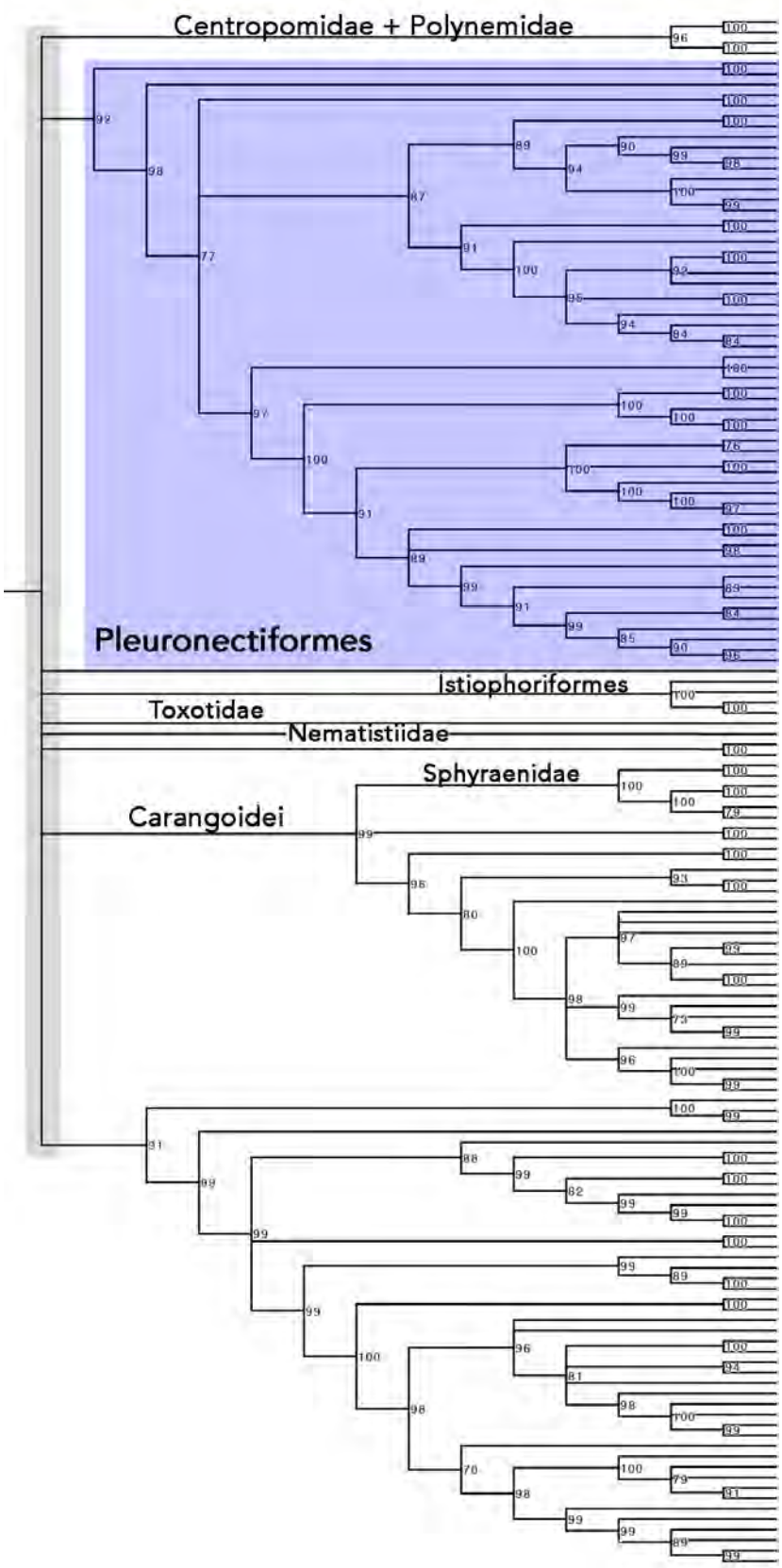
Psettodidae, here as well. In the case of Nandidae, when sampling within the family is increased (Near et al., 2013) its position does stabilize (although, in that case the support is low for the monophyly of its sister group, Anabantiformes).

The placement of the monotypic Nematistiidae, however, remains uncertain, with increases in sampling impossible due to the low diversity. The ML Con tree recovered a Nematistiidae + Sphyraenidae + Toxotidae clade, but with only 66% bootstrap support, and with even lower support at the node uniting that clade to a very well supported Carangoidei. When all nodes in the Con tree with a bootstrap support below 70% are collapsed (Fig. 3-5) it becomes clear that these data, in addition to the datasets analyzed by many others, are unable to resolve the placement of Carangoidei, Toxotidae, Nematistidae, Sphyraenidae, and Istiophoriformes within Carangimorpha, making their phylogenetic placement an important key in determining the sister group to Pleuronectiformes.

A well supported carangoid clade comprised of Echeneidae + Coryphaenidae + Rachycentridae (100% bootstrap) sister to a poorly supported (61%) “Carangidae”, that did not include *Oligoplites saurus,* was recovered. *Oligoplites saurus* is variously placed in the hypotheses, sister to Bathyclupeidae in the Con tree, and with other carangids in the *rho* tree, but not the *rnf213* or *tmem22* trees (sequence for *chst2* was not generated for *O. saurus*).

A clade made up of all the remaining taxa sampled was recovered with high statistical support and other than the node uniting Bathyclupeidae and *Oligoplites* to Scombriformes, it is highly structured. Within this clade is a highly supported Latidae + Lobitidae, then a menid lineage, followed by a large group that has Anabantiformes

Figure 3-5. Con tree with all nodes with a bootstrap proportion < 70 are collapsed. Blue represents Pleuronectiformes. Note comb (in gray) at base of tree.

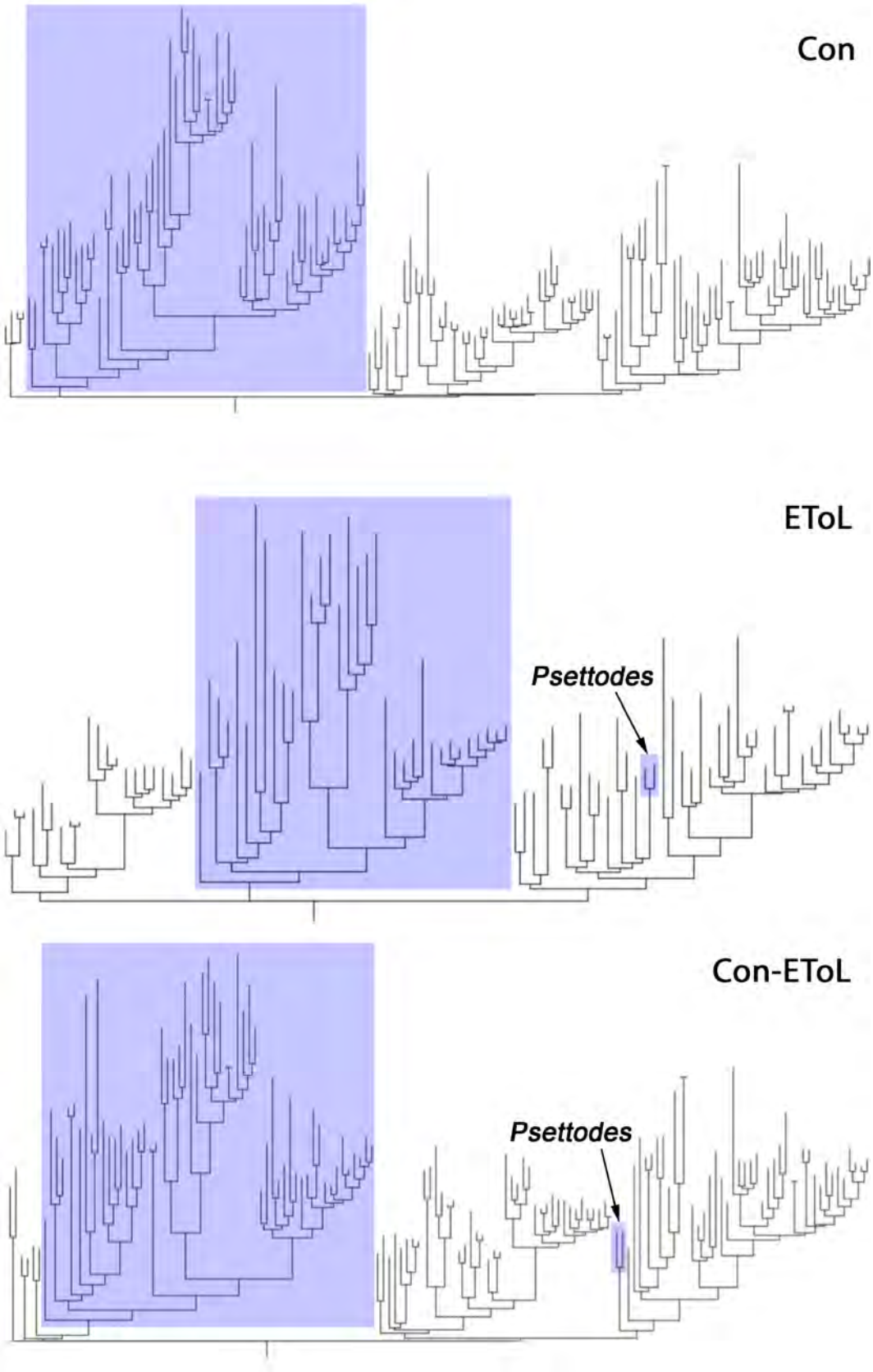

Centropomidae + Polynemidae
Pleuronectiformes
Istiophoriformes
Toxotidae
Nematistiidae
Sphyraenidae
Carangoidei

(note: Helostomatidae as well as the non-anabantiform Anabantimorphariae: order Synbranchiformes (*sensu* Betancur-R et al., 2013a,b) were not included in these data and analyses) recovered as sister to Polycentridae. A well supported Mondactylidae + Inermidae + Emmelichthyidae clade was recovered as sister to a highly supported Scombriformes (*sensu* Betancur-R et al., 2013a).

Phylogenetic analysis of the EToL data (using just the Betancur-R et al. (2013a) trimmed alignment), did not recover a monophyletic Pleuronectiformes (Figs. 3-6 and 3-7). Instead, *Psettodes* was recovered in a well supported, but distantly related (to the other flatfishes), clade with Nematistiidae, Rachycentridae, Coryphaenidae, Toxotidae, Centropomidae and Polynemidae that lacked structure at the base. A monophyletic Pleuronectoidei, however, was recovered with 100% bootstrap support. As with the Con tree, there is no resolution as to which taxon is the sister group of Pleuronectoidei (assuming *Psettodes* is not). All of these findings are consistent with Betancur-R et al.'s (2013a) results inferred using subsets of their much larger dataset.

The phylogeny inferred from all the data discussed here, the Con-EtoL tree (Fig. 3-8), was similar to the EToL tree in that *Psettodes* was not recovered with Pleuronectoidei, however, the addition of the Con data increased the affinity of *Psettodes* to *Nematistius* and recovered them as sister with a bootstrap proportion 85. The Con-EToL data suggest *Toxotes* is actually distantly related to *Nematistius* (and *Psettodes*) and is recovered in a clade with Menidae and Istiophoriformes. The addition of the Con data to the EToL data increased the ambiguity of the placement of Sphyraenidae as well as the monophyletic status of Carangidae, as there was low support for alternative phylogenetic hypotheses. This is somewhat surprising, given that their alignment (consisting of 20

Figure 3-6. Comparison on the reconstruction of Pleuronectiformes (blue) among the Con, EToL and Con-EToL trees; note that scales of branch lengths are not equivalent across trees.


Con
EToL
Psettodes
Con-EToL
Psettodes

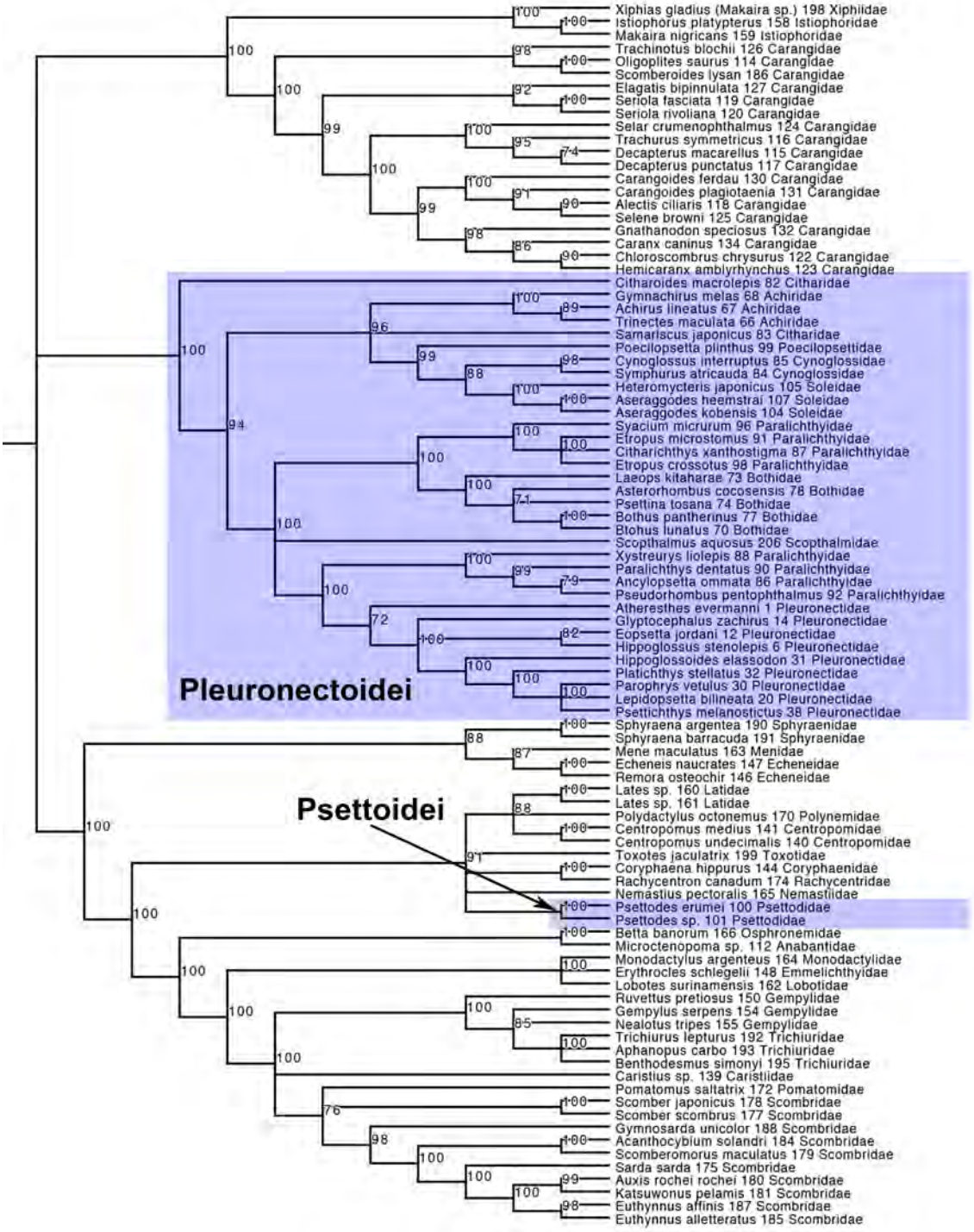
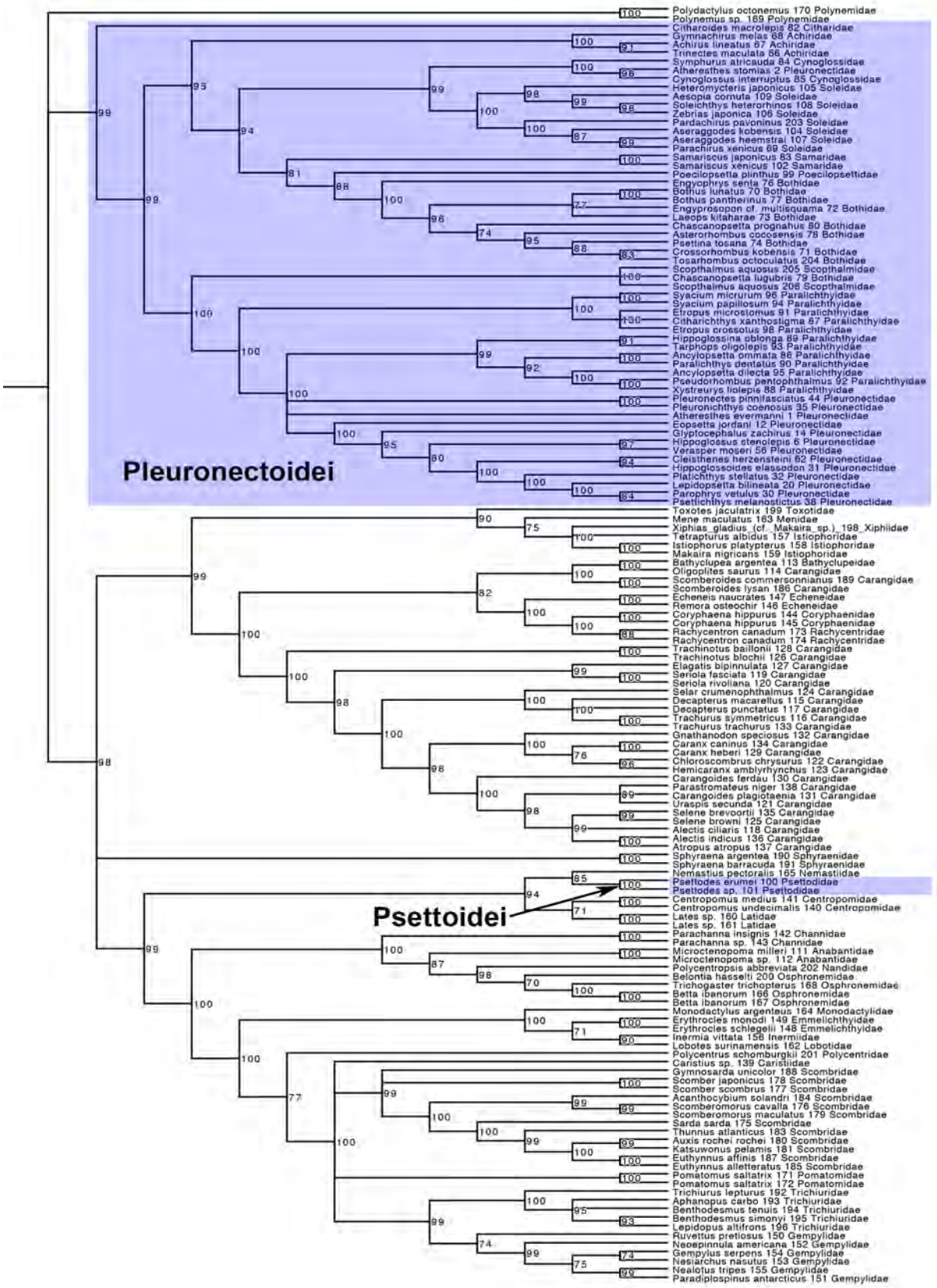


Figure 3-7. Maximum likelihood EToL tree of the trimmed EToL dataset. Blue represents the pleuronectiform suborders. All nodes with a bootstrap proportion < 70 are collapsed.

Figure 3-8. Maximum likelihood Con-EToL tree inferred with all 24 loci. Blue represents pleuronectiform suborders. Nodes with a bootstrap proportion < 70 are collapsed.


Pleuronectoidei
Psettoidei
Polydactylus octonemus 170 Polynemidae
Polynemus sp. 169 Polynemidae
Citharoides macrolepis 82 Citharidae
Gymnachirus melas 88 Achiridae
Achirus lineatus 87 Achiridae
Trinectes maculata 86 Achiridae
Symphurus atricauda 84 Cynoglossidae
Atheresthes stomias 2 Pleuronectidae
Cynoglossus interruptus 85 Cynoglossidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 89 Soleidae
Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Poecilopsetta plinthus 99 Poecilopsettidae
Engyophrys senta 76 Bothidae
Bothus lunatus 70 Bothidae
Bothus pantherinus 77 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta prognathus 80 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Scophthalmus aquosus 205 Scophthalmidae
Chascanopsetta lugubris 79 Bothidae
Scophthalmus aquosus 206 Scophthalmidae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Atheresthes evermanni 1 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
Toxotes jaculatrix 199 Toxotidae
Mene maculatus 163 Menidae
Xiphias_gladius_(cf._Makaira_sp.)_198_Xiphiidae
Tetrapturus albidus 157 Istiophoridae
Istiophorus platypterus 158 Istiophoridae
Makaira nigricans 159 Istiophoridae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Elagatis bipinnulata 127 Carangidae
Seriola fasciata 119 Carangidae
Seriola rivoliana 120 Carangidae
Selar crumenophthalmus 124 Carangidae
Decapterus macarellus 115 Carangidae
Decapterus punctatus 117 Carangidae
Trachurus symmetricus 116 Carangidae
Trachurus trachurus 133 Carangidae
Gnathanodon speciosus 132 Carangidae
Caranx caninus 134 Carangidae
Caranx heberi 129 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Carangoides ferdau 130 Carangidae
Parastromateus niger 138 Carangidae
Carangoides plagiotaenia 131 Carangidae
Uraspis secunda 121 Carangidae
Selene brevoortii 135 Carangidae
Selene browni 125 Carangidae
Alectis ciliaris 118 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Nemastius pectoralis 165 Nemastiidae
Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Centropomus medius 141 Centropomidae
Centropomus undecimalis 140 Centropomidae
Lates sp. 160 Latidae
Lates sp. 161 Latidae
Parachanna insignis 142 Channidae
Parachanna sp. 143 Channidae
Microctenopoma milleri 111 Anabantidae
Microctenopoma sp. 112 Anabantidae
Polycentropsis abbreviata 202 Nandidae
Belontia hasselti 200 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Betta ibanorum 166 Osphronemidae
Betta ibanorum 167 Osphronemidae
Monodactylus argenteus 164 Monodactylidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Inermia vittata 158 Inermiidae
Lobotes surinamensis 162 Lobotidae
Polycentrus schomburgkii 201 Polycentridae
Caristius sp. 139 Caristiidae
Gymnosarda unicolor 188 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Acanthocybium solandri 184 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Trichiurus lepturus 192 Trichiuridae
Aphanopus carbo 193 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Ruvettus pretiosus 150 Gempylidae
Neoepinnula americana 152 Gempylidae
Gempylus serpens 154 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae

markers) provided so much more data, but it is likely that the increased coverage offered by the Con alignment sufficiently dampened some signal, or lack there of.

None of the individual ML gene trees (*rho, rnf213, tmem22* or *chst2*) recovered a monophyletic Pleuronectiformes and all exhibit overall low statistical support and short internodes at the base and, overall, a large variance in branch lengths (Fig. 3-9). Two of the gene trees, *rho* and *chst2*, appear to provide the strongest signal for monophyly, but neither were able to recover *Psettodes* with any other pleuronectoids. The opposite pattern is evident in the *rnf213* and *tmem22* trees: low support for monophyly of the order, but *Psettodes* recovered with other flatfishes. It appears that the combination of the two types of signal and the fact that each provided almost identical amounts of data (*rho* + *chst2* = 1889 sites (52% informative) and *rnf213* + *tmem22* = 1892 sites (55% informative)) to the Con alignment was the reason that support for monophyly in the Con phylogeny was so high as well as the placement of Psettoidei as sister to Pleuronectoidei. The gene tree discordance seems to result in a harmonious and stable node once all the data are concatenated and analyzed together, regardless of whether the data was partitioned by gene or codon.

In addition to the gene trees, each two-gene and three-gene partition was analyzed to ascertain at which point, if one exists, there is enough data to recover the Con topology where Pleuronectiformes is monophyletic; the two-gene trees are summarized in Figs. 3-10 and the three gene trees in Fig. 3-11. None of the ML trees inferred from these combinations were able to recover a monophyletic Pleuronectiformes, but as data were added, the topologies began to converge on the Con tree. However, no particular gene or gene combination provides high support for monophyly. In this case, it is only the

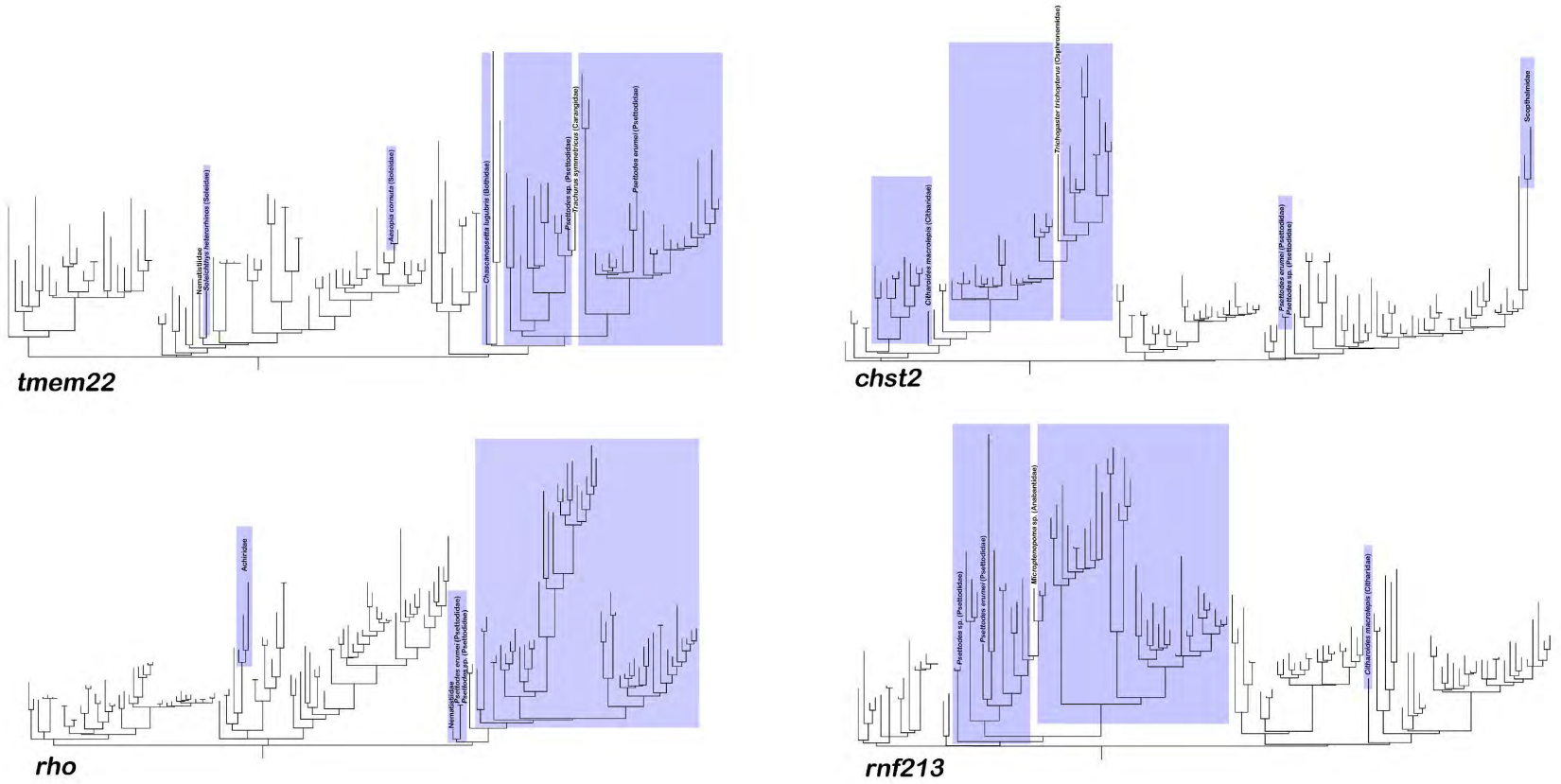


Figure 3-9. Comparison of gene tree topologies with branch lengths. Note that none recover a monophyletic Pleuronectiformes (blue) and all have short internodes at the base.

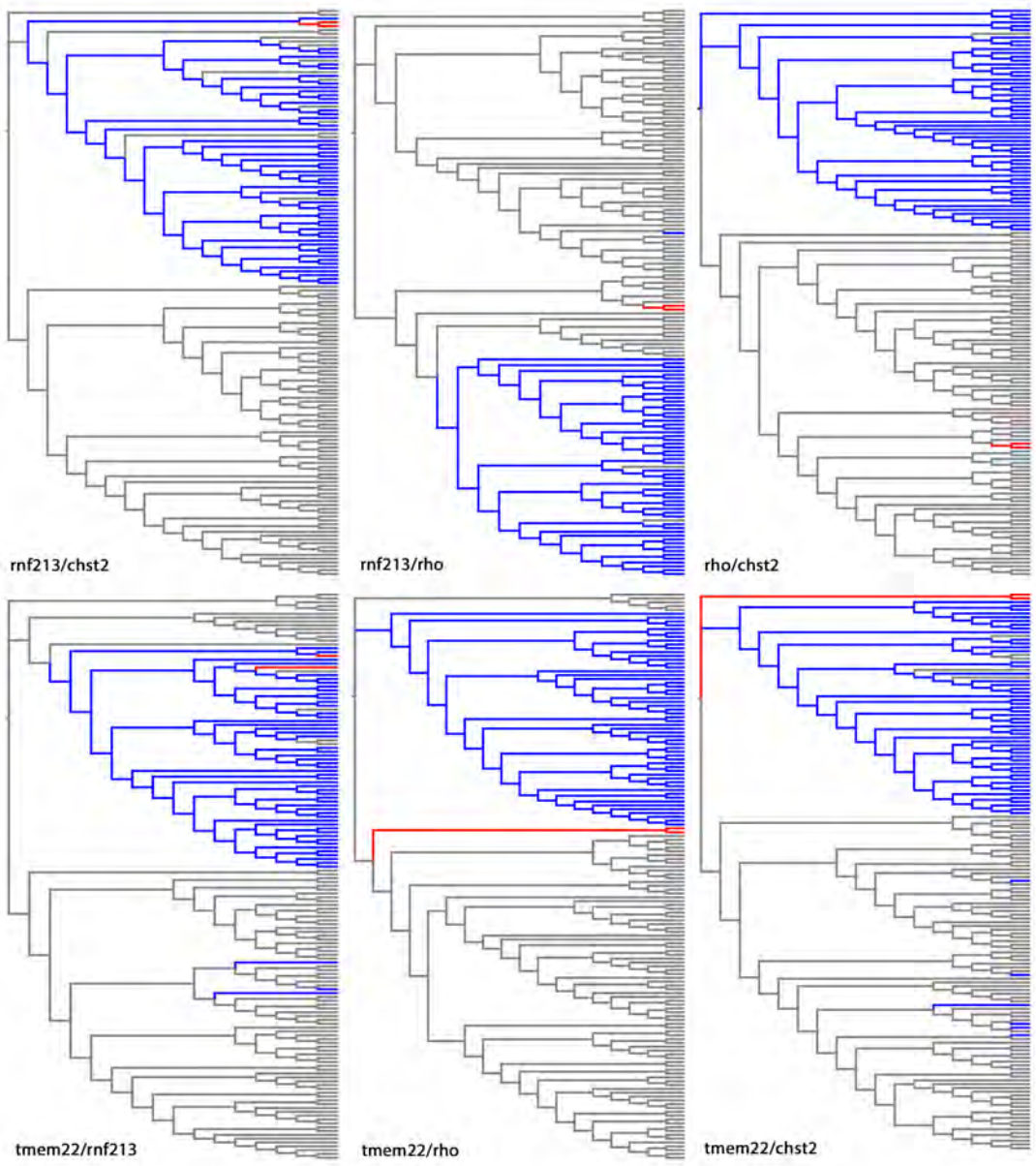


Figure 3-10. Topology comparison of each three-partition tree. Blue represents Pleuronectoidei and red represents Psettoidei.

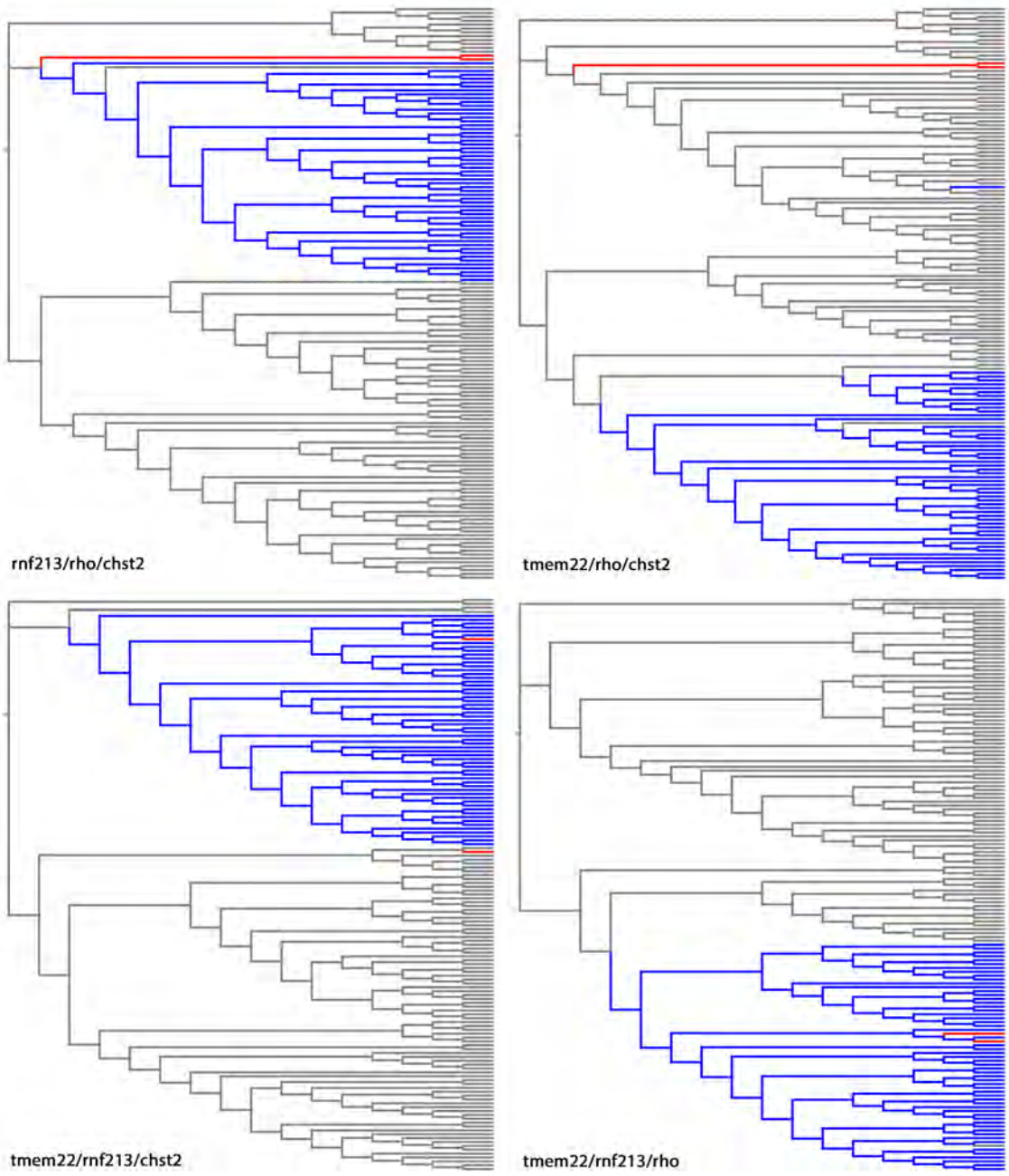


Figure 3-11. Topology comparison of each three-partition tree. Blue represents Pleuronectoidei and red represents Psettoidei.

analysis of the entire, concatenated, dataset of these four genes that resolves a monophyletic Pleuronectiformes with Psettoidei sister to Pleuronectoidei.

3.4 TOPOLOGY TESTING

All possible pairwise comparisons between the individual gene, Con, EToL and the Con-EToL datasets and ML topologies were carried out and the results are summarized in Table 3-3. There were only a few instances where the null hypothesis (H_0: all topologies for comparison are equally good explanations of the data) was not rejected ($P > 0.05$): the SH tests comparing the Con tree to the *chst2* alignment ($P = 0.19466$), the Con and Con-EToL trees to the *rho* alignment ($P = 0.0695$ and 0.04592, respectively), the Con tree to the Con-EToL data ($P = 0.0695$) and the EToL data to the Con-EToL tree ($P = 1$).

Surprisingly, the P-value for the latter SH test as well as the KH and AU tests, was much higher than the P-values from the comparison of the EToL data to its own ML topology ($P = 0.42422$). This comparison may indicate that the signal for flatfish monophyly is in the EToL data, but was masked by the effect of so much missing data, and fewer terminals (94 versus 148 in the trimmed alignment). Except in the case of the EToL data being compared to the Con-EToL tree and its own topology, the SH tests were the only ones that resulted in P-values greater than zero. This is not surprising since of the three topology tests, the SH test, although very good at controlling its type I error, is the most conservative (Shimodaira, 2002; Jobb; 2008). Its bias, however, increases with the number of input trees (Strimmer and Rambaut, 20002), and so only seven trees were used for comparison here.

Table 3-3. Results of the KH, AU and SH topology tests. The null hypothesis (H_0: all topologies for comparison are equally good explanations of the data) was rejected when $P < 0.05$. Boldface rows represent instances where at least one test failed to reject the null hypothesis. The bolded and italicized row represents the case where another topology (Con-ETоL) was a better explanation of the ETоL data its own ML topology.

Table 3-3 (continued).

Data	Topology	KH Test	AU Test	SH Test
tmem22	*tmem22*	—	—	—
	chst2	0.000	0.000	0.000
	rnf213	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	EToL	0.000	0.000	0.000
	Con	0.000	0.000	0.000
	Con-EToL	0.000	0.000	0.000
chst2	*tmem22*	0.000	0.000	0.000
	chst2	—	—	—
	rnf213	0.000	0.000	0.000
	rho	0.000	0.000	0.00038
	EToL	0.000	0.000	0.000
	Con	**0.000**	**0.000**	**0.19466**
	Con-EToL	0.000	0.000	0.03058
rnf213	*tmem22*	0.000	0.000	0.000
	chst2	0.000	0.000	0.000
	rnf213	—	—	—
	rho	0.000	0.000	0.000
	EToL	0.000	0.000	0.000
	Con	0.000	0.000	0.000
	Con-EToL	0.000	0.000	0.000
rho	*tmem22*	0.000	0.000	0.000
	chst2	0.000	0.000	0.000
	rnf213	0.000	0.000	0.000
	rho	—	—	—
	EToL	0.000	0.000	0.000
	Con	**0.000**	**0.000**	**0.0695**
	Con-EToL	**0.000**	**0.000**	**0.04592**
Con	*tmem22*	0.000	0.000	0.000
	chst2	0.000	0.000	0.000
	rnf213	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	EToL	0.000	0.000	0.000
	Con	—	—	—
	Con-EToL	**0.000**	**0.000**	**0.0695**
EToL	*tmem22*	0.000	0.000	0.000
	chst2	0.000	0.000	0.000
	rnf213	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	EToL	***0.00044***	***0.000***	***0.42422***
	Con	0.000	0.000	0.000
	Con-EToL	**1**	**0.9617**	**1**
Con-EToL	*tmem22*	0.000	0.000	0.000
	chst2	0.000	0.000	0.000
	rnf213	0.000	0.000	0.000
	rho	0.000	0.000	0.000
	EToL	0.000	0.000	0.000
	Con	0.000	0.000	0.000
	Con-EToL	—	—	—

As mentioned above, based on the gene trees it appears that the phylogenetic signal provided by the four genes making up the Con dataset can be partitioned into two types, with each providing the same quantity of data: 1) a stronger signal for flatfish monophyly, but the inability to recover *Psettodes* with pleuronectoids (*rho* and *chst2*) and 2) a weak signal for monophyly of flatfishes, but the ability to recover *Psettodes* with other flatfishes (*rnf213* and *tmem22*). This discordance is likely the cause of the non-zero P-value generated for the SH test when comparing the *rho* topology to the *chst2* data. Even though it was greater than zero, it is still much too low (0.00038) to justify rejection of the null hypothesis.

All other topology comparisons to individual gene alignments resulted in P-values of zero, except for the SH test comparing the *chst2* data to the Con topology; that value was 0.03058, much closer to a P-value indicating rejection of the null hypotheses. If a similar result were observed in the other two, less conservative tests, rejection of H_0 would be acceptable, but that is not the case here. This result highlights the contribution of this new SEG marker to test phylogenetic hypotheses such as flatfish monophyly since its signal is congruent with that of the Con tree, the best estimate of phylogeny given all the data analyzed for this study.

3.5 Base Compositional Bias and Gene Tree Incongruence

For the four markers sequenced here, base composition was mapped as a continuous character on each individual gene tree (Figs. 3-12 to 3-15), as well as the Con tree (Fig. 3-16) and compositional bias was observed in all four markers. Overall, GC content was low in the *rnf213* data (Fig. 3-14) and the lowest values were observed in

Figure 3-12. Base compositional bias mapped on *rho* tree with bootstrap support at nodes. Cool to warm color gradient represent lower to higher proportions of G+C.

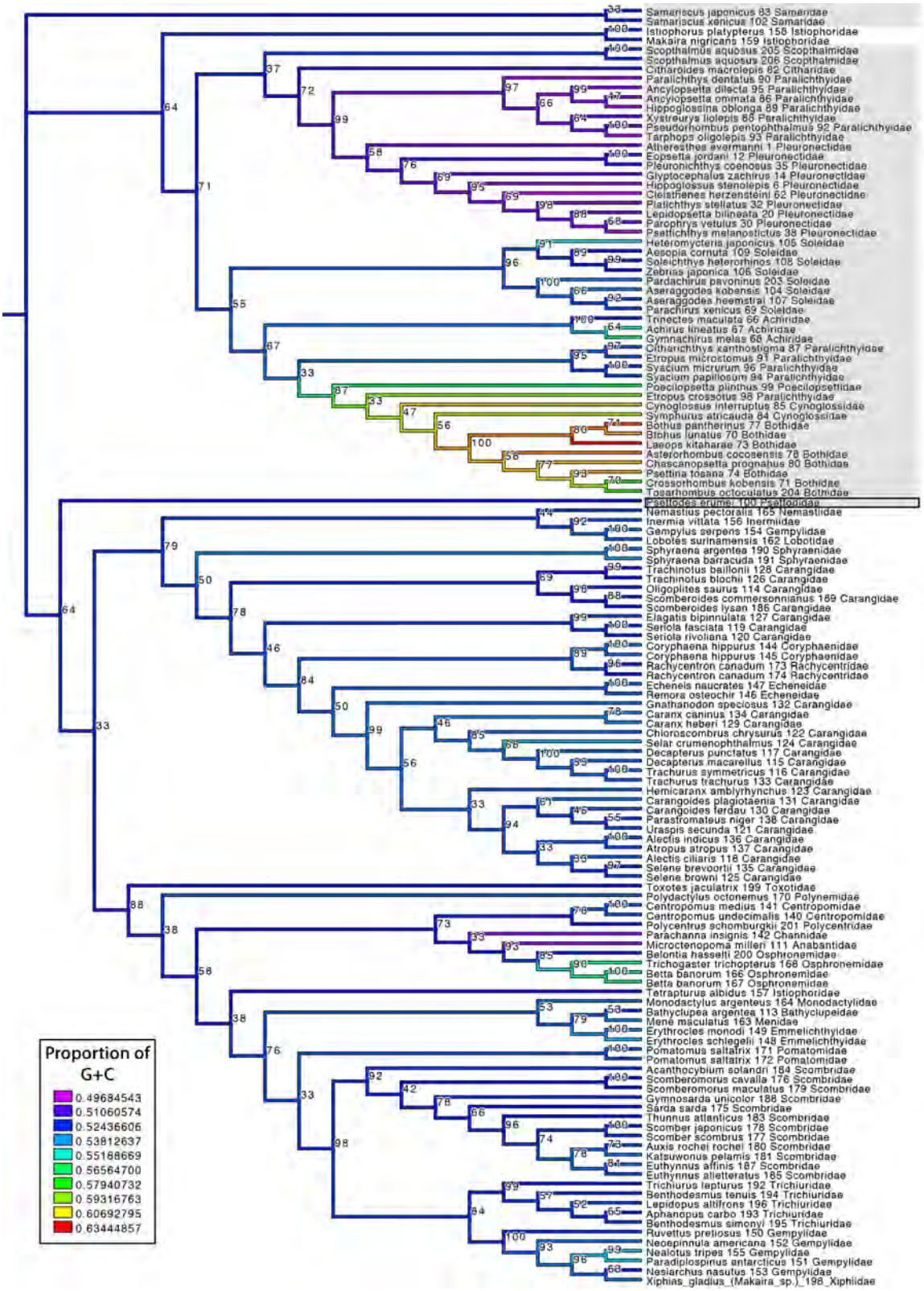

Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Istiophorus platypterus 158 Istiophoridae
Makaira nigricans 159 Istiophoridae
Scopthalmus aquosus 205 Scopthalmidae
Scopthalmus aquosus 206 Scopthalmidae
Citharoides macrolepis 82 Citharidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Trinectes maculata 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 68 Achiridae
Otharchthys xanthostigma 87 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Etropus crossotus 98 Paralichthyidae
Cynoglossus interruptus 85 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Bothus pantherinus 77 Bothidae
Bothus lunatus 70 Bothidae
Laeops kitaharae 73 Bothidae
Asterorhombus cocosensis 78 Bothidae
Chascanopsetta prognathus 80 Bothidae
Psettina tosana 74 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Psettodes erumei 100 Psettodidae
Nemastius pectoralis 165 Nemastiidae
Inermia vittata 156 Inermiidae
Gempylus serpens 154 Gempylidae
Lobotes surinamensis 162 Lobotidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Oligoplites saurus 114 Caranglidae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Elagatis bipinnulata 127 Carangidae
Seriola fasciata 119 Carangidae
Seriola rivoliana 120 Carangidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Gnathanodon speciosus 132 Carangidae
Caranx caninus 134 Carangidae
Caranx heberi 129 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Selar crumenophthalmus 124 Carangidae
Decapterus punctatus 117 Carangidae
Decapterus macarellus 115 Carangidae
Trachurus symmetricus 116 Carangidae
Trachurus trachurus 133 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Carangoides plagiotaenia 131 Carangidae
Carangoides ferdau 130 Carangidae
Parastromateus niger 138 Carangidae
Uraspis secunda 121 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Alectis ciliaris 118 Carangidae
Selene brevoortii 135 Carangidae
Selene browni 125 Carangidae
Toxotes jaculatrix 199 Toxotidae
Polydactylus octonemus 170 Polynemidae
Centropomus medius 141 Centropomidae
Centropomus undecimalis 140 Centropomidae
Polycentrus schomburgkii 201 Polycentridae
Parachanna insignis 142 Channidae
Microctenopoma milleri 111 Anabantidae
Belontia hasselti 200 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Betta banorum 166 Osphronemidae
Betta banorum 167 Osphronemidae
Tetrapturus albidus 157 Istiophoridae
Monodactylus argenteus 164 Monodactylidae
Bathyclupea argentea 113 Bathyclupeidae
Mene maculatus 163 Menidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Acanthocybium solandri 184 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Gymnosarda unicolor 188 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Trichiurus lepturus 192 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Aphanopus carbo 193 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Ruvettus pretiosus 150 Gempylidae
Neoepinnula americana 152 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Xiphias_gladius_(Makaira_sp.)_198_Xiphiidae
Proportion of G+C
0.49684543
0.51060574
0.52436606
0.53812637
0.55188669
0.56564700
0.57940732
0.59316763
0.60692795
0.63444857

Figure 3-13. Base compositional bias mapped on *chst2* tree with bootstrap support at nodes. Cool to warm color gradient represent lower to higher proportions of G+C.

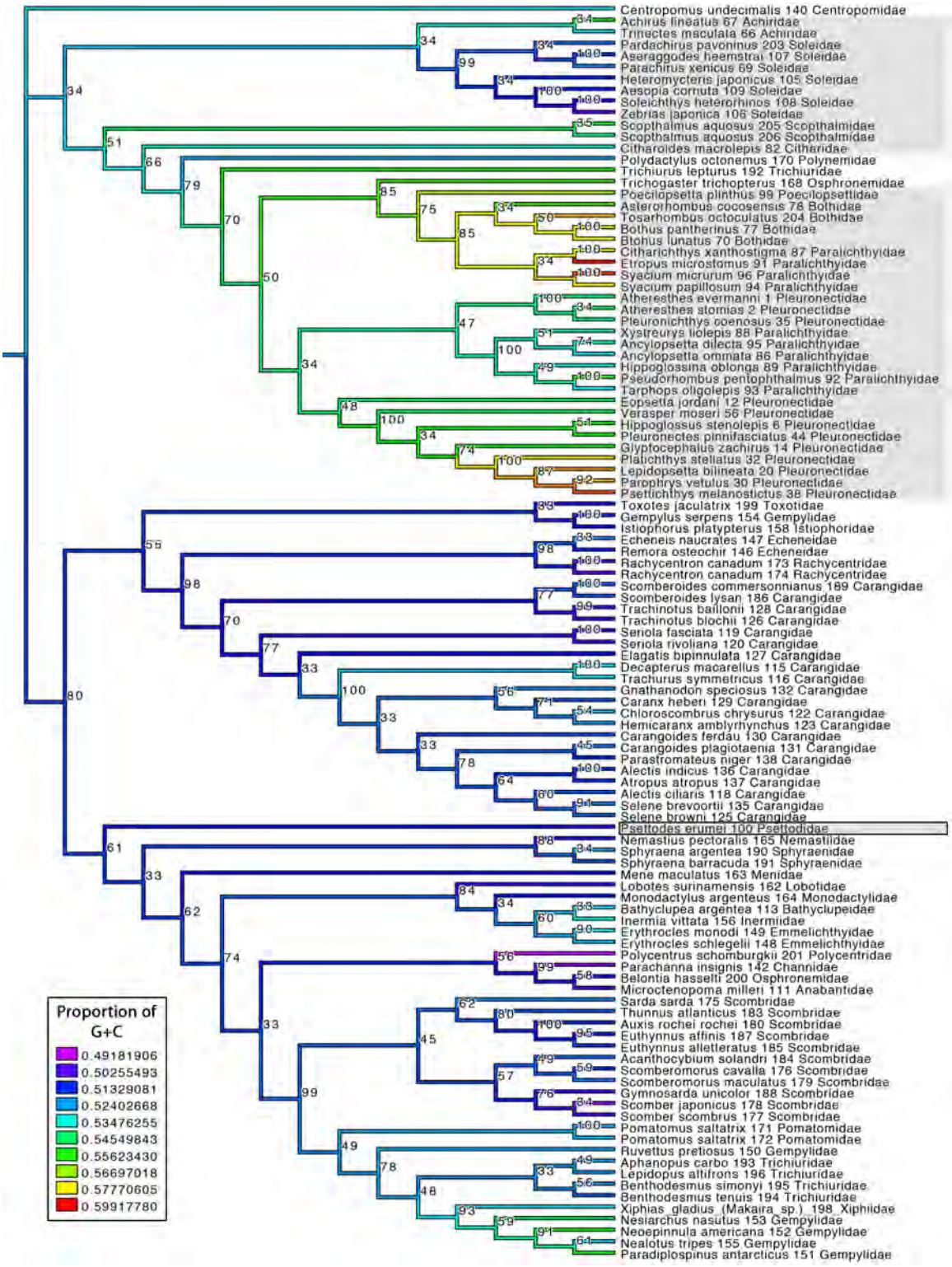

Centropomus undecimalis 140 Centropomidae
Achirus lineatus 67 Achiridae
Trinectes maculata 66 Achiridae
Pardachirus pavoninus 203 Soleidae
Aseraggodes haemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Scopthalmus aquosus 205 Scopthalmidae
Scopthalmus aquosus 206 Scopthalmidae
Citharoides macrolepis 82 Citharidae
Polydactylus octonemus 170 Polynemidae
Trichiurus lepturus 192 Trichiuridae
Trichogaster trichopterus 168 Osphronemidae
Poecilopsetta plinthus 99 Poecilopsettidae
Asterorhombus cocosensis 78 Bothidae
Tosarhombus octoculatus 204 Bothidae
Bothus pantherinus 77 Bothidae
Btohus lunatus 70 Bothidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Xystreurys liolepis 88 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Eopsetta jordani 12 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
Toxotes jaculatrix 199 Toxotidae
Gempylus serpens 154 Gempylidae
Istiophorus platypterus 158 Istiophoridae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Seriola fasciata 119 Carangidae
Seriola rivoliana 120 Carangidae
Elagatis bipinnulata 127 Carangidae
Decapterus macarellus 115 Carangidae
Trachurus symmetricus 116 Carangidae
Gnathanodon speciosus 132 Carangidae
Caranx heberi 129 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Carangoides ferdau 130 Carangidae
Carangoides plagiotaenia 131 Carangidae
Parastromateus niger 138 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Alectis ciliaris 118 Carangidae
Selene brevoortii 135 Carangidae
Selene browni 125 Carangidae
Psettodes erumei 100 Psettodidae
Nemastius pectoralis 165 Nemastiidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Mene maculatus 163 Menidae
Lobotes surinamensis 162 Lobotidae
Monodactylus argenteus 164 Monodactylidae
Bathyclupea argentea 113 Bathyclupeidae
Inermia vittata 156 Inermiidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Polycentrus schomburgkii 201 Polycentridae
Parachanna insignis 142 Channidae
Belontia hasselti 200 Osphronemidae
Microctenopoma milleri 111 Anabantidae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Auxis rochei rochei 180 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Acanthocybium solandri 184 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Gymnosarda unicolor 188 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Ruvettus pretiosus 150 Gempylidae
Aphanopus carbo 193 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Xiphias gladius (Makaira sp.) 198 Xiphiidae
Nesiarchus nasutus 153 Gempylidae
Neoepinnula americana 152 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae
Proportion of G+C
0.49181906
0.50255493
0.51329081
0.52402668
0.53476255
0.54549843
0.55623430
0.56697018
0.57770605
0.59917780

Figure 3-14. Base compositional bias mapped on *rnf213* tree with bootstrap support at nodes. Cool to warm color gradient represent lower to higher proportions of G+C.

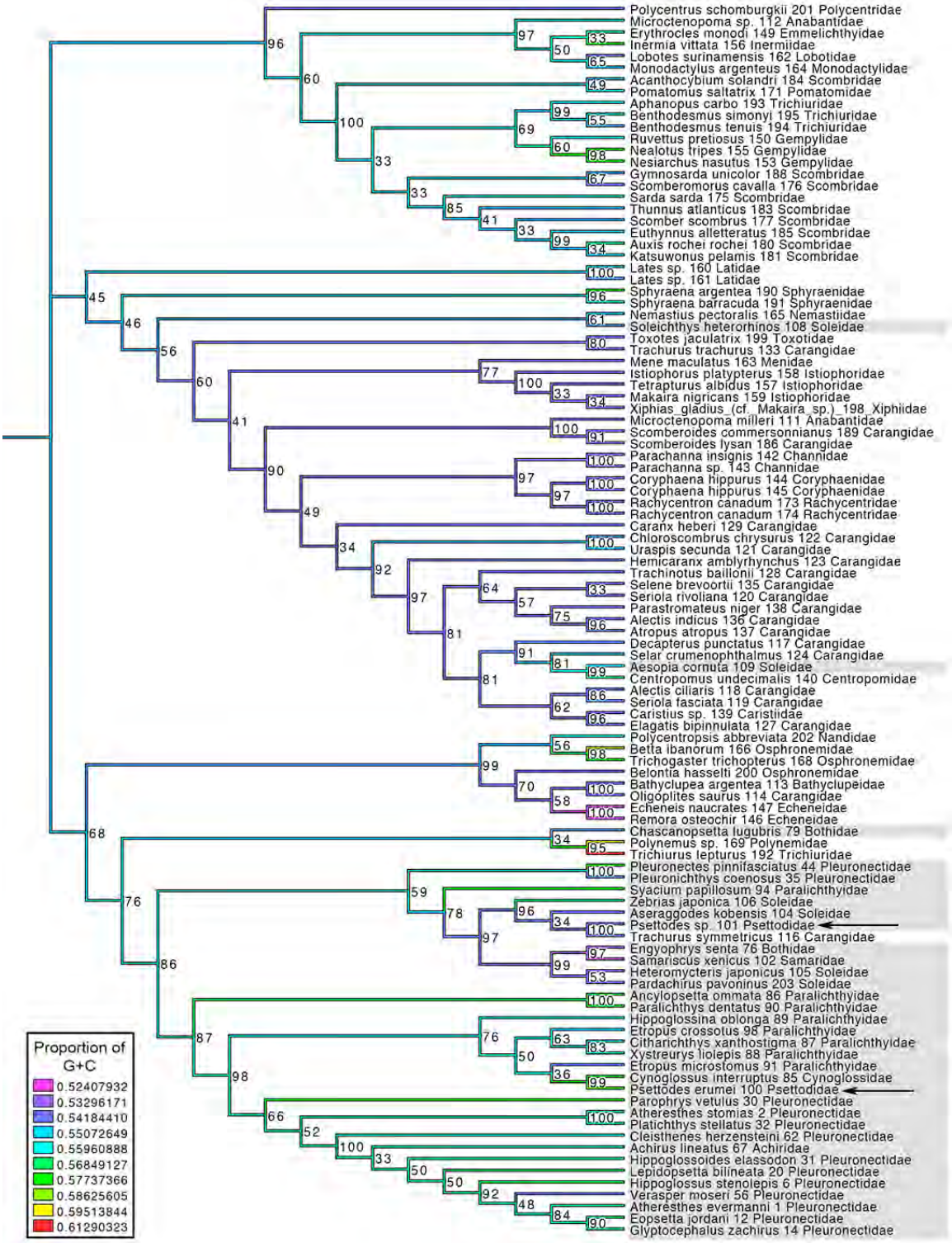
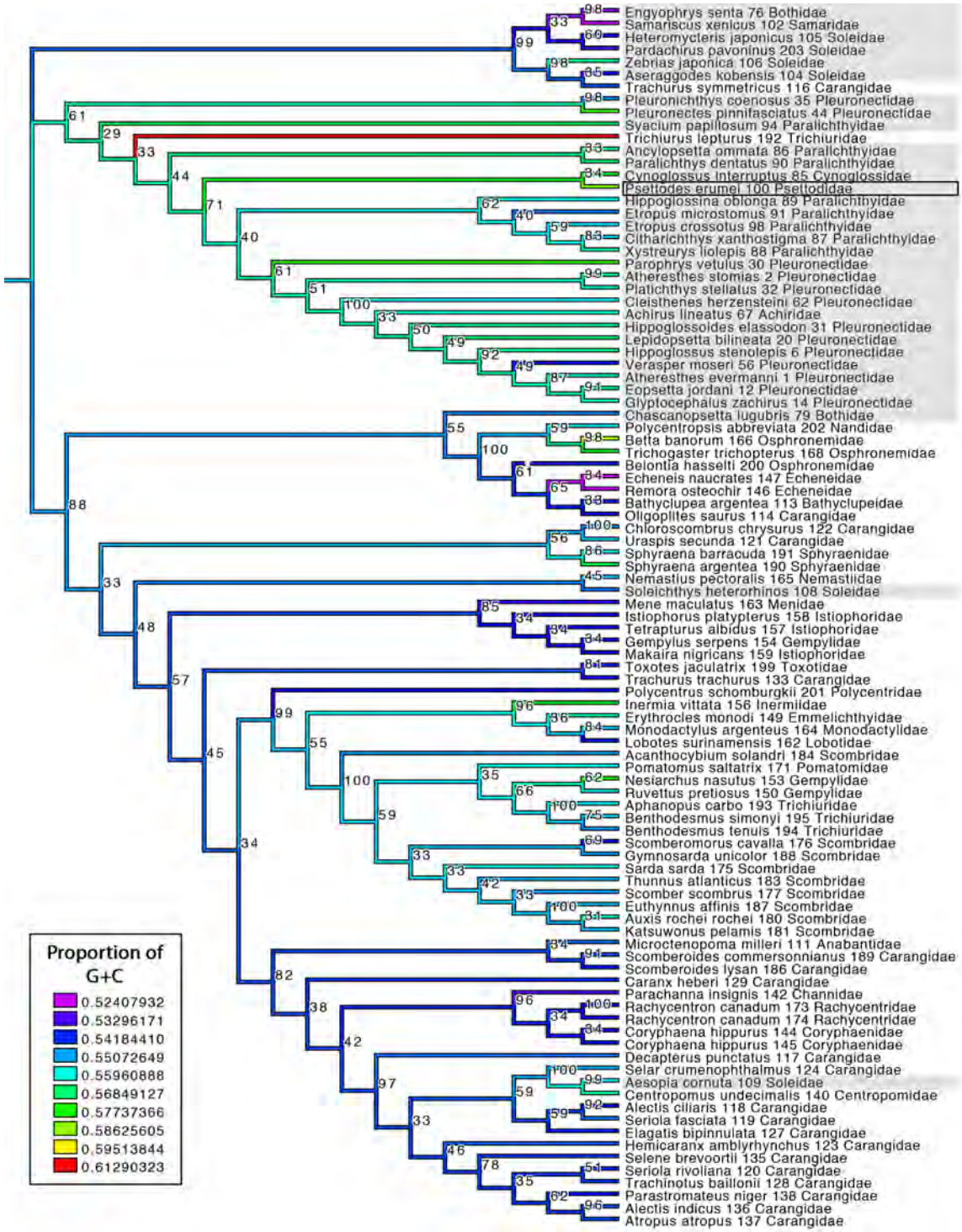

Polycentrus schomburgkii 201 Polycentridae
Microctenopoma sp. 112 Anabantidae
Erythrocles monodi 149 Emmelichthyidae
Inermia vittata 156 Inermiidae
Lobotes surinamensis 162 Lobotidae
Monodactylus argenteus 164 Monodactylidae
Acanthocybium solandri 184 Scombridae
Pomatomus saltatrix 171 Pomatomidae
Aphanopus carbo 193 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Ruvettus pretiosus 150 Gempylidae
Nealotus tripes 155 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Gymnosarda unicolor 188 Scombridae
Scomberomorus cavalla 176 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Scomber scombrus 177 Scombridae
Euthynnus alletteratus 185 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Lates sp. 160 Latidae
Lates sp. 161 Latidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Nemastius pectoralis 165 Nemastiidae
Soleichthys heterorhinos 108 Soleidae
Toxotes jaculatrix 199 Toxotidae
Trachurus trachurus 133 Carangidae
Mene maculatus 163 Menidae
Istiophorus platypterus 158 Istiophoridae
Tetrapturus albidus 157 Istiophoridae
Makaira nigricans 159 Istiophoridae
Xiphias_gladius_(cf._Makaira_sp.)_198_Xiphiidae
Microctenopoma milleri 111 Anabantidae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Parachanna insignis 142 Channidae
Parachanna sp. 143 Channidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Caranx heberi 129 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Uraspis secunda 121 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Trachinotus baillonii 128 Carangidae
Selene brevoortii 135 Carangidae
Seriola rivoliana 120 Carangidae
Parastromateus niger 138 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Decapterus punctatus 117 Carangidae
Selar crumenophthalmus 124 Carangidae
Aesopia cornuta 109 Soleidae
Centropomus undecimalis 140 Centropomidae
Alectis ciliaris 118 Carangidae
Seriola fasciata 119 Carangidae
Caristius sp. 139 Caristiidae
Elagatis bipinnulata 127 Carangidae
Polycentropsis abbreviata 202 Nandidae
Betta ibanorum 166 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Belontia hasselti 200 Osphronemidae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Chascanopsetta lugubris 79 Bothidae
Polynemus sp. 169 Polynemidae
Trichiurus lepturus 192 Trichiuridae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Syacium papillosum 94 Paralichthyidae
Zebrias japonica 106 Soleidae
Aseraggodes kobensis 104 Soleidae
Psettodes sp. 101 Psettodidae
Trachurus symmetricus 116 Carangidae
Engyophrys senta 76 Bothidae
Samariscus xenicus 102 Samaridae
Heteromycteris japonicus 105 Soleidae
Pardachirus pavoninus 203 Soleidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Cynoglossus interruptus 85 Cynoglossidae
Psettodes erumei 100 Psettodidae
Parophrys vetulus 30 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Achirus lineatus 67 Achiridae
Hippoglossoides elassodon 31 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Atheresthes evermanni 1 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Proportion of G+C
0.52407932
0.53296171
0.54184410
0.55072649
0.55960888
0.56849127
0.57737366
0.58625605
0.59513844
0.61290323

Figure 3-15. Base compositional bias mapped on *tmem22* tree with bootstrap support at nodes. Cool to warm color gradient represent lower to higher proportions of G+C.


Engyophrys senta 76 Bothidae
Samariscus xenicus 102 Samaridae
Heteromycteris japonicus 105 Soleidae
Pardachirus pavoninus 203 Soleidae
Zebrias japonica 106 Soleidae
Aseraggodes kobensis 104 Soleidae
Trachurus symmetricus 116 Carangidae
Pleuronichthys coenosus 35 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Syacium papillosum 94 Paralichthyidae
Trichiurus lepturus 192 Trichiuridae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Cynoglossus interruptus 85 Cynoglossidae
Psettodes erumei 100 Psettodidae
Hippoglossina oblonga 89 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Parophrys vetulus 30 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Achirus lineatus 67 Achiridae
Hippoglossoides elassodon 31 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Atheresthes evermanni 1 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Chascanopsetta lugubris 79 Bothidae
Polycentropsis abbreviata 202 Nandidae
Betta banorum 166 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Belontia hasselti 200 Osphronemidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Uraspis secunda 121 Carangidae
Sphyraena barracuda 191 Sphyraenidae
Sphyraena argentea 190 Sphyraenidae
Nemastius pectoralis 165 Nemastiidae
Soleichthys heterorhinos 108 Soleidae
Mene maculatus 163 Menidae
Istiophorus platypterus 158 Istiophoridae
Tetrapturus albidus 157 Istiophoridae
Gempylus serpens 154 Gempylidae
Makaira nigricans 159 Istiophoridae
Toxotes jaculatrix 199 Toxotidae
Trachurus trachurus 133 Carangidae
Polycentrus schomburgkii 201 Polycentridae
Inermia vittata 156 Inermiidae
Erythrocles monodi 149 Emmelichthyidae
Monodactylus argenteus 164 Monodactylidae
Lobotes surinamensis 162 Lobotidae
Acanthocybium solandri 184 Scombridae
Pomatomus saltatrix 171 Pomatomidae
Nesiarchus nasutus 153 Gempylidae
Ruvettus pretiosus 150 Gempylidae
Aphanopus carbo 193 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Scomberomorus cavalla 176 Scombridae
Gymnosarda unicolor 188 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Scomber scombrus 177 Scombridae
Euthynnus affinis 187 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Microctenopoma milleri 111 Anabantidae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Caranx heberi 129 Carangidae
Parachanna insignis 142 Channidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Decapterus punctatus 117 Carangidae
Selar crumenophthalmus 124 Carangidae
Aesopia cornuta 109 Soleidae
Centropomus undecimalis 140 Centropomidae
Alectis ciliaris 118 Carangidae
Seriola fasciata 119 Carangidae
Elagatis bipinnulata 127 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Selene brevoortii 135 Carangidae
Seriola rivoliana 120 Carangidae
Trachinotus baillonii 128 Carangidae
Parastromateus niger 138 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Proportion of G+C
0.52407932
0.53296171
0.54184410
0.55072649
0.55960888
0.56849127
0.57737366
0.58625605
0.59513844
0.61290323

Soleidae, as well as *Parachanna, Bathyclupea, Trichogaster* (all recovered with *Betta*), and one *Coryphaena hippurus,* (recovered with other carangiforms and *Polydactylus*).

Higher bias towards GC in *rnf213* was clear in a few flatfish taxa, namely *Xystreurys, Poecilopsetta* and *Pseudorhombus,* which were recovered together. This bias may explain the unusual recovery of *Poecilopsetta* (nested within the *Paralichthys* + *Pseudorhombus* group) in the Con tree. A similar result was seen in *rho* where *Poecilopsetta* is recovered with other taxa showing high GC bias, but in this case most of the clade is comprised of bothids and cynoglossids not paralichthyids as in the Con tree. *Psettodes* exhibited low bias towards GC in *rho,* as did most other taxa (flatfish and non-flatfish), including the istiophorids, which replace *Psettodes* as sister to Pleuronectoidei (minus *Samariscus*) in the *rho* ML tree, albeit with low support.

Both the SEG trees, in addition to the others, do not recover a monophyletic Pleuronectiformes, however, in the case of *tmem22* it is not *Psettodes* that is recovered apart from flatfishes, but two soleids: *Soleichthys* and *Aesopia*. Convergence towards in GC content does not explain this placement, as both genera exhibit similar biases as those observed in the other soleids, recovered together. However, the soleids that did group together were not recovered with the large flatfish clade that showed higher bias overall.

The other SEG, *chst2*, was similar to *tmem22* in the overall distribution of base compositional bias on the tree, with the flatfishes having higher biases towards GC than the non-flatfishes; *Psettodes* and Soleidae were the exceptions. In this case, however, *Psettodes* was not recovered with the majority of flatfishes, instead grouping within a large group of non-flatfishes, most of which also showed low bias towards GC. If base compositional bias were the only cause for the placement of *Psettodes* outside of

flatfishes, than the same would be expected of the soleids, and that was not observed. In addition, the scopthalmids exhibited GC content more similar to the flatfishes (except Soleidae and *Psettodes*), but were recovered with high support with scombriforms.

When base compositional bias was mapped on the Con tree (Fig. 3-16) it is clear that, overall, Pleuronectiformes exhibit higher bias towards GC (especially the bothids and pleuronectids). But many, including the psettoids, *Citharoides* and the soleids are more similar to outgroup taxa in GC content, and are still recovered within Pleuronectiformes with high support. In some cases base compositional bias can explain gene tree discordance (or at least part of the discordance), but the pattern observed in these data is not clear-cut, suggesting that even if base compositional bias is the cause for some systematic error, it is not the only factor.

3.6 Substitution Saturation and Gene Tree Incongruence

Of the all saturation curves, the third position of *rho* exhibits the highest degree of saturation both in transitions and transversions (Fig. 3-12). All four sets of data, however, had a level 1 change-point detected with 100% confidence level for *rho*-3, *chst2*-3 and *tmem22*-3, and 99% for *rnf213*. The TPWD values for the transitions (the blue "X"s) are placed on each of the curves in Figure 3-17. The phylogenetic signal contributed by the four markers can be categorized as either 1) providing a strong signal for pleuronectoid monophyly or 2) providing low support for monophyly of the order, but recover *Psettodes* with pleuronectoids. Because of the tendency of *Psettodes* to act like a "rogue" taxon the F84 distances from all other taxa (except *Psettodes* sp.) to *Psettodes erumei* were plotted to see if the two patterns observed in the gene trees could be explained by saturation of transitions at third codon positions. All F84 distance matrices

Figure 3-16. Base compositional bias mapped on Con tree with bootstrap support at nodes. Cool to warm color gradient represent lower to higher proportions of G+C.

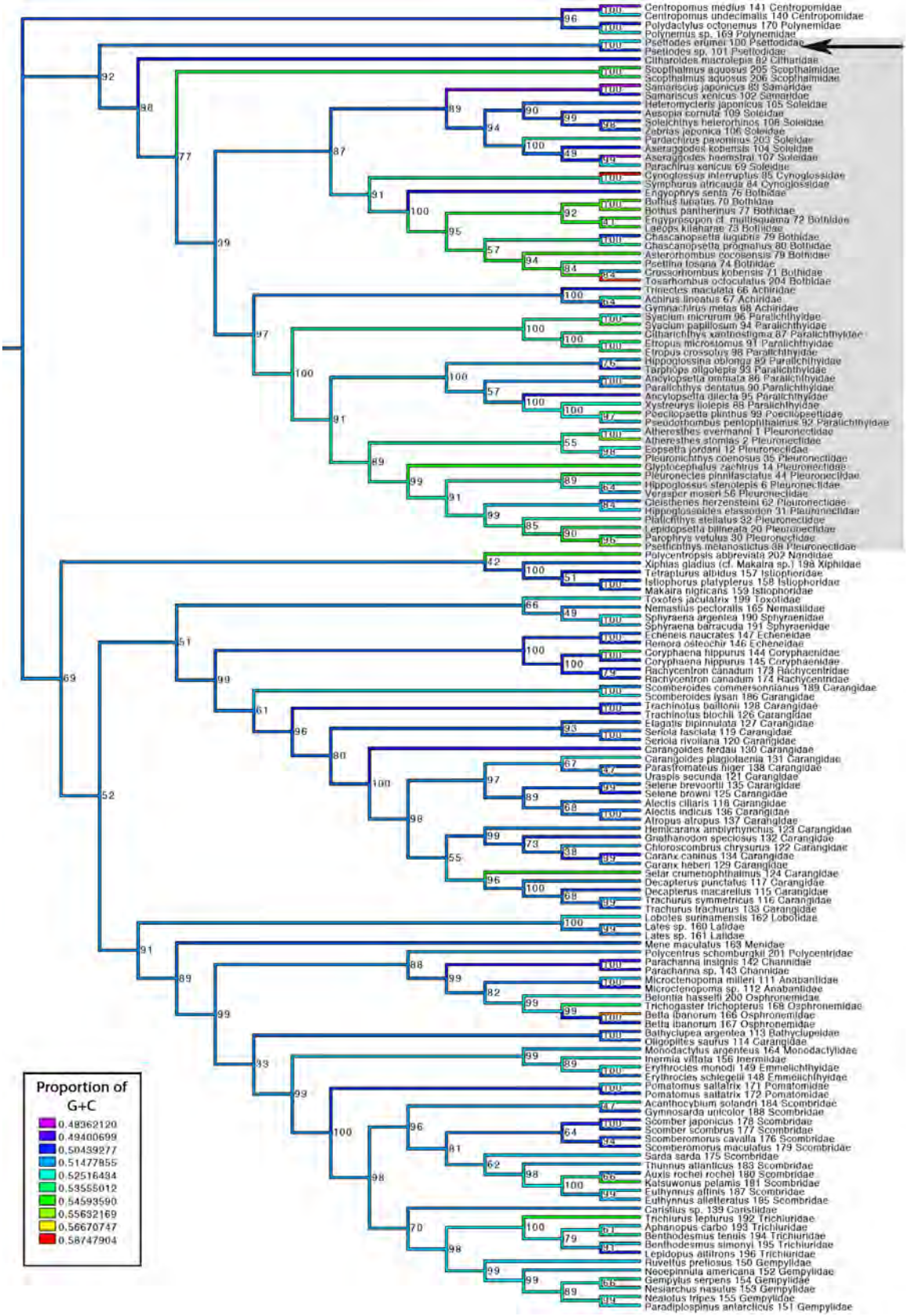
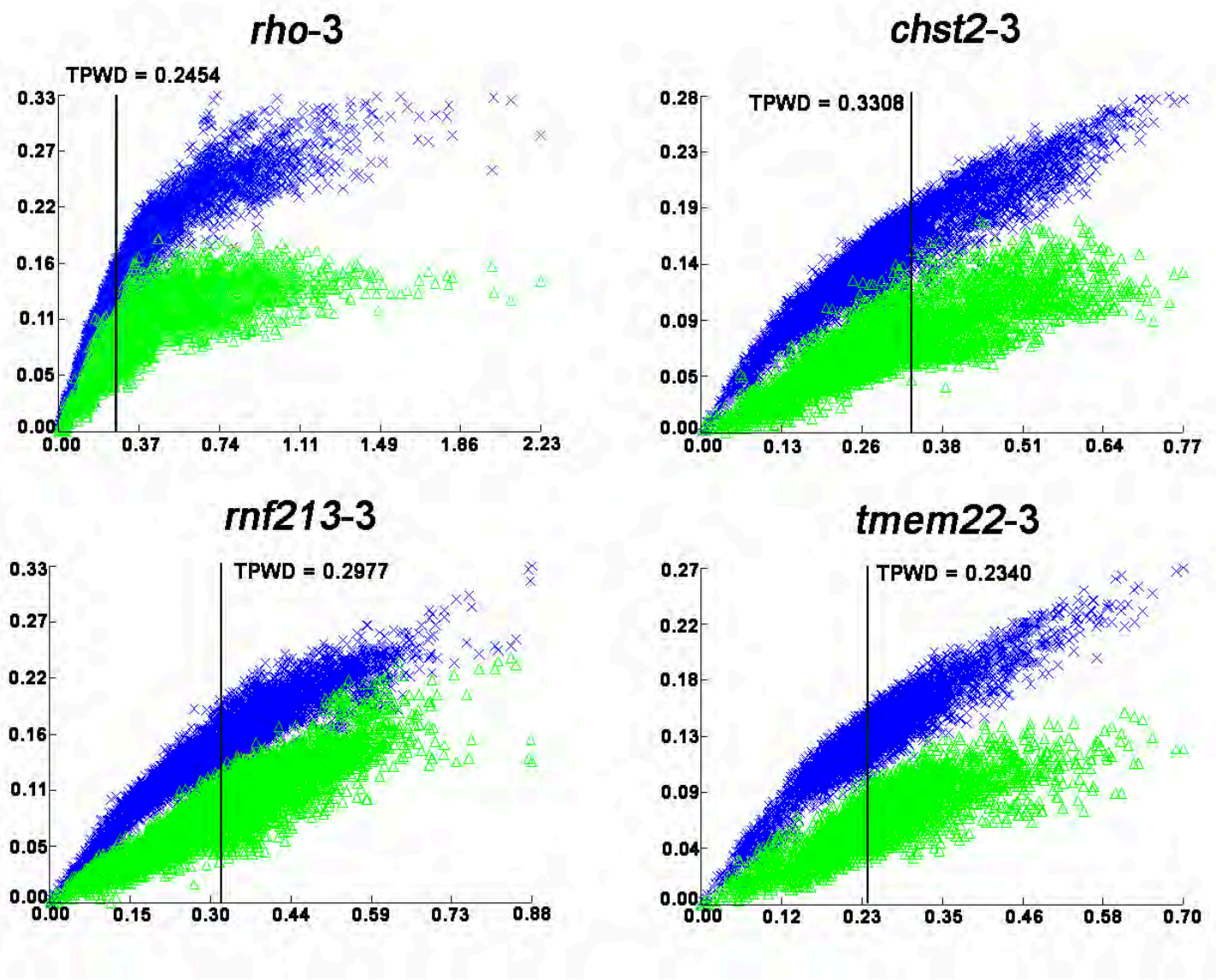

Centropomus medius 141 Centropomidae
Centropomus undecimalis 140 Centropomidae
Polydactylus octonemus 170 Polynemidae
Polynemus sp. 169 Polynemidae
Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Citharoides macrolepis 82 Citharidae
Scopthalmus aguosus 205 Scopthalmidae
Scopthalmus aguosus 206 Scopthalmidae
Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes haemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Cynoglossus interruptus 85 Cynoglossidae
Symphurus africauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Bothus lunatus 70 Bothidae
Bothus pantherinus 77 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta lugubris 79 Bothidae
Chascanopsetta prognatus 80 Bothidae
Asterorhombus cocosensis 79 Bothidae
Psettina fossaria 74 Bothidae
Grossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Trinectes maculata 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 98 Achiridae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Etropus crossotus 86 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Pseudopleuronectes herzensteini 28 Pleuronectidae
Polycentropsis abbreviata 202 Nandidae
Xiphias gladius (cf. Makaira sp.) 198 Xiphiidae
Tetrapturus albidus 157 Istiophoridae
Istiophorus platypterus 158 Istiophoridae
Makaira nigricans 159 Istiophoridae
Toxotes jaculatrix 199 Toxotidae
Nemastius pectoralis 165 Nemastiidae
Sphyraena argentea 190 Sphyraenidae
Sphyraena barracuda 191 Sphyraenidae
Echeneis naucrates 147 Echeneidae
Remora osteochir 146 Echeneidae
Coryphaena hippurus 144 Coryphaenidae
Coryphaena hippurus 145 Coryphaenidae
Rachycentron canadum 173 Rachycentridae
Rachycentron canadum 174 Rachycentridae
Scomberoides commersonnianus 189 Carangidae
Scomberoides lysan 186 Carangidae
Trachinotus baillonii 128 Carangidae
Trachinotus blochii 126 Carangidae
Elagatis bipinnulata 127 Carangidae
Seriola fasciata 119 Carangidae
Seriola rivoliana 120 Carangidae
Carangoides ferdau 130 Carangidae
Carangoides plagiotaenia 131 Carangidae
Parastromateus niger 138 Carangidae
Uraspis secunda 121 Carangidae
Selene brevoortii 135 Carangidae
Selene browni 125 Carangidae
Alectis ciliaris 118 Carangidae
Alectis indicus 136 Carangidae
Atropus atropus 137 Carangidae
Hemicaranx amblyrhynchus 123 Carangidae
Gnathanodon speciosus 132 Carangidae
Chloroscombrus chrysurus 122 Carangidae
Caranx caninus 134 Carangidae
Caranx heberi 129 Carangidae
Selar crumenophthalmus 124 Carangidae
Decapterus punctatus 117 Carangidae
Decapterus macarellus 115 Carangidae
Trachurus symmetricus 116 Carangidae
Trachurus trachurus 133 Carangidae
Lobotes surinamensis 162 Lobotidae
Lates sp. 160 Latidae
Lates sp. 161 Latidae
Mene maculatus 163 Menidae
Polycentrus schomburgkii 201 Polycentridae
Parachanna insignis 142 Channidae
Parachanna sp. 143 Channidae
Microctenopoma milleri 111 Anabantidae
Microctenopoma sp. 112 Anabantidae
Belontia hasselti 200 Osphronemidae
Trichogaster trichopterus 168 Osphronemidae
Betta ibanorum 166 Osphronemidae
Betta ibanorum 167 Osphronemidae
Bathyclupea argentea 113 Bathyclupeidae
Oligoplites saurus 114 Carangidae
Monodactylus argenteus 164 Monodactylidae
Inermia vittata 156 Inermiidae
Erythrocles monodi 149 Emmelichthyidae
Erythrocles schlegelii 148 Emmelichthyidae
Pomatomus saltatrix 171 Pomatomidae
Pomatomus saltatrix 172 Pomatomidae
Acanthocybium solandri 184 Scombridae
Gymnosarda unicolor 188 Scombridae
Scomber japonicus 178 Scombridae
Scomber scombrus 177 Scombridae
Scomberomorus cavalla 176 Scombridae
Scomberomorus maculatus 179 Scombridae
Sarda sarda 175 Scombridae
Thunnus atlanticus 183 Scombridae
Auxis rochei rochei 180 Scombridae
Katsuwonus pelamis 181 Scombridae
Euthynnus affinis 187 Scombridae
Euthynnus alletteratus 185 Scombridae
Caristius sp. 139 Carangidae
Trichiurus lepturus 192 Trichiuridae
Aphanopus carbo 193 Trichiuridae
Benthodesmus tenuis 194 Trichiuridae
Benthodesmus simonyi 195 Trichiuridae
Lepidopus altifrons 196 Trichiuridae
Ruvettus preciosus 150 Gempylidae
Neoepinnula americana 152 Gempylidae
Gempylus serpens 154 Gempylidae
Nesiarchus nasutus 153 Gempylidae
Nealotus tripes 155 Gempylidae
Paradiplospinus antarcticus 151 Gempylidae
Proportion of G+C
0.48362120
0.49400699
0.50439277
0.51477855
0.52516434
0.53555012
0.54593590
0.55632169
0.56670747
0.58747904

Figure 3-17. Saturation plots for third codon position of each gene. Transitions (blue Xs) and transversions (green triangles) are plotted against F84 distances. Black vertical line represents the level 1 change-point, treated as threshold pairwise distance (TPWD) for saturation.


rho-3
TPWD = 0.2454
0.33
0.27
0.22
0.16
0.11
0.05
0.00
0.00
0.37
0.74
1.11
1.49
1.86
2.23
chst2-3
TPWD = 0.3308
0.28
0.23
0.19
0.14
0.09
0.05
0.00
0.00
0.13
0.26
0.38
0.51
0.64
0.77
rnf213-3
TPWD = 0.2977
0.33
0.27
0.22
0.16
0.11
0.05
0.00
0.00
0.15
0.30
0.44
0.59
0.73
0.88
tmem22-3
TPWD = 0.2340
0.27
0.22
0.18
0.13
0.09
0.04
0.00
0.00
0.12
0.23
0.35
0.46
0.58
0.70

are available as one .xlsx file with four sheets in the American Museum of Natural History Dissertation Database.

Figure 3-18 confirms the saturation of third codon position transitions as a potential source of bias, as the distances between *Psettodes erumei* and the other (blue) flatfishes are well beyond the TPWD for *rho*-3 and *chst2*-3, and below that threshold distance from the (red) outgroups. The opposite pattern can be seen in the plots derived from the *tmem22*-3 and *rnf213* data, suggesting that the close relationship between *Psettodes* and other flatfishes observed in those gene trees was inferred with unsaturated data and that the inability of *rho* and *chst2* to recover *Psettodes* with the pleuronectoids was caused, at least in part, by randomized signal due to saturation at the third codon position of those genes.

DISCUSSION

4.1 NEW SINGLE EXON MARKERS FOR PHYLOGENIC INFERENCE

The results of the alignment of *socs6* indicate that the SEGs listed by Tine et al. (2011) should not be assumed to only contain one exon in all fishes even if they are SEGs in most annotated fish genomes. The structure of those genes may vary widely or be highly conserved across Teleostei and should be vetted as sequences that can be aligned with minimal ambiguity before including them in any phylogenetic analyses. In contrast, both the *tmem22* and *chst2* alignments, show that these markers are most likely SEGs for many more percomorph groups. Moreover, they 1) can be sequenced for a diverse array of taxa, 2) supply a maximal amount of data per sequencing reaction, 3) are entirely protein coding and conserved for carangimorphs so alignment is unambiguous 4) provide a large number of informative sites and 5) are no more prone to systematic bias than *rho*

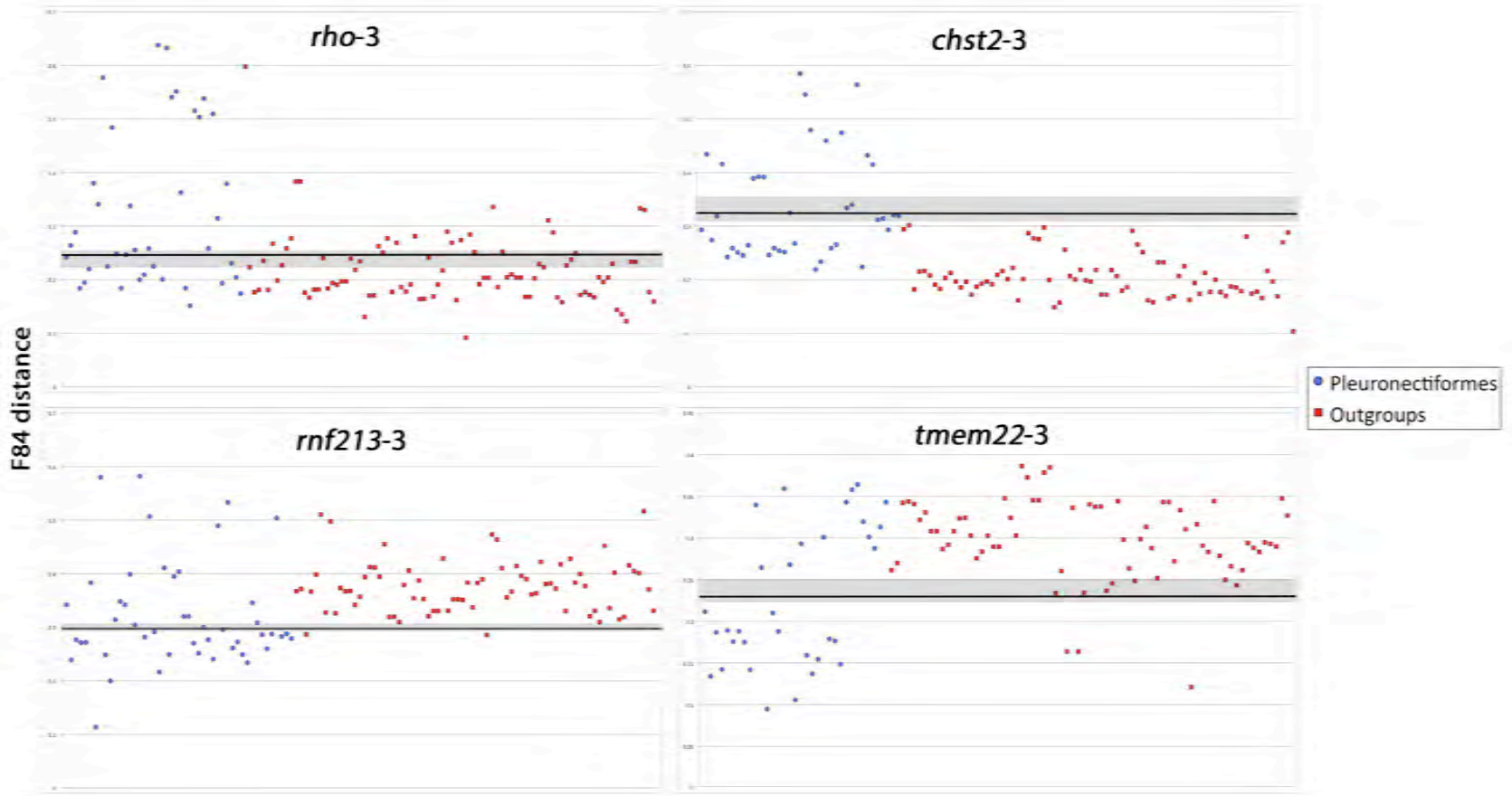


Figure 3-18. Plot of the F84 distance (y-axis) of all taxa (x-axis: blue are pleuronectiform taxa, red are outgroups) from *Psettodes erumei*. The black horizontal line is the level 1 change-point, or threshold pairwise distance (TPWD), calculated for transitions at each third codon position. The gray box represents the 95% confidence interval.

of *rnf213*. These attributes suggest they are useful markers for similar level phylogenetic studies of percomorph lineages.

The addition of the new SEGs in this analysis has made it possible to recover a highly supported monophyletic Pleuronectiformes, with Psettoidei sister to Pleuronectoidei (recovered by both Chapleau [1993] and Berendzen and Dimmick [2002]). These data yielded that result regardless of which partitioning scheme was used and without any manipulation of the data, however, this topology is only recovered when all four markers are analyzed together; none of the ML gene trees, two-gene trees (including the *tmemm22*/*chst2*, the "SEGs" tree) or the three-gene trees are able to recover a monophyletic Pleuronectiformes—it is the analysis of all the data combined, not the SEGs alone, that result in high support at that node.

4.2 Resolving Flatfish Monophyly With Molecular Data: Why Has it Been so Difficult?

When using multiple genetic markers to infer phylogenies it is expected that gene tree discordance will occur. Ideally, the large amount of data will ameliorate any confounding discordance, and mitigate effects of the unique histories of different genes or any phylogenic noise that may be present in a gene or partition (Baker and Desalle, 1997; Rokas et al., 2003; Phillipe et al., 2011; Betancur-R et al., 2013b). Unfortunately, it is not possible to predict how many genes or which genes will provide the most robust phylogenetic hypothesis beforehand, as this has only been accomplished by simulation studies (Wortley et al., 2005). Once genes have been sequenced, it is possible and desirable to investigate any causes for low statistical support at nodes, incongruence, possible long-branch attraction and others, all of which may be correlated and/or share a

causal relationship. This is especially true when results are obtained that are incongruent with those obtained from other independent sets of data, as is the case with flatfish monophyly.

Such data analyses were lacking in the Betancur-R et al., 2013a ("EToL") study that was unable to resolve the base of Carangimorpha, including monophyly and placement of flatfishes. Based on the results of analyses of the trimmed EToL dataset, missing data appears to have contributed to poor and misleading resolution in the hypothesis. This explanation is supported by the results of the topology tests (Table 3-3), where the KH and AU tests failed to reject the null that all topologies for comparison are equally good explanations of the data when comparing the EToL data to its own ML topology.

Additionally, when the EToL data was compared to the Con-EToL topology all tests rejected the null, meaning that the addition of the four genes with superior coverage increased the likelihood of the Con-EToL topology (4 plus 20 genes) so much so that the 20-gene tree was no longer the best explanation for the data. The number of genes was only increased by 16.7% yet the additional coverage provided by the Con dataset contributed a signal strong enough to significantly decrease the likelihood of the EToL ML topology that was inferred with 20 (versus four) loci. Furthermore, the opposite pattern is seen when comparing the Con data to all other topologies. Only in the SH test comparing the Con data and the Con-EToL topology (of which it is contributing a strong signal) was the P-value high enough to fail to reject the null, but this value is only 0.0195 greater than the threshold to reject (0.05) and is only recovered with the most conservative test.

The high clade support for the EToL topology and the significant increase in likelihood when those data are used to explain the Con tree, is consistent with the results of Roure et al. (2013) who found that a large proportion of missing data in a large matrix can cause topological incongruence due to disparate estimates of model parameters, which directly effects the likelihood score. This is also consistent with Wiens's (2006) description of abundant missing data masking the "good" signal provided by complete sites.

The EToL study was especially broad in scope, designed to test phylogenetic hypotheses pertaining to most teleost diversity, which flatfishes and their relatives are nested well within. Betancur-R et al. (2013b) were the first to test monophyly by increasing sampling within flatfishes, limiting outgroups for comparison to closely related taxa and sequencing 20 loci with little missing data. That study design was intended to resolve the issue that was apparent in larger studies, namely the placement of *Psettodes*, but was unable to. Even though the new data and analyses described here recovered a monophyletic Pleuronectiformes with high support, many of the same issues described by Betancur-R et al. (2013b) were still identified. As expected, there was gene tree discordance: none of the four ML gene trees were able to recover monophyly and all exhibited overall low bootstrap support. In some cases this could be explained by base compositional bias, but substitution saturation may also be the cause.

The results of the GC mapping and change point analysis show that saturation is a better explanation for the "rogue" nature of *Psettodes*—a taxon whose phylogenetic position has been difficult to resolve using molecular data (Betancur-R et al., 2013a,b; Near et al., 2013; and others). Because substitution saturation appears to be uncorrelated

(or weakly correlated) to base compositional bias and there is a trend in phylogenetics towards attempting to resolve deeper and deeper nodes using more and more loci and even whole genomes, substitution saturation needs to be revisited and studies designed to accurately asses its potential deleterious effects on phylogeny estimation, as well as possible ways to dampen those effects, are not only warranted, but necessary.

Finally, while missing data, base compositional bias and substitution saturation may explain topological discordance, both these results and those of Betancur-R et al. (2013b) show that the addition of data improves stationarity, validating the common practice of analyzing multiple markers simultaneously, and adding additional resolution to this very difficult phylogenetic question.

4.3 Resolution of the Sister Group to Pleuronectiformes

The results of the phylogenetic analyses of the concatenated alignment generated for this study are the first time monophyly of the order has been recovered without the removal of data and it is also novel in that it was also recovered with such high statistical support. This, along with the results of Chapleau (1993) and Betancur-R et al. (2013b) is convincing evidence for monophyly. Unfortunately, even though so many outgroups were included, all of which are representative of taxa that have consistently been recovered with or near pleuronectiforms (Li et al., 2009; Betancur-R et al., 2013a; Near et al., 2013) the support at the base of the Con tree is low and when all nodes with values less than 70 are collapsed (Fig. 3-5) the lack of resolution regarding sister-group hypotheses becomes apparent.

Much like the tendency of *Psettodes* to exhibit variable placement in the percomorph tree of life given analysis of each unique dataset, many of the carangimorph

taxa whose exact position is unresolved here have been known to do the same. In particular, Nematistiidae, Sphyraenidae, Toxotidae, Centropomidae, Polynemidae and Istiophoriformes (Xiphiidae + Istiophoridae) are variably placed in different hypotheses. Most of these taxa share something else in common with *Psettodes*: low extant taxonomic diversity, including one or a few genera, usually with only a few species. Sphyraenidae is comprised of only one genus (27 species) as is Toxtotidae (10 species), and Centropomidae (12 species). Istiophoriformes only has five genera (13 species) and Nematistiidae (one species) has the lowest taxonomic diversity of all.

An asymmetrical topology, with a mix of low and high diversity taxa, demands more sequence data to recover it and has been shown to be more sensitive to artifacts such as saturation (Xia, 2003). This may be the reason recovering *Psettodes* (one genus, three species) with the sister Pleuronectoidei (~700 species) has been so difficult. If asymmetrical amounts of diversity among sister groups make it difficult to recover phylogeny using sequence data, than it is worthwhile to test for this within Carangimorpha assuming one of the groups mentioned above is sister to Pleuronectiformes rather than a larger group such as Carangoidei. Low extant diversity may have an effect on our ability to recover relationships of these lineages with high support, and this may not only be a problem resolving the base of Carangimorpha. Nelson (2006) noted that the former "Perciformes" (representing ~10,000 percomorph species) is comprised of 160 families, 52 of which only include only one genus and 23 of which are monotypic, constituting 46% of families. Because resolving relationships within the major percomorph clades is usually contingent upon resolving interfamilial relationships, this statistic may be indicative of a taxonomic asymmetry that is widespread and may

effect our ability to recover relationships among deep nodes. In these cases, morphological information, including fossil, larval, and adult characters, could provide evidence that can help resolve the phylogenetic placement of Pleuronectiformes. Furthermore, optimizing new molecular markers, adding data to existing matrices with poor coverage, and reexamining the possible role nucleotide substitution saturation may play in the poor estimation of these phylogenies and how we can dampen its deleterious effects, can all help resolve the base of the carangimorph tree, allowing for meaningful comparisons that will inform our understanding of the evolution of many aspects of the biology of these fishes, most notably the evolution of the unique and truly amazing flatfish metamorphosis and bilateral asymmetry.

Ahlstrom, E. H., K. Amaoka, D. A. Hensley, H. G. Moser, B. Y. Sumida, and others. 1984. “Pleuronectiformes: Development.” *Ontogeny and Systematics of Fishes* 1: 640–70.

Amaoka, Kunio. 1969. “Studies on the Sinistral Flounders Found in the Waters around Japan. Taxonomy, Anatomy and Phylogeny.” *J. Shimonoseki Univ. Fish.* 18: 65–340.

Arratia, Gloria, Adriana López-Arbarello, Guntupalli VR Prasad, Varun Parmar, and Jürgen Kriwet. 2004. “Late Cretaceous–Paleocene Percomorphs (Teleostei) from India—early Radiation of Perciformes.” *Recent Advances in the Origin and Early Radiation of Vertebrates*, 635–63.

Baker, Richard H., and Rob DeSalle. 1997. “Multiple Sources of Character Information and the Phylogeny of Hawaiian Drosophilids.” *Systematic Biology* 46 (4): 654–73.

Berendzen, Peter B., and Walter Wheaton Dimmick. 2002. “Phylogenetic Relationships of Pleuronectiformes Based on Molecular Evidence.” *Journal Information* 2002 (3).

Bergsten, Johannes. 2005. “A Review of Long-Branch Attraction.” *Cladistics* 21 (2): 163–93.

Betancur-R., Ricardo, Richard E. Broughton, Edward O. Wiley, Kent Carpenter, J. Andres Lopez, Chenhong Li, Nancy I. Holcroft, et al. 2013a. “The Tree of Life and a New Classification of Bony Fishes.” *PLoS Currents* 5 (April). doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, Ricardo, Chenhong Li, Thomas A. Munroe, Jesus A. Ballesteros, and Guillermo Ortí. 2013b. “Addressing Gene Tree Discordance and Non-Stationarity to Resolve a Multi-Locus Phylogeny of the Flatfishes (Teleostei: Pleuronectiformes).” *Systematic Biology*, syt039.

Brower, A. V. Z., R. DeSalle, and A. Vogler. 1996. “Gene Trees, Species Trees, and Systematics: A Cladistic Perspective.” *Annual Review of Ecology and Systematics*, 423–50.

Chabanaud, P. 1949. “Le Probleme de La Phylogenese Des Heterosomata.” *Bull. Inst. Oceanogr.*, no. 950: 1–24.

Chapleau, François. 1993. “Pleuronectiform Relationships: A Cladistic Reassessment.” *Bulletin of Marine Science* 52 (1): 516–40.

Chen, Wei-Jen, Céline Bonillo, and Guillaume Lecointre. 2003. “Repeatability of Clades as a Criterion of Reliability: A Case Study for Molecular Phylogeny of Acanthomorpha (Teleostei) with Larger Number of Taxa.” *Molecular Phylogenetics and Evolution* 26 (2): 262–88. doi:10.1016/S1055-7903(02)00371-8.

Cusimano, Natalie, and Susanne S. Renner. 2010. “Slowdowns in Diversification Rates from Real Phylogenies May Not Be Real.” *Systematic Biology* 59 (4): 458–64.

Dávalos, Liliana M., and Susan L. Perkins. 2008. “Saturation and Base Composition Bias Explain Phylogenomic Conflict in Plasmodium.” *Genomics* 91 (5): 433–42. doi:10.1016/j.ygeno.2008.01.006.

Degnan, James H., and Noah A. Rosenberg. 2006. “Discordance of Species Trees with Their Most Likely Gene Trees.” *PLoS Genetics* 2 (5): e68.

Dettai, Agnès, and Guillaume Lecointre. 2005. “Further Support for the Clades Obtained by Multiple Molecular Phylogenies in the Acanthomorph Bush.” *Comptes Rendus Biologies* 328 (7): 674–89. doi:10.1016/j.crvi.2005.04.002.

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, J. Heled, M. Kearse, et al. 2010. *Geneious v5. 1. 2010*.

Felsenstein, Joseph. 1978. “Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading.” *Systematic Zoology* 27 (4): 401. doi:10.2307/2412923.

Friedman, Matt. 2008. “The Evolutionary Origin of Flatfish Asymmetry.” *Nature* 454 (7201): 209–12.

———. 2012. “Osteology Of† Heteronectes Chaneti (Acanthomorpha, Pleuronectiformes), an Eocene Stem Flatfish, with a Discussion of Flatfish Sister-Group Relationships.” *Journal of Vertebrate Paleontology* 32 (4): 735–56.

Gosline, William A., and Vernon E. Brock. 1960. *Handbook of Hawaiian Fishes*. University of Hawaii Press Honolulu. http://library.wur.nl/WebQuery/clc/533791.

Graybeal, Anna. 1998. “Is It Better to Add Taxa or Characters to a Difficult Phylogenetic Problem?” *Systematic Biology* 47 (1): 9–17. doi:10.1080/106351598260996.

Haaker, Peter L. “The Biology of the California Halibut, Paralichthys Californicus (Ayres) in Anaheim Bay.” Calisphere a website by the University of California. *The Biology of the California Halibut, Paralichthys Californicus (Ayres) in Anaheim Bay*. http://content.cdlib.org/view?docId=kt6n39n885&chunk.id=d0e5205&brand=calisphere&doc.view=entire_text.

Heled, Joseph, and Alexei J. Drummond. 2010. “Bayesian Inference of Species Trees from Multilocus Data.” *Molecular Biology and Evolution* 27 (3): 570–80. doi:10.1093/molbev/msp274.

Hillis, David M., David D. Pollock, Jimmy A. McGuire, and Derrick J. Zwickl. 2003. “Is Sparse Taxon Sampling a Problem for Phylogenetic Inference?” *Systematic Biology* 52 (1): 124–26. doi:10.1080/10635150390132911.

Huang, Huateng, Qixin He, Laura S. Kubatko, and L. Lacey Knowles. 2010. “Sources of Error Inherent in Species-Tree Estimation: Impact of Mutational and Coalescent Effects on Accuracy and Implications for Choosing among Different Methods.” *Systematic Biology* 59 (5): 573–83.

Jobb, G. 2008. “TREEFINDER Version of October 2008.” *Distributed by the Author at Www. Treefinder. de*.

Katoh, K., G. Asimenos, and H. Toh. 2009. “Multiple Alignment of DNA Sequences with MAFFT.” *Methods Mol Biol* 537: 39–64.

Kelly, Neil, and Gene Kira. 1993. *The Baja Catch: A Fishing, Travel & Remote Camping Manual for Baja California*. Apples & Oranges, Incorporated.

Kishino, Hirohisa, and Masami Hasegawa. 1989. “Evaluation of the Maximum Likelihood Estimate of the Evolutionary Tree Topologies from DNA Sequence Data, and the Branching Order in Hominoidea.” *Journal of Molecular Evolution* 29 (2): 170–79.

Larmuseau, Maarten HD, Tine Huyse, Kim Vancampenhout, Jeroen KJ Van Houdt, and Filip AM Volckaert. 2010. “High Molecular Diversity in the Rhodopsin Gene in

Closely Related Goby Fishes: A Role for Visual Pigments in Adaptive Speciation?" *Molecular Phylogenetics and Evolution* 55 (2): 689–98.

Li, B., A. Dettaī, C. Cruaud, A. Couloux, M. Desoutter-Meniger, and G. Lecointre. 2009. "RNF213, a New Nuclear Marker for Acanthomorph Phylogeny." *Molecular Phylogenetics and Evolution* 50 (2): 345–63.

Little, A. G., S. C. Lougheed, and C. D. Moyes. 2010. "Evolutionary Affinity of Billfishes (Xiphiidae and Istiophoridae) and Flatfishes (Plueronectiformes): Independent and Trans-Subordinal Origins of Endothermy in Teleost Fishes." *Molecular Phylogenetics and Evolution* 56 (3): 897–904.

Maddison, Wayne P. 1997. "Gene Trees in Species Trees." *Systematic Biology* 46 (3): 523–36.

Maddison, W. P., and D. R. Maddison. 2011. *Mesquite 2.75: A Modular System for Evolutionary Analysis*.

Milinkovitch, Michel C., and James Lyons-Weiler. 1998. "Finding Optimal Ingroup Topologies and Convexities When the Choice of Outgroups Is Not Obvious." *Molecular Phylogenetics and Evolution* 9 (3): 348–57.

Near, Thomas J., Alex Dornburg, Ron I. Eytan, Benjamin P. Keck, W. Leo Smith, Kristen L. Kuhn, Jon A. Moore, et al. 2013. "Phylogeny and Tempo of Diversification in the Superradiation of Spiny-Rayed Fishes." *Proceedings of the National Academy of Sciences* 110 (31): 12738–43.

Nelson, Joseph S. 2006. *Fishes of the World*. John Wiley & Sons. Nichols, Richard. 2001. "Gene Trees and Species Trees Are Not the Same." *Trends in Ecology & Evolution* 16 (7): 358–64.

Nosenko, Tetyana, Fabian Schreiber, Maja Adamska, Marcin Adamski, Michael Eitel, Jörg Hammel, Manuel Maldonado, et al. 2013. "Deep Metazoan Phylogeny: When Different Genes Tell Different Stories." *Molecular Phylogenetics and Evolution* 67 (1): 223–33. doi:10.1016/j.ympev.2013.01.010.

Ortí, Guillermo, and others. 2014. "Molecular Evidence for the Monophyly of Flatfishes (Carangimorpharia: Pleuronectiformes)." *Molecular Phylogenetics and Evolution* 73: 18–22.

Pamilo, Pekka, and Masatoshi Nei. 1988. “Relationships between Gene Trees and Species Trees.” *Molecular Biology and Evolution* 5 (5): 568–83.

Patterson, Colin. 1993. “An Overview of the Early Fossil Record of Acanthomorphs.” *Bulletin of Marine Science* 52 (1): 29–59.

Philippe, Hervé, Henner Brinkmann, Dennis V. Lavrov, D. Timothy J. Littlewood, Michael Manuel, Gert Wörheide, and Denis Baurain. 2011. “Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough.” *PLoS Biology* 9 (3): e1000602.

Pick, K. S., H. Philippe, F. Schreiber, D. Erpenbeck, D. J. Jackson, P. Wrede, M. Wiens, et al. 2010. “Improved Phylogenomic Taxon Sampling Noticeably Affects Nonbilaterian Relationships.” *Molecular Biology and Evolution* 27 (9): 1983–87. doi:10.1093/molbev/msq089.

Platt, Adam R., Robert W. Woodhall, and Alfred L. George. 2007. “Improved DNA Sequencing Quality and Efficiency Using an Optimized Fast Cycle Sequencing Protocol.” *Biotechniques* 43 (1): 58.

Poe, Steven, and David L. Swofford. 1999. “Taxon Sampling Revisited.” *Nature* 398 (6725): 299–300. doi:10.1038/18592.

Rannala, Bruce. 2002. “Identifiability of Parameters in MCMC Bayesian Inference of Phylogeny.” *Systematic Biology* 51 (5): 754–60.

Revell, Liam, Luke Harmon, and Richard Glor. 2005. “Under-Parameterized Model of Sequence Evolution Leads to Bias in the Estimation of Diversification Rates from Molecular Phylogenies.” *Systematic Biology* 54 (6): 973–83. doi:10.1080/10635150500354647.

Roje, Dawn M. 2014. “Evaluating the Effects of Non-Neutral Molecular Markers on Phylogeny Inference.” *PloS One* 9 (2): e87428.

Rokas, Antonis, Barry L. Williams, Nicole King, and Sean B. Carroll. 2003. “Genome-Scale Approaches to Resolving Incongruence in Molecular Phylogenies.” *Nature* 425 (6960): 798–804.

Rosenberg, Michael S., and Sudhir Kumar. 2003. “Taxon Sampling, Bioinformatics, and Phylogenomics.” *Systematic Biology* 52 (1): 119–24.

Roure, Béatrice, Denis Baurain, and Hervé Philippe. 2013. “Impact of Missing Data on Phylogenies Inferred from Empirical Phylogenomic Data Sets.” *Molecular Biology and Evolution* 30 (1): 197–214. doi:10.1093/molbev/mss208.

Shimodaira, Hidetoshi. 2002. “An Approximately Unbiased Test of Phylogenetic Tree Selection.” *Systematic Biology* 51 (3): 492–508. doi:10.1080/10635150290069913.

Shimodaira, Hidetoshi, and Masami Hasegawa. 1999. “Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference.” *Molecular Biology and Evolution* 16: 1114–16.

Smith, William Leo, and Ward C. Wheeler. 2006. “Venom Evolution Widespread in Fishes: A Phylogenetic Road Map for the Bioprospecting of Piscine Venoms.” *Journal of Heredity* 97 (3): 206–17.

Smith, W.L., and M.T. Craig. 2007. “Casting the Percomorph Net Widely: The Importance of Broad Taxonomic Sampling in the Search for the Placement of Serranid and Percid Fishes.” *Copeia* 2007 (1).

Sullivan, Jack, and Paul Joyce. 2005. “Model Selection in Phylogenetics.” *Annual Review of Ecology, Evolution, and Systematics*, 445–66.

Swofford, David L., Gary J. Olsen, Peter J. Waddell, David M. Hillis, D. M. Hillis, C. Moritz, and B. K. Mable. 1996. “Molecular Systematics.” *Molecular Systematics*.

Tarrío, Rosa, Francisco Rodríguez-Trelles, and Francisco J. Ayala. 2000. “Tree Rooting with Outgroups When They Differ in Their Nucleotide Composition from the Ingroup” *Molecular Phylogenetics and Evolution* 16 (3): 344–49.

Taylor, W.A. 2000. *Change-Point Analyzer Shareware Program* (version 2.3). Libertyville, IL: Taylor Enterprises. http://www.variation.com/cpa/.

———. “Change-Point Analysis: A Powerful New Tool for Detecting Changes.” http://www.variation.com/cpa/tech/changepoint.html.

Tine, M., H. Kuhl, A. Beck, L. Bargelloni, and R. Reinhardt. 2011. “Comparative Analysis of Intronless Genes in Teleost Fish Genomes: Insights into Their Evolution and Molecular Function.” *Marine Genomics*.

Tornabene, Luke, Yongjiu Chen, and Frank Pezold. 2013. “Gobies Are Deeply Divided: Phylogenetic Evidence from Nuclear DNA (Teleostei: Gobioidei: Gobiidae).” *Systematics and Biodiversity* 11 (3): 345–61.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C. Faircloth, Maido Remm, and Steven G. Rozen. 2012. “Primer3—new Capabilities and Interfaces.” *Nucleic Acids Research* 40 (15): e115–e115.

Wiens, John J. 2003. “Missing Data, Incomplete Taxa, and Phylogenetic Accuracy.” *Systematic Biology* 52 (4): 528–38.

———. 2005. “Can Incomplete Taxa Rescue Phylogenetic Analyses from Long-Branch Attraction?” *Systematic Biology* 54 (5): 731–42.

———. 2006. “Missing Data and the Design of Phylogenetic Analyses.” *Journal of Biomedical Informatics* 39 (1): 34–42.

Wortley, Alexandra H., Paula J. Rudall, David J. Harris, and Robert W. Scotland. 2005. “How Much Data Are Needed to Resolve a Difficult Phylogeny? Case Study in Lamiales.” *Systematic Biology* 54 (5): 697–709. doi:10.1080/10635150500221028.

Xia, Xuhua, Zheng Xie, Marco Salemi, Lu Chen, and Yong Wang. 2003. “An Index of Substitution Saturation and Its Application.” *Molecular Phylogenetics and Evolution* 26 (1): 1–7.

Xia, X., and Z. Xie. 2001. “DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution.” *Journal of Heredity* 92 (4): 371–73.

CHAPTER IV

MORPHOLOGY AND MOLECULES: RESOLVING PLEURONECTIFORM RELATIONSHIPS

ABSTRACT

Adult flatfishes (order Pleuronectiformes) are asymmetrical fishes that are generally benthic; however, pretransformation larvae are symmetrical, swimming upright in the water column. A previous phylogenetic analysis, utilizing characters of adult morphology, recovered seven major lineages of Pleuronectiformes (Chapleau, 1993). In contrast, evidence from external larval morphology indicates an alternative hypothesis (Ahlstrom, 1984; Hensley and Ahlstrom, 1984). To resolve intraordinal relationships, I analyzed DNA sequences (totaling 3781 bp) for the protein-coding loci rho, rnf213, tmem22 and chst2 for 58 flatfishes and 90 putative outgroups using the ML optimality criterion. Additionally, I recoded a previously published matrix of characters of adult morphology and combined it with new larval characters to test whether life history is correlated with phylogenetic signal. Further, I assessed the accuracy of ML ancestral character state estimation (ACE) to determine whether or not morphological characters provide additional support for hypotheses of relationships among major pleuronectiform groups. These results suggest that larval characters should not be treated as a source of independent data, but do provide resolution and additional support for novel relationships within Pleuronectiformes, although they may be in violation of the condition of low rates on ML ACE.

INTRODUCTION

Flatfishes (Pleuronectiformes), the commercially important soles, flounders and halibuts, exhibit one of the most remarkable vertebrate metamorphoses, where one eye of a bilaterally symmetrical larva migrates to the opposite side of the cranium, resulting in highly asymmetrical juvenile and adult forms. Post-metamorphosis, the eyeless (blind) side exhibits either greatly reduced pigmentation, or lacks pigmentation entirely. Adult flatfishes generally occupy marine benthic habitats throughout all of the world's oceans, but have adopted a life history strategy that involves a pelagic larval stage. Because these planktonic larvae have evolved in a dissimilar environment from the adults and are bilaterally symmetrical, they exhibit a suite of physiologically, ecologically, behaviorally, and anatomically distinct characters. These seemingly bizarre morphological characters include, but are not limited to, elaborate spines of the head region, elongate fin rays, and melanophores organized in complex arrays. All (approximately) 700 species in 14 families (Nelson, 2004) share this life history—one that is extremely common for marine teleosts.

Independent origins of flatfish asymmetry (polyphyly) have been proposed (Chabanaud, 1949; Amaoka, 1969), but never evaluated in a phylogenetic context. Chapleau (1993) was the first to carry out a cladistic assessment of the order, but stated that until any derived character states between flatfishes and any other group(s) of symmetrical fishes could be indentified that it was unwarranted to regard them as polyphyletic and, moreover, provided the following three synapomorphies uniting them: 1) metamorphosis with eye migration leading to an asymmetrical cranium, 2) a dorsal fin that overlaps the neurocranium and 3) the presence of the recessus orbitalis (an accessory

organ of the eye).

Because all characters of adult flatfish morphology that Chapleau (1993) recognized are either autapomorphic or plesiomorphic, and therefore parsimony uninformative, he was not able to root his tree using an outgroup and instead used characters that he had ordered and polarized using *Psettodes*, which he deemed the pleisiomorphic flatfish because of the high number of pleisiomorphic character states it shares with other percoids (and not other flatfishes). He analyzed his matrix of 39 characters of adult morphology utilizing the parsimony criterion and the result was a consensus of 18 most parsimonious trees (Fig. 4-1).

That consensus tree was largely unresolved, but he was able to identify two major lineages (I and II) that he designated as the sister suborders: Psettoidei (*Psettodes*) and Pleuronectoidei (all other flatfish taxa). He recovered a tricotomy at the base of Pleuronectiodei with *Lepidoblepharon*, *Citharoides* and the rest (lineage III). That large group, lineage III, lacked structure at the base, but synapomoprhies were established uniting four lineages within it, with the most derived being the highly asymmetrical Soleidae, sister to the family with the highest degree of asymmetry, Cynoglossidae. Therefore, Chapleau (1993) concluded that an extremely asymmetrical morphology was the derived state for the order.

More recently, competing hypotheses of intraordinal relationships have been presented (Berendzen and Dimmick, 2002 and Betancur-R et al., 2013b) based on analyses of DNA sequence data using likelihood-based methods of phylogenetic inference. Both studies recovered the two suborders established by Chapleau (1993), but, unlike Chapleau (1993), they did not find a highly asymmetrical (adult) morphology as

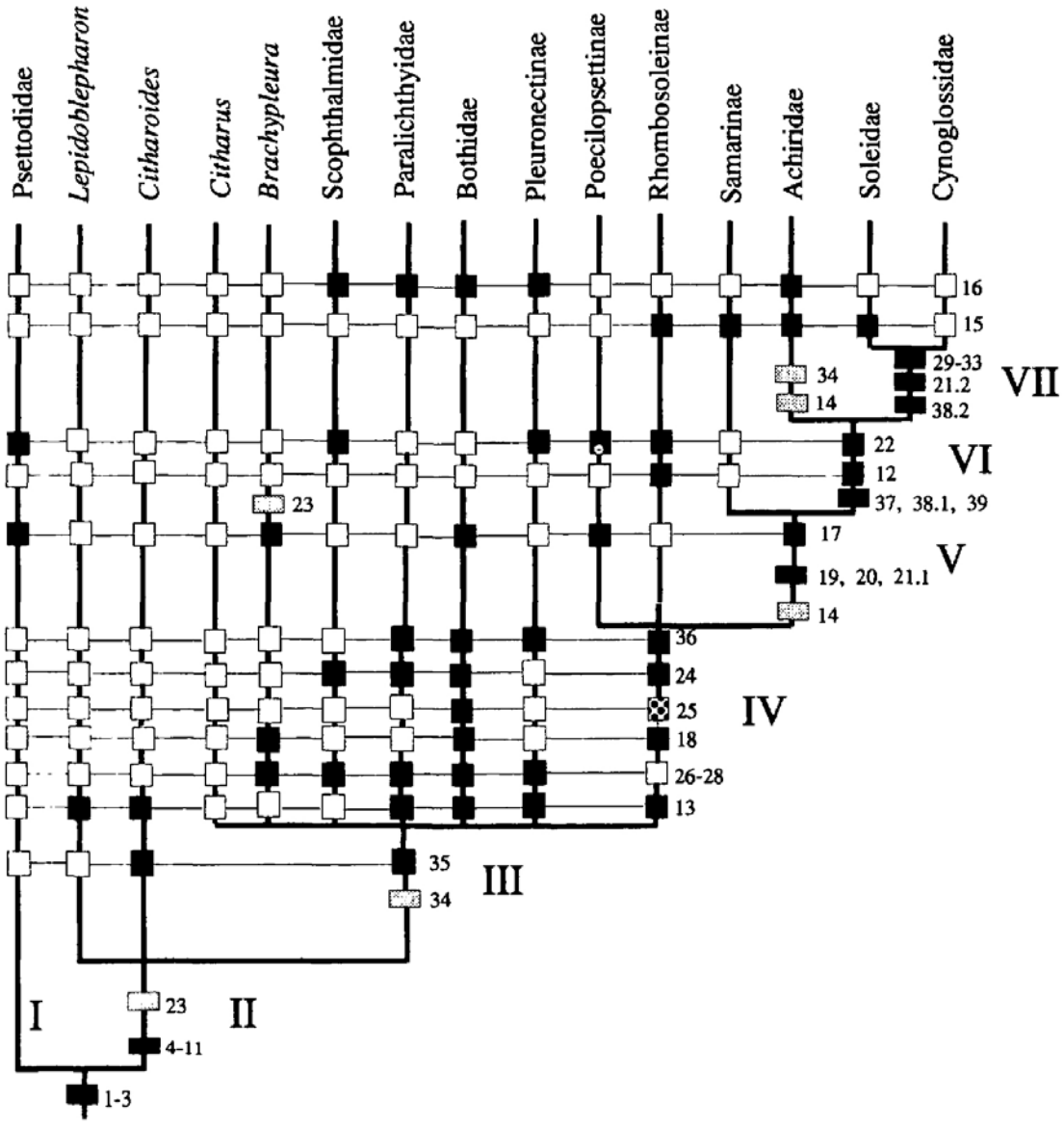


Figure 4-1. Consensus of 18 equally parsimonious trees for Pleuronectiformes; reproduced from Chapleau (1993), Figure 7. Black rectangles represent uniquely derived character states; shaded rectangles are derived character states with one reversal. Squares represent polarized and ordered character states: empty squares are plesiomorphic states, black squares are first apomorphic states and dotted squares are second apomorphic states. Roman numerals indicate lineages (I-VII) and decimals are order of apomorphic states.

being the derived state, but instead recovered pleuronectids and some paralichthyids (many of which have symmetrical jaws and all which maintain well developed blind-side pectoral fins) as the most derived taxon.

Even though Chapleau (1993) treated Paralichthyidae as monophyletic he admitted that was speculative as no synapomorphies of adult morphology united the family and similarity in larval morphology, identified by Hensley and Ahlstrom (1984), suggested an alternative hypothesis. Both Berendzen and Dimmick (2002) and Betancur-R et al. (2013b) confirmed that Paralichthyidae is not monophyletic, but were not consistent in the reconstruction of paralichthyid genera. Clearly there is some convergence, but interfamilial relationships—the deeper nodes—are not entirely resolved.

Ahlstrom (1984) and Hensley and Ahlstrom (1984) provided a thorough description of larval flatfishes and identified many similarities that could potentially inform their phylogeny and even though the use of larval characters in systematic ichthyology has been heralded at least in part for its assumed independence from adult characters (Cohen, 1984; Roje, 2010 and others), they have yet to be analyzed in a phylogenetic study of the order. Beyond flatfish phylogenetics, ichthyoplankton are often overlooked, probably because larvae are difficult to collect and work with, as they are small, delicate and go through multiple ontogenetic stages. Because both larval and adult morphology is well described for most flatfish taxa it follows that given a well supported phylogeny this would be an appropriate group to 1) test whether larval and adult characters are less likely to be correlated with each other than either would be to characters of their own life history stage and 2) investigate the ability of larval characters

to provide additional support for novel phylogenetic hypotheses.

Because of this and because there is still very little resolution regarding flatfish phylogeny (monophyly, sister group hypotheses and interrelationships), I analyzed new molecular markers along with previously established protein coding loci for a large number of flatfish taxa and putative outgroups. Additionally, I combine the morphological matrix published by Chapleau (1993) with larval characters and analyze them to see if their addition increases resolution in a phylogenetic analysis of morphological characters. I also use these data to test hypotheses regarding prior relationships as well as assumptions that larval characters are "independent" sets of data. Finally, I carry out ancestral character estimation (ACE) to investigate the presence of apomorphy of both the adult and larval characters and to evaluate if larval characters may help unite flatfish (pleisiomorphy) with other carangimorphs.

MATERIALS AND METHODS

2.1 TAXON SAMPLING

Because support for monophyly of the order has been lacking (Chen, 2003; Dettai and Lecointre, 2005; Smith and Craig, 2007; Li et al, 2009; Betancur-R et al., 2013a,b; Near et al., 2013) a large phylogeny including many flatfish and putative outgroups was generated to test monophyly, sister group hypotheses, and within-order relationships. The details as to outgroup sampling and marker selection are described in Chapter III (p. 21). As for taxon sampling within Pleuronectiformes, because Chapleau (1993) recognized three distinct paralichthyid groups: 1) a *Cyclopsetta* group (including *Cyclopsettta*, *Syacium*, *Citharichthys* and *Etropus*), 2) a *Pseudorhombus* group (including *Pseudorhombus*, *Tarhops* and *Cephalopsetta*) and 3) "the remaining Paralichthyidae" (p.

528) that I will refer to as the *Paralichthys* group—all of which have unknown relationships to the rest of the Pleuronectoidei—sampling within Paralichthyidae was designed to represent all three groups and is summarized as follows: two species of *Syacium*, one *Citharichthys*, one *Pseudorhombus*, one *Tarhops*, two *Etropus*, one *Hippoglossina*, one *Ancyclopsetta*, one *Paralichthys* and one *Xystreurys*.

Additionally, both suborders, Psettoidei (consisting only of the three species of *Psettode*s) and Pleuronectoidei (the remainder) were represented, as well as ten additional families all thought to be monophyletic (Chapleau, 1993; Berendzen and Dimmick, 2002; Betancur-R et al., 2013b). The breakdown of taxa sampled by family is as follows: Achiridae (four species, four genera), Bothidae (11 species, nine genera), Citharidae (two species, two genera), Cynoglossidae (two species, two genera), Paralichthyidae (12 species, nine genera), Pleuronectidae (14 species, 13 genera), Poecilopsettidae (one species), Psettodidae (one *Psettodes* sp. and one *P. erumei*), Samaridae (one species), Scopthalmidae (one species) and Soleidae (seven species, six genera). Therefore, the total taxonomic diversity sampled across Pleuronectiformes includes 57 species (58 individuals) representing both suborders, 11 families (including all three acknowledged paralichthyid groups) and 49 genera.

2.2 Morphological Character Matrix

Adult characters were taken from Chapleau (1993) and consisted of 39 characters most of which were osteological. Chapleau coded the characters mostly at the family level (except for *Lepidoblepharon*, *Citharus*, *Citharoides* and *Brachypleura*), so all Pleuronectidae were treated in one row of the matrix. For this study the families were not assumed to be monophyletic, therefore Chapleau's (1993) coding was extrapolated out to

all species Chapleau (1993) recognized as members of a certain family (or genus). Additionally, any characters that Chapleau (1993) described in the appendix as being "observed, but not encoded" (or equivalent) were coded here according to the observed distribution of the character. Because some of the downstream analyses carried out in this study require discrete, binary characters and Chapleau (1993) treated and analyzed his characters as being ordered and polarized, any states he coded as "2" (of 0, 1 or 2) were changed to "1". See Table 4-1 for a summary of the entire matrix, including the new coding of Chapleau's (1993) characters.

I included eleven larval characters, coded as binary characters according to their description in Ahlstrom (1984) and Hensley and Ahlstrom (1984). In addition to requiring binary coding, some downstream analyses also require no ambiguous coding or missing data, therefore, characters were chosen for which their distribution across the taxa sampled here was unambiguous in Hensley and Ahlstrom (1984). These characters include elongate anterior dorsal rays (EDR), preopercular spines (POSp), frontal spines (FrSp), size at transformation (SaT), elongate pelvic (sometimes referred to as ventral) rays (EPR), types of gut (G), otic spines (OtSp; coded per Roje (2010) for Pleuronectidae), urohyal spines (USp), basipterygial spines (BSp), cleithral spines (CSp) and body spines (BodySp). I coded all larval characters as present (1) or absent (0), except SaT (0 = $\leq$ 9mm, 1 = > 9 mm) and G (simple = 0, elongate or trailing = 1). These characters and their coding for the taxa sampled are presented in Table 4-1. Finally, the early-life history of the monotypic *Clidoderma* is unknown, therefore I coded all larval characters for that species according to the most common state observed for the rest of its family, Pleuronectidae.

Table 4-1. Morphological character matrix. Characters 1-39 correspond to Chapleau's (1993) characters of adult morphology; characters with an asterisk were coded as per the text in the corresponding appendix (i.e. when a character state was state was "observed, but not encoded" it was coded here). Additionally, all Chapleau's (1993) three-state characters were coded as 0 or 1(=2). Larval characters are: EDR (elongate dorsal rays), POSp (preopercular spines), FrSp (frontal spines), SaT (size at transformation; 0 = ≤ 9mm, 1 = > 9 mm), EPR (elongate pelvic rays), G (Gut; simple = 0, elongate or trailing = 1), OtSp (otic spines), USp (urohyal spines), BSp (basipterygial spines), CSp (cleithral spines) and BodySp (body spines). Larval characters were coded according to Hensley and Ahlstrom (1984), except OtSp, which was coded per Roje (2010) for Pleuronectidae. All larval characters, except SaT and G (see above) are coded as absent (0) or present (1). Highlighted characters corresponded to ACEs in Figures 4-8 to 4-15.

Character	1	2	3	4	5	6	7	8	9	10	11	12*	13	14*	15
Psettodes erumei 100 Psettodidae	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Psettodes sp. 101 Psettodidae	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Samariscus japonicus 83 Samaridae	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Samariscus xenicus 102 Samaridae	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Heteromycteris japonicus 105 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aesopia cornuta 109 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Soleichthys heterorhinos 108 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Zebrias japonica 106 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aseraggodes kobensis 104 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pardachirus pavoninus 203 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aseraggodes heemstrai 107 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Parachirus xenicus 69 Soleidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Citharoides macrolepis 82 Citharidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Scopthalmus aquosus 205 Scopthamlmidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Scopthalmus aquosus 206 Scopthamlmidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Trinectes maculata 66 Achiridae	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Achirus lineatus 67 Achiridae	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Gymnachirus melas 68 Achiridae	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Syacium micrurum 96 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Syacium papillosum 94 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Citharichthys xanthostigma 87 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Etropus crossotus 98 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Etropus microstomus 91 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Cynoglossus interuptus 84 Cynoglossidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Symphurus atricauda 84 Cynoglossidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Engyophrys senta 76 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Bothus pantherinus 77 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Bothus lunatus 70 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Engyprosopon cf. multisquama 72 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Laeops kitaharae 73 Bothidae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Chascanopsetta lugubris 79 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Chascanopsetta prognahus 80 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Asterorhombus cocosensis 78 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Psettina tosana 74 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Crossorhombus kobensis 71 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Tosarhombus octoculatus 204 Bothidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Hippoglossina oblonga 89 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Tarphops oligolepis 93 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Ancylopsetta ommata 86 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Paralichthys dentatus 90 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Ancylopsetta dilecta 95 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Xystreurys liolepis 88 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Poecilopsetta plinthus 99 Poecilopsettidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Pseudorhombus pentophthalmus 92 Paralichthyidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Atheresthes evermanni 1 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Atheresthes stomias 2 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Eopsetta jordani 12 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Pleuronichthys coenosus 35 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Glyptocephalus zachirus 14 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Pleuronectes pinnifasciatus 44 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Hippoglossus stenolepis 6 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Verasper moseri 56 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Cleisthenes herzensteini 62 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Hippoglossoides elassodon 31 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Platichthys stellatus 32 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Lepidopsetta bilineata 20 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Parophrys vetulus 30 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Psettichthys melanostictus 38 Pleuronectidae	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0

Character (cont.)	16	17	18*	19	20	21	22*	23*	24	25*	26*	27	28*	29*	30	31*	32
Psettodes erumei 100 Psettodidae	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Psettodes sp. 101 Psettodidae	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Samariscus japonicus 83 Samaridae	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0
Samariscus xenicus 102 Samaridae	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0
Heteromycteris japonicus 105 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Aesopia cornuta 109 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Soleichthys heterorhinos 108 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Zebrias japonica 106 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Aseraggodes kobensis 104 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Pardachirus pavoninus 203 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1
Aseraggodes heemstrai 107 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Parachirus xenicus 69 Soleidae	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
Citharoides macrolepis 82 Citharidae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Scopthalmus aquosus 205 Scopthamlmidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Scopthalmus aquosus 206 Scopthamlmidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Trinectes maculata 66 Achiridae	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Achirus lineatus 67 Achiridae	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Gymnachirus melas 68 Achiridae	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Syacium micrurum 96 Paralichthyidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Syacium papillosum 94 Paralichthyidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Citharichthys xanthostigma 87 Paralichthyidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Etropus crossotus 98 Paralichthyidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Etropus microstomus 91 Paralichthyidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Cynoglossus interuptus 84 Cynoglossidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Symphurus atricauda 84 Cynoglossidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Engyophrys senta 76 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Bothus pantherinus 77 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Bothus lunatus 70 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Engyprosopon cf. multisquama 72 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Laeops kitaharae 73 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Chascanopsetta lugubris 79 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Chascanopsetta prognahus 80 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Asterorhombus cocosensis 78 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Psettina tosana 74 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Crossorhombus kobensis 71 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Tosarhombus octoculatus 204 Bothidae	1	0	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0
Hippoglossina oblonga 89 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Tarphops oligolepis 93 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Ancylopsetta ommata 86 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Paralichthys dentatus 90 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Ancylopsetta dilecta 95 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Xystreurys liolepis 88 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Poecilopsetta plinthus 99 Poecilopsettidae	0	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0
Pseudorhombus pentophthalmus 92 Paralichthyidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Atheresthes evermanni 1 Pleuronectidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Atheresthes stomias 2 Pleuronectidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Eopsetta jordani 12 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Pleuronichthys coenosus 35 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Glyptocephalus zachirus 14 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Pleuronectes pinnifasciatus 44 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Hippoglossus stenolepis 6 Pleuronectidae	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0
Verasper moseri 56 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Cleisthenes herzensteini 62 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Hippoglossoides elassodon 31 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Platichthys stellatus 32 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Lepidopsetta bilineata 20 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Parophrys vetulus 30 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0
Psettichthys melanostictus 38 Pleuronectidae	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0

Character (cont.)	33	34	35	36	37	38	39	EDR	POSp	FrSp	SaT	EPR
Psettodes erumei 100 Psettodidae	0	0	0	0	0	0	0	1	1	1	1	0
Psettodes sp. 101 Psettodidae	0	0	0	0	0	0	0	1	1	1	1	0
Samariscus japonicus 83 Samaridae	0	1	1	1	0	0	0	1	0	0	1	0
Samariscus xenicus 102 Samaridae	0	1	1	1	0	0	0	1	0	0	1	0
Heteromycteris japonicus 105 Soleidae	1	1	1	1	0	0	1	0	0	0	0	0
Aesopia cornuta 109 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Soleichthys heterorhinos 108 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Zebrias japonica 106 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Aseraggodes kobensis 104 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Pardachirus pavoninus 203 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Aseraggodes heemstrai 107 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Parachirus xenicus 69 Soleidae	1	1	1	1	1	1	1	0	0	0	0	0
Citharoides macrolepis 82 Citharidae	0	0	1	0	0	0	0	1	1	1	1	1
Scopthalmus aquosus 205 Scopthamlmidae	0	1	1	1	0	0	0	0	1	1	1	0
Scopthalmus aquosus 206 Scopthamlmidae	0	1	1	1	0	0	0	0	1	1	1	0
Trinectes maculata 66 Achiridae	0	0	1	1	1	1	1	1	0	0	1	0
Achirus lineatus 67 Achiridae	0	0	1	1	1	1	1	1	1	1	0	0
Gymnachirus melas 68 Achiridae	0	0	1	1	1	1	1	1	1	1	0	0
Syacium micrurum 96 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	1
Syacium papillosum 94 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	1
Citharichthys xanthostigma 87 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	1
Etropus crossotus 98 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	1
Etropus microstomus 91 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	1
Cynoglossus interuptus 84 Cynoglossidae	0	1	1	1	0	0	0	1	1	1	1	0
Symphurus atricauda 84 Cynoglossidae	0	1	1	1	1	0	0	1	1	1	1	0
Engyophrys senta 76 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Bothus pantherinus 77 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Bothus lunatus 70 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Engyprosopon cf. multisquama 72 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Laeops kitaharae 73 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Chascanopsetta lugubris 79 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Chascanopsetta prognahus 80 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Asterorhombus cocosensis 78 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Psettina tosana 74 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Crossorhombus kobensis 71 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Tosarhombus octoculatus 204 Bothidae	0	1	1	1	0	0	0	1	0	0	1	0
Hippoglossina oblonga 89 Paralichthyidae	0	1	1	1	0	0	0	1	0	0	1	0
Tarphops oligolepis 93 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	0
Ancylopsetta ommata 86 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	0
Paralichthys dentatus 90 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	0
Ancylopsetta dilecta 95 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	1	0
Xystreurys liolepis 88 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	0	0
Poecilopsetta plinthus 99 Poecilopsettidae	0	1	1	1	0	0	0	0	0	0	1	0
Pseudorhombus pentophthalmus 92 Paralichthyidae	0	1	1	1	0	0	0	1	1	1	0	0
Atheresthes evermanni 1 Pleuronectidae	0	1	1	1	0	0	0	0	1	1	1	0
Atheresthes stomias 2 Pleuronectidae	0	1	1	1	0	0	0	0	1	1	1	0
Eopsetta jordani 12 Pleuronectidae	0	1	1	1	0	0	0	0	1	1	0	0
Pleuronichthys coenosus 35 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	0	0
Glyptocephalus zachirus 14 Pleuronectidae	0	1	1	1	0	0	0	0	1	1	1	0
Pleuronectes pinnifasciatus 44 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	0	0
Hippoglossus stenolepis 6 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0
Verasper moseri 56 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0
Cleisthenes herzensteini 62 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0
Hippoglossoides elassodon 31 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0
Platichthys stellatus 32 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0
Lepidopsetta bilineata 20 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	0	0
Parophrys vetulus 30 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0
Psettichthys melanostictus 38 Pleuronectidae	0	1	1	1	0	0	0	0	0	0	1	0

Character (cont.)	G	OtSp	Usp	BSp	CSp	BodySp
Psettodes erumei 100 Psettodidae	0	0	0	0	0	0
Psettodes sp. 101 Psettodidae	0	0	0	0	0	0
Samariscus japonicus 83 Samaridae	1	0	0	0	0	0
Samariscus xenicus 102 Samaridae	1	0	0	0	0	0
Heteromycteris japonicus 105 Soleidae	1	0	0	0	0	0
Aesopia cornuta 109 Soleidae	1	0	0	0	0	0
Soleichthys heterorhinos 108 Soleidae	1	0	0	0	0	0
Zebrias japonica 106 Soleidae	1	0	0	0	0	0
Aseraggodes kobensis 104 Soleidae	1	0	0	0	0	0
Pardachirus pavoninus 203 Soleidae	1	0	0	0	0	0
Aseraggodes heemstrai 107 Soleidae	1	0	0	0	0	0
Parachirus xenicus 69 Soleidae	1	0	0	0	0	0
Citharoides macrolepis 82 Citharidae	0	0	0	0	0	0
Scopthalmus aquosus 205 Scopthamlmidae	0	1	0	0	0	0
Scopthalmus aquosus 206 Scopthamlmidae	0	1	0	0	0	0
Trinectes maculata 66 Achiridae	0	1	0	0	0	1
Achirus lineatus 67 Achiridae	1	1	0	0	0	1
Gymnachirus melas 68 Achiridae	1	1	0	0	0	1
Syacium micrurum 96 Paralichthyidae	0	1	0	0	0	0
Syacium papillosum 94 Paralichthyidae	0	1	0	0	0	0
Citharichthys xanthostigma 87 Paralichthyidae	0	1	0	0	0	0
Etropus crossotus 98 Paralichthyidae	0	1	0	0	0	0
Etropus microstomus 91 Paralichthyidae	0	1	0	0	0	0
Cynoglossus interuptus 84 Cynoglossidae	1	0	0	0	0	0
Symphurus atricauda 84 Cynoglossidae	1	0	0	0	0	0
Engyophrys senta 76 Bothidae	0	0	0	0	0	0
Bothus pantherinus 77 Bothidae	0	0	1	1	1	1
Bothus lunatus 70 Bothidae	0	0	1	1	1	1
Engyprosopon cf. multisquama 72 Bothidae	0	0	1	1	1	1
Laeops kitaharae 73 Bothidae	0	0	1	1	1	1
Chascanopsetta lugubris 79 Bothidae	1	0	1	1	1	1
Chascanopsetta prognahus 80 Bothidae	1	0	1	1	1	1
Asterorhombus cocosensis 78 Bothidae	0	0	1	1	1	1
Psettina tosana 74 Bothidae	0	0	1	1	1	1
Crossorhombus kobensis 71 Bothidae	0	0	1	1	1	1
Tosarhombus octoculatus 204 Bothidae	0	0	1	1	1	1
Hippoglossina oblonga 89 Paralichthyidae	0	1	0	0	0	0
Tarphops oligolepis 93 Paralichthyidae	0	1	0	0	0	0
Ancylopsetta ommata 86 Paralichthyidae	0	1	0	0	0	0
Paralichthys dentatus 90 Paralichthyidae	0	1	0	0	0	0
Ancylopsetta dilecta 95 Paralichthyidae	0	1	0	0	0	0
Xystreurys liolepis 88 Paralichthyidae	0	1	0	0	0	0
Poecilopsetta plinthus 99 Poecilopsettidae	0	0	0	0	0	0
Pseudorhombus pentophthalmus 92 Paralichthyidae	0	1	0	0	0	0
Atheresthes evermanni 1 Pleuronectidae	0	0	0	0	0	0
Atheresthes stomias 2 Pleuronectidae	0	0	0	0	0	0
Eopsetta jordani 12 Pleuronectidae	0	0	0	0	0	0
Pleuronichthys coenosus 35 Pleuronectidae	0	1	0	0	0	0
Glyptocephalus zachirus 14 Pleuronectidae	0	1	0	0	0	0
Pleuronectes pinnifasciatus 44 Pleuronectidae	0	0	0	0	0	0
Hippoglossus stenolepis 6 Pleuronectidae	0	0	0	0	0	0
Verasper moseri 56 Pleuronectidae	0	0	0	0	0	0
Cleisthenes herzensteini 62 Pleuronectidae	0	0	0	0	0	0
Hippoglossoides elassodon 31 Pleuronectidae	0	0	0	0	0	0
Platichthys stellatus 32 Pleuronectidae	0	0	0	0	0	0
Lepidopsetta bilineata 20 Pleuronectidae	0	0	0	0	0	0
Parophrys vetulus 30 Pleuronectidae	0	0	0	0	0	0
Psettichthys melanostictus 38 Pleuronectidae	0	0	0	0	0	0

2.3 PHYLOGENETIC INFERENCE

I inferred the phylogeny used for all downstream comparative analyses with the entire dataset, including four genes (*rho, rnf213, tmem22* and *chst2*), that when sequenced for flatfishes and 90 outgroups resulted in a concatenated alignment totaling 3,781 sites (53% were variable). This dataset is referred to as the "Con" matrix and tree. The concatenated matrix was partitioned by gene and analyzed in Treefinder (Jobb, 2008) with 1000 bootstrap replicates. The ML analysis of the entire dataset resulted in a well-supported, monophyletic Pleuronectiformes with Psettoidei sister to Pleuronectoidei. I pruned and used this portion of the tree for all downstream analyses and discussions. A detailed description and justification of the materials and methods used to infer this phylogeny for flatfishes is given in Chapter III (p. 21).

To investigate whether the alternative coding and/or addition of larval characters to Chapleau's (1993) matrix would result in an increase or decrease in resolution (compared to his results shown in Fig. 4-1 here), I carried out separate MP and ML analyses of the 50 character matrix. To obtain the tree with shortest tree length, I performed two MP (Wagner parsimony) analyses separately in the programs Pars and Penny, both available in the Phylip 3.7a package (Felsenstein, 2009). In Pars, I used the "more thorough search" option and left all other choices as default. In Penny, I increased the run length to 1,000,000 million trees generated (all sampled every 100 trees) to see if a tree (or tree) with a shorter length than the one(s) inferred using Pars could be obtained; other than this I used default settings. To create a majority-rule consensus of all MP trees, I used Consense, software that is also available in the Phylip 3.7a package.

I performed the ML reconstruction of the morphological matrix using the GAMMA model of rate heterogeneity for binary characters in RAxML (Stamatakis, 2006) with bootstrap proportions derived using the rapid bootstrapping algorithm (Stamatakis et al., 2008). I completed the ML analysis through the CIPRES portal (Miller et al., 2010) available at http://www.phylo.org/sub_sections/portal/.

2.4 TOPOLOGY TESTS

To test alternative phylogenic hypotheses regarding relationships within Pleuronectiformes given the molecular dataset described in Chapter III (p.21), I trimmed the alignment to include only Pleuronectiformes. I then selected nucleotide substitution models for each gene partition (*rho*, *rnf213*, *tmem22*, *chst2*) using the AICc carried out in Treefinder (Jobb, 2008). Because topology tests require that *the* ML tree be included in all comparisons, I analyzed the trimmed alignment (without bootstrap replicates) and rooted the phylogeny using *Psettodes*, to determine if that topology differed from the one inferred using all the data. The topologies did not differ, except in branch length estimates.

I then carried out KH (Kishina and Hasegawa, 1989), AU (Shimodaira, 2002) and SH (Shimodaira and Hasegawa, 1999) tests using Treefinder (2008), comparing the resultant topologies to the one inferred using the four genes: Chapleau's (1993) consensus tree, Berendzen and Dimmick's (2002) Bayesian (consensus) tree (Fig. 4-2), Betancur-R et al.'s (trimmed) topology (2013b; reproduced in Figure 4-3 here) as well as the MP tree(s) and ML tree inferred using the 50 character matrix described above. Only Berendzen and Dimmick's (2002) and Betancur-R et al.'s (2013b) trees required pruning to match the Con topology for Pleuronectiformes.

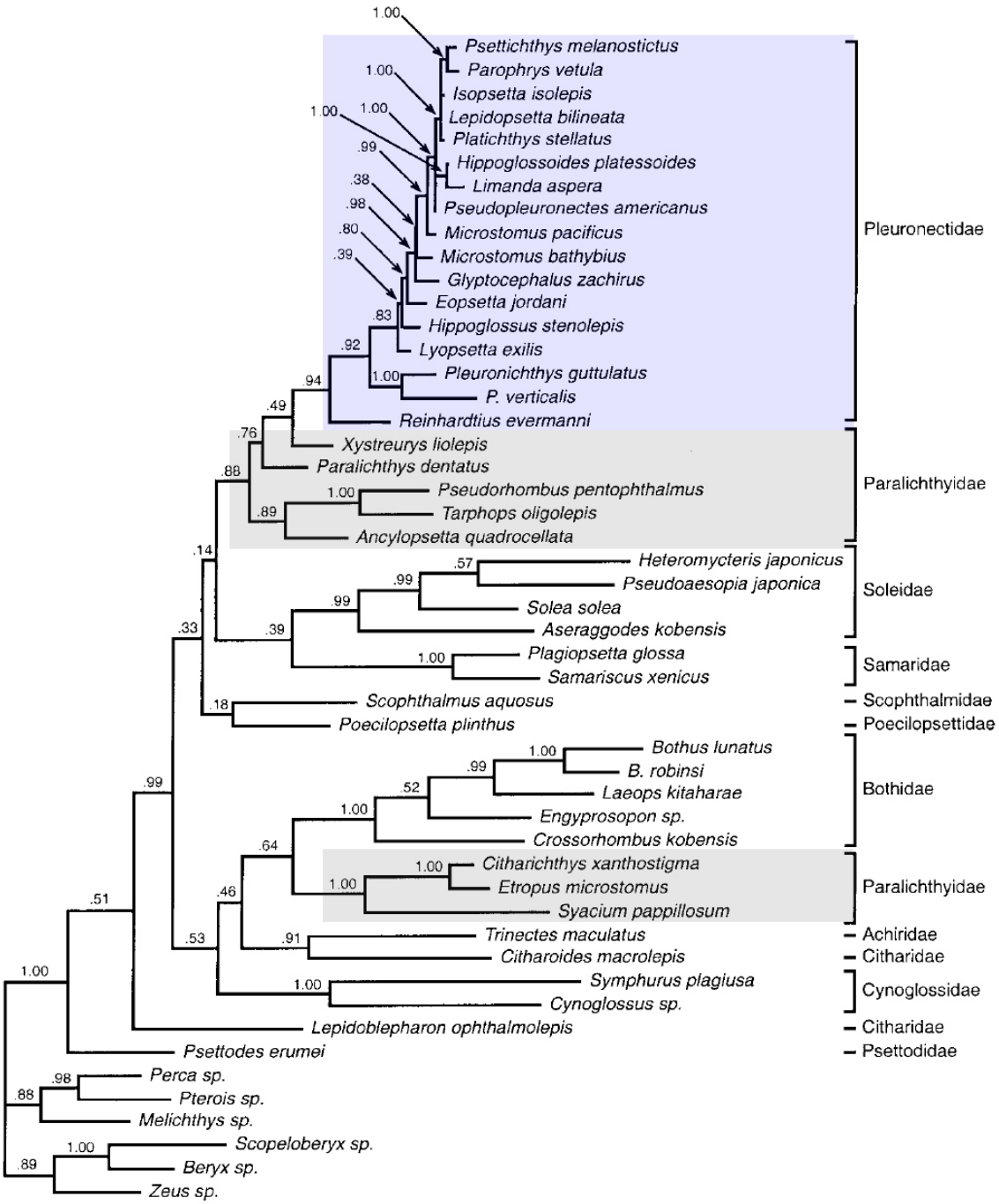


Figure 4-2. The 50% majority rule consensus of the posterior distribution of trees inferred using 12S and 16S sequence data for Pleuronectiformes; posterior probabilities are on the branches. Reproduced from Berendzen and Dimmick (2002), Figure 5. Blue clade is Pleuronectidae and gray are the paralichthyid groups, with a monophyletic *Cyclopsetta* group sister to Bothidae.

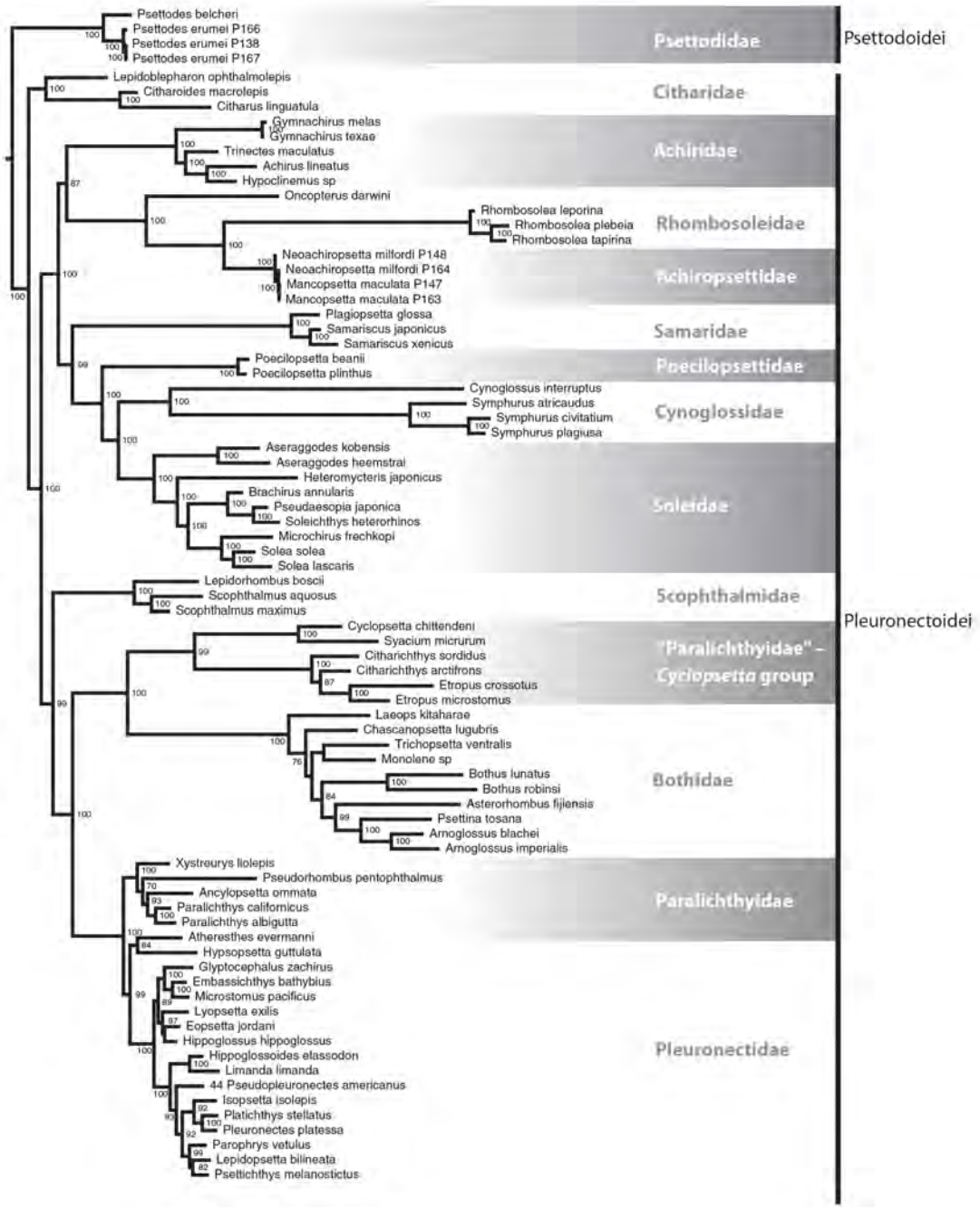


Figure 4-3. The favored phylogenetic hypothesis of Betancur-R et al. (2013a) for Pleuronectiformes—the ML tree using the GRT3 model for AGY-recoded partitions—inferred from the complete data set (214 taxa and 20 genes); numbers at nodes indicate support values ≥ 75, estimated with the rapid bootstrapping algorithm in RAxML. Reproduced from Betancur-R (2013b) Figure 3, part (c).

Because the main goal of these studies was to identify relationships among major groups (the families and "Paralichthyidae") and not the relationships within them, I carried out two rounds of topology tests: one where trees were fully resolved (if applicable) and one where monophyletic families (or the *Cyclopsetta* group) were collapsed. It was expected that increasing the ambiguity within families would also decrease the difference in likelihoods among topologies, possibly resulting in P-values that fail to reject the null.

I implemented hypothesis tests in Treefinder (Jobb, 2008) using the RELL (resample estimated log-Likelihood) nonparametric bootstrap method with 50,000 replicates used to generate the null distributions for all three tests. I used the same models to calculate the likelihoods for each topology as those chosen using the AICc in Treefinder (Jobb, 2008). The null hypothesis (H_0: all topologies for comparison are equally good explanations of the data) was rejected when $P < 0.05$.

2.4 Comparative Analyses

To carry out all downstream analyses of character evolution, I generated a rooted ultrametric tree for Pleuronectiformes. I trimmed the large alignment (four genes, 58 flatfish plus 90 outgroups) described above to include only flatfish taxa and analyzed it in BEAST 201 v1.6.2 (Drummond and Rambaut, 2007) using a template from BEAUTI v1.6.2, with results visualized in TRACER v.1.4. The ML topology for flatfishes that was inferred using the entire dataset was used to constrain the ultrametric topology. I estimated mean substitution rates under a relaxed uncorrelated lognormal clock, allowing for independent rates along each branch (Drummond et al., 2006). I carried out one analysis consisting of 50,000,000 generations with a burnin of 20% with parameters and

trees sampled every 1,000 generations. I then used TRACER v.1.4 to assess stationarity and the effective sample size of all parameters (Drummond and Rambaut, 2007).

To test for independent evolution, I carried out all possible (unique) pairwise comparisons for all 50 characters (((50 x 50)-50)/2 = 1,225 tests) using a modified version of Pagel's (1994) statistical test of non-independent evolution of two discrete, binary characters implemented using the Correl module (Midford and Maddison, 2008) in Mesquite v. 2.75 (Madisson and Maddison, 2011). This version uses likelihood ratio tests comparing a four-parameter model, where the rates of change in each character are independent of the state of the other, to an eight-parameter mode, in which the rates depend on the state of the other character. To generate P-values, I performed 20 rounds of initial likelihood iterations, followed by 200 rounds of simulations and used a Chi-square test to approximate distribution with the null hypothesis (independence) being rejected when $P < 0.05$.

Before I implemented the ancestral character estimation, I evaluated the level and type of phylogenetic signal using *D*, a measure of phylogenetic signal of binary characters (Fritz and Purvis, 2010), for each character using the R package caper (Orme, 2013). I also generated probabilities of two models of character evolution (random and Brownian motion). Negative D-values correspond to an extremely clumped distribution of a character on a tree and a large positive D value corresponds to an over dispersed model where no predictions regarding ancestral state can be made. A D-value of 0 represents a Brownian motion model; a value of 1.0 represents the random model. I then carried out ancestral character state estimation (ACE) for those characters with a D-value ≤ 0. I implemented the ACE analysis in the R package ape (Paradis et al., 2004), using

ML with an equal rates Mk 1 model (Lewis, 2001): two character states and equal rates of transition.

Under the one parameter Mk 1 model, expected rates of character change are constant along every branch and throughout time and the rates of instantaneous change are equal. To reconstruct the ancestral state of a discrete binary character, the ML estimate at each node is dependent on the observed character states, topology, and branch lengths (treated as time). The ML reconstruction is that which is maximizing the probability of the data given the model. Therefore, when all branch lengths are assigned to 1.0, the rate parameters will be small—the MP condition (minimizing change).

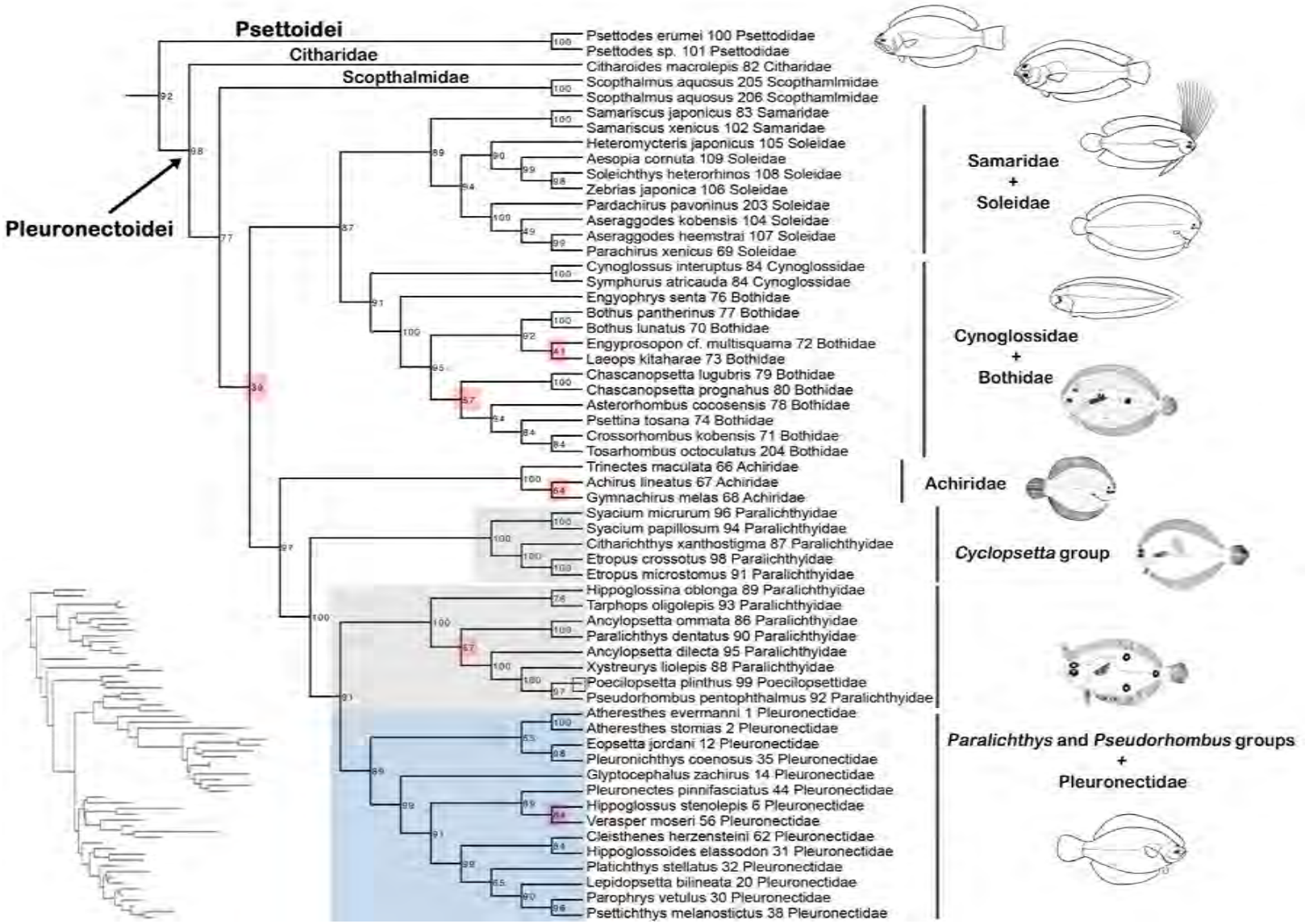
To investigate the impact high rates and branch length information had on the accuracy of ACEs, I calculated estimated rates using the ultrametric tree with branch length information (the ML case) and then using the same tree with all branch lengths assigned to 1.0 (the MP case). I then calculated differences in individual likelihoods (ML-MP) and their rates for each character given the two analyses. Because the MP "model" assumes low rates of change, but cannot estimate whether rates are high, if the difference in rates is large (always greater for ML) than most likely the assumption of low rates is violated and the accuracy of the ML ACE is also expected to be low.

RESULTS

3.1 PHYLOGENY OF PLEURONECTIFORMES USING SEQUENCE DATA

I inferred a monophyletic Pleuronectiformes with 92% bootstrap support for the Con dataset using the four-partition scheme, comprised of Psettoidei (*Psettodes* spp.) sister to Pleuronectoidei (all remaining flatfishes). Figure 4-4 shows an annotated tree of just the flatfishes. The basal split in Pleuronectoidei is Citharidae, followed by

Figure 4-4. Phylogeny of Pleuronectiformes, with bootstrap support at nodes, estimated with Treefinder of the concatenated, Con, dataset for all taxa that was partitioned by gene; bootstrap values less than 70 are shaded red and branch lengths are shown in the bottom left corner. Traditional Paralichthyidae is shaded gray (with Poecilopsettidae in yellow) showing the *Cyclopsetta* group is monophyletic and sister to a larger clade comprised of Pleuronectidae sister to a well-supported clade containing the remaining paralichthyids. Illustrations are reproduced from Nelson (2006), except Bothidae, Achiridae, *Cyclopsetta* and *Pseudorhombus* examples that are reproduced from Lopez-Martinez et al. (2010).


Psettoidei
Citharidae
Scopthalmidae
Pleuronectoidei
Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Citharoides macrolepis 82 Citharidae
Scopthalmus aquosus 205 Scopthamlmidae
Scopthalmus aquosus 206 Scopthamlmidae
Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Cynoglossus interuptus 84 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Bothus pantherinus 77 Bothidae
Bothus lunatus 70 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta lugubris 79 Bothidae
Chascanopsetta prognahus 80 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Trinectes maculata 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 68 Achiridae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
Samaridae
+
Soleidae
Cynoglossidae
+
Bothidae
Achiridae
Cyclopsetta group
Paralichthys and Pseudorhombus groups
+
Pleuronectidae

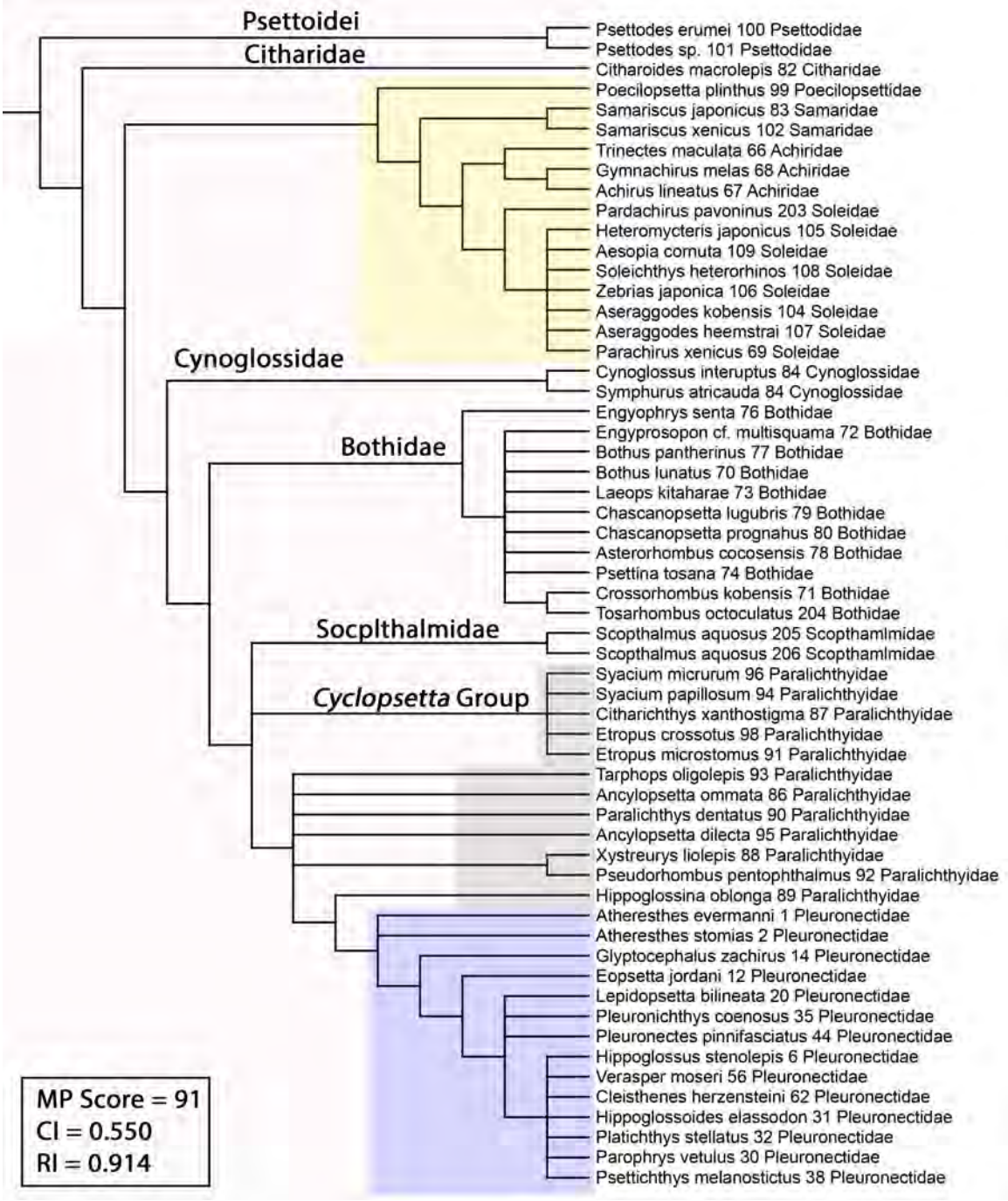
Scopthalmidae, recovered as sister to a large, poorly supported, clade that is split in to two large, highly supported, groups. The first group includes samarids + soleids, sister to cynoglossids + bothids, all with high bootstrap values. The other group consists of achirids sister to a paraphyletic Paralichthyide with the *Paralichthys* and *Pseudorhombus* groups more closely related to a monophyletic Pleuronectidae than to the *Cyclopsetta* group.

Poecilopsettidae is represented here by one (of 20) species and is recovered as sister to *Pseudorhombus pentopthalmus* (nested well within the *Paralichthys* group + *Pseudorhombus* group clade). This appears to be an artifact due to long-branch attraction (see Fig. 4-4), probably caused by high branch length heterogeneity due to low taxon sampling within the family (Felsenstein, 1978; Bergsten, 2005) and/or substitution saturation at the third codon position of *rho*; see Chapter III, Fig. 3-17 (p. 79) where this is described in detail. It may also be due to convergence caused by base compositional bias in *rnf213*, (again, see Chapter III, p. 75), or all of the above. Unfortunately, only *rho* and *rnf213* were sequenced for *P. pentopthalmus* so there was no additional data that may have resolved these problems.

3.2 Phylogeny Inferred with Morphological Characters

Although the Penny runs were extended, they never resulted in a tree with fewer steps than the Pars tree, with the best Penny tree having a score of 93. The Pars run resulted in six most parsimonious trees each with a score (tree length) of 91, a consistency index (CI) of 0.544 and a retention index (RI) of 0.912; the majority rule consensus is shown in Fig. 4-5. I recovered pleuronectoids with Citharidae sister to the rest, and all families except for Paralichthyidae monophyletic. Within the paralichthyids,

Figure 4-5. Majority-rule consensus of six MP trees estimated in Pars (available in Phylip 3.695) using the morphological data matrix of 50 (39 adult and 11 larval) discrete binary characters. Traditional Paralichthyidae is shaded gray showing a monophyletic *Cyclopsetta* Group sister to a large clade with a paraphyletic *Paralichthys* + *Pseudorhombus* groups with a monophyletic Pleuronectidae (blue). Yellow clade shows the MP reconstruction of Poecilopsettidae sister to Samaridae + Achiridae + Soleidae. Bottom left box lists tree length (MP score), the consistency index (CI) and the retention index (RI).


Psettoidei
Citharidae
Cynoglossidae
Bothidae
Socplthalmidae
Cyclopsetta Group
Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Citharoides macrolepis 82 Citharidae
Poecilopsetta plinthus 99 Poecilopsettidae
Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Trinectes maculata 66 Achiridae
Gymnachirus melas 68 Achiridae
Achirus lineatus 67 Achiridae
Pardachirus pavoninus 203 Soleidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Cynoglossus interuptus 84 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Bothus pantherinus 77 Bothidae
Bothus lunatus 70 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta lugubris 79 Bothidae
Chascanopsetta prognahus 80 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Scopthalmus aquosus 205 Scopthamlmidae
Scopthalmus aquosus 206 Scopthamlmidae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
MP Score = 91
CI = 0.550
RI = 0.914

the *Cyclopsetta* group was recovered in one clade, but the *Paralichthys* group and *Pseudorhombus* group taxa were almost entirely unresolved in a clade with a monophyletic Pleuronectidae. Here, *Poecilopsetta* was recovered as sister to a mostly resolved clade with Soleidae + Achiridae sister to Samaridae.

The ML analysis of the same matrix implemented in RAxML recovered a tree similar to the MP phylogeny, but with very low bootstrap support overall and with the base being "comb-like" (Fig. 4-6). The branch lengths were shallow when compared to the outgroup, *Psettodes*, and showed a high degree of variance across the tree. Regardless, the analysis resulted in a ML tree with the same *Pocilopsetta* + Samaridae +Achiridae + Soleidae clade as the MP tree and was also able to recover a monophyletic *Cyclopsetta* group.

In the ML tree, however, the placement of the *Paralichthys* group and *Pseudorhombus* group taxa were almost entirely unresolved. All families except Pleuronectidae (*Hippglossina*, a paralichthyid, is nested within the pleuronectids) and Paralichthyidae were recovered as monophyletic, but the statistical support provided in this analysis is very low at the branch leading to Bothidae, as well as the internodes at the base. The tree length, CI and RI of the fully resolved ML tree are 93, 0.550 and 0.914.

The following models were chosen by Treefinder using the AICc for each partition that was trimmed to include only flatfish taxa: J3[Optimum,Empirical]:G[Optimum]:5 for *rho;* J2 [Optimum,Empirical]:G[Optimum]:5, for *rnf213* and GTR[Optimum,Empirical]:G[Optimum]:5 for *tmem22* and *chst2*. I used those models with that partitioning scheme for all three (KH, AU and SH) tests and topologies for comparison. All pairwise tests of both rounds of tests (resolved trees and

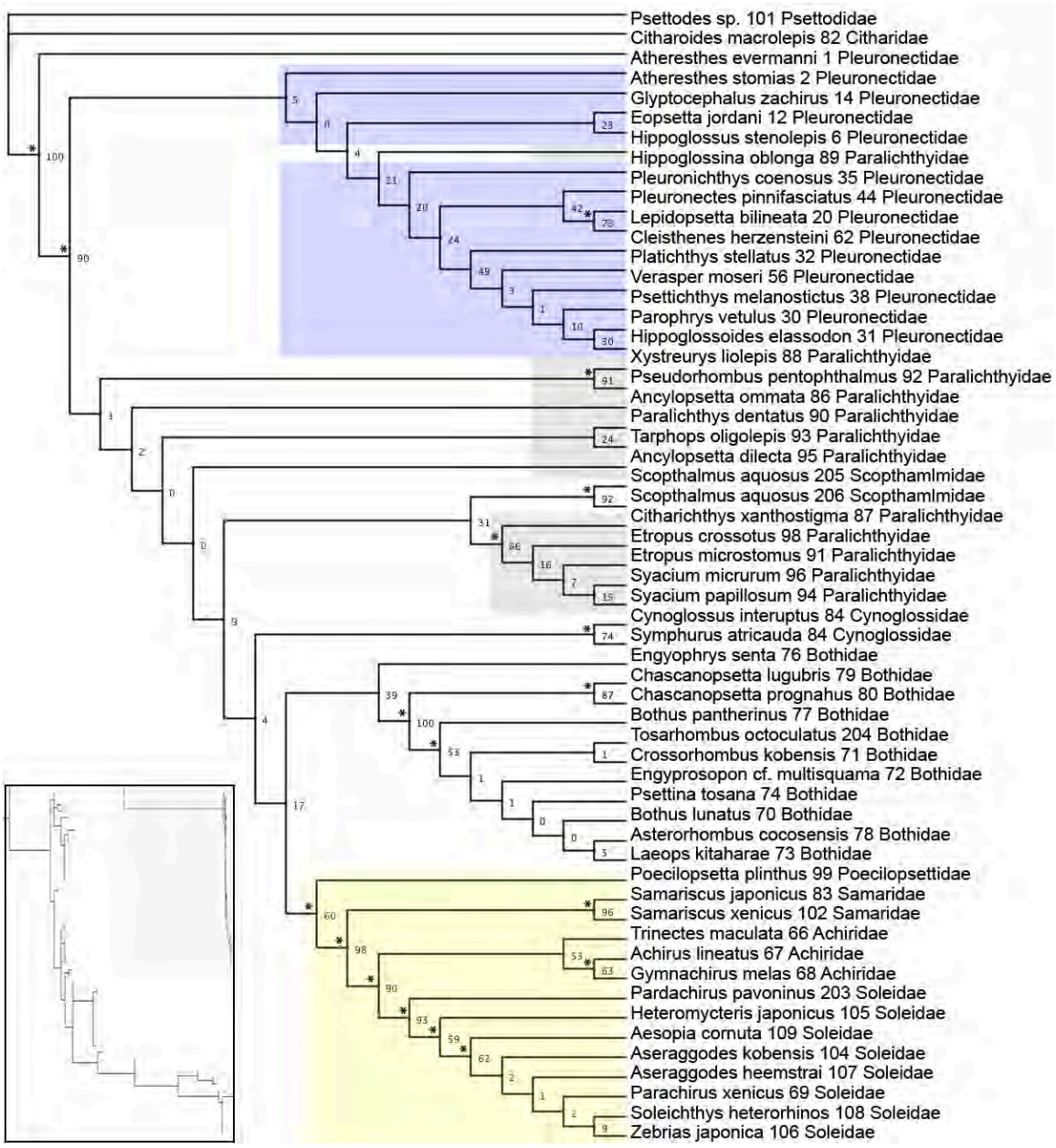


Figure 4-6. ML reconstruction of the morphological matrix using the GAMMA model of rate heterogeneity for binary characters in RAxML. Bootstrap proportions, derived using the rapid bootstrapping algorithm, are shown at nodes; those with an asterisk represent support values that are ≥ 50. Blue is Pleuronectidae, gray are paralichthyids and yellow is the Poecilopsettidae + Samaridae + Achiridae + Soleidae clade (same as the MP topology). Bottom-left box shows branch lengths after Psettodes is trimmed and before (the tree shaded gray inside the box).

collapsed families) rejected the null hypothesis (H_0: all topologies for comparison are equally good explanations of the data), with the only non-zero P-value (0.0036) being generated using the SH test when comparing these data to the collapsed topology of Berendzen and Dimmick (2002).

3.3 Comparative Phylogenetic Analyses

I confirmed convergence of the ultrametric trees generated in BEAST (with the Con ML topology constrained on the analyses) with the trace in TRACER v.1.4, with no significant fluctuations post-burnin. I chose the maximum clade credibility tree and it, with branch lengths, is summarized in Fig.4-7. I used that tree in all downstream analyses.

All pairwise tests of non-independent evolution for all 50 characters, *all* 1225 tests, resulted in P-values of zero. Therefore, in all cases the null hypothesis (independence) was rejected (i.e. when $P < 0.05$). This implies that adult/adult, larval/larval and adult/larval character pairs, given the ultrametric tree and the observed states (in Table 4-1), evolved independently.

All D-values generated were negative and ranged from -2.3977 (character 6) to -0.076 (SaT). All probabilities of random signal were zero and all probabilities of evolution via Brownian motion were near 1.0, the exception was the P(BM) for SaT at 0.389. Overall, the larval characters showed the least negative D-values and lowest P(BM). These values are listed for all characters in Table 4-2.

I carried out ACE on all characters and as expected, based on D-values and P(BM), the larval characters, opposed to the adult, had the highest uncertainty at the internal nodes (Figures 4-8 to 4-15). Quite a few character states (adult and larval)

Figure 4-7. Mirror tree image showing change in branch lengths from original ML tree of Pleuronectiformes (left) to the ultrametric tree (right) inferred in BEAST, constrained by the ML topology, with mean substitution rates estimated under a relaxed uncorrelated lognormal clock allowing for independent rates along each branch (note: the two phylogenies do not share the same scale).

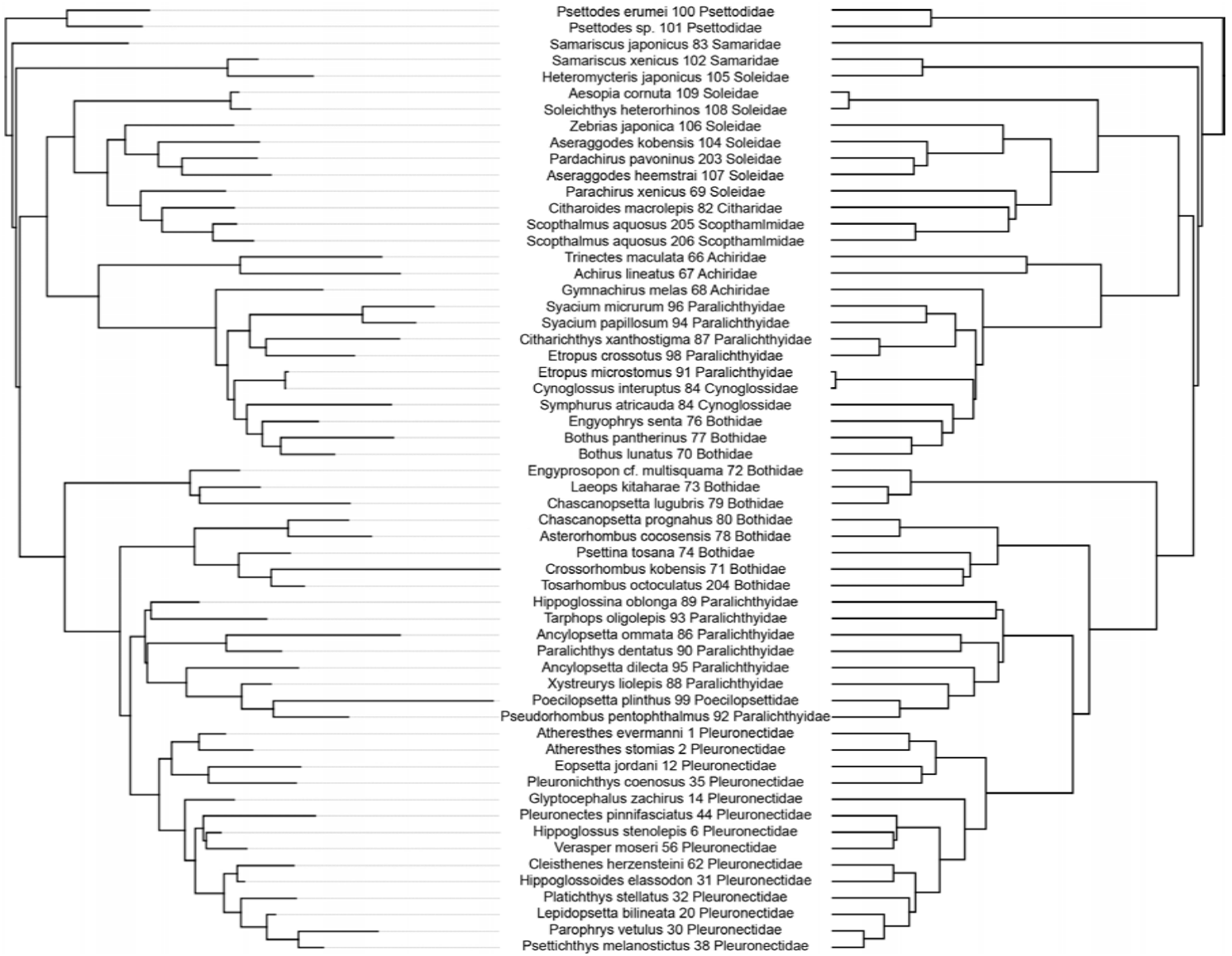
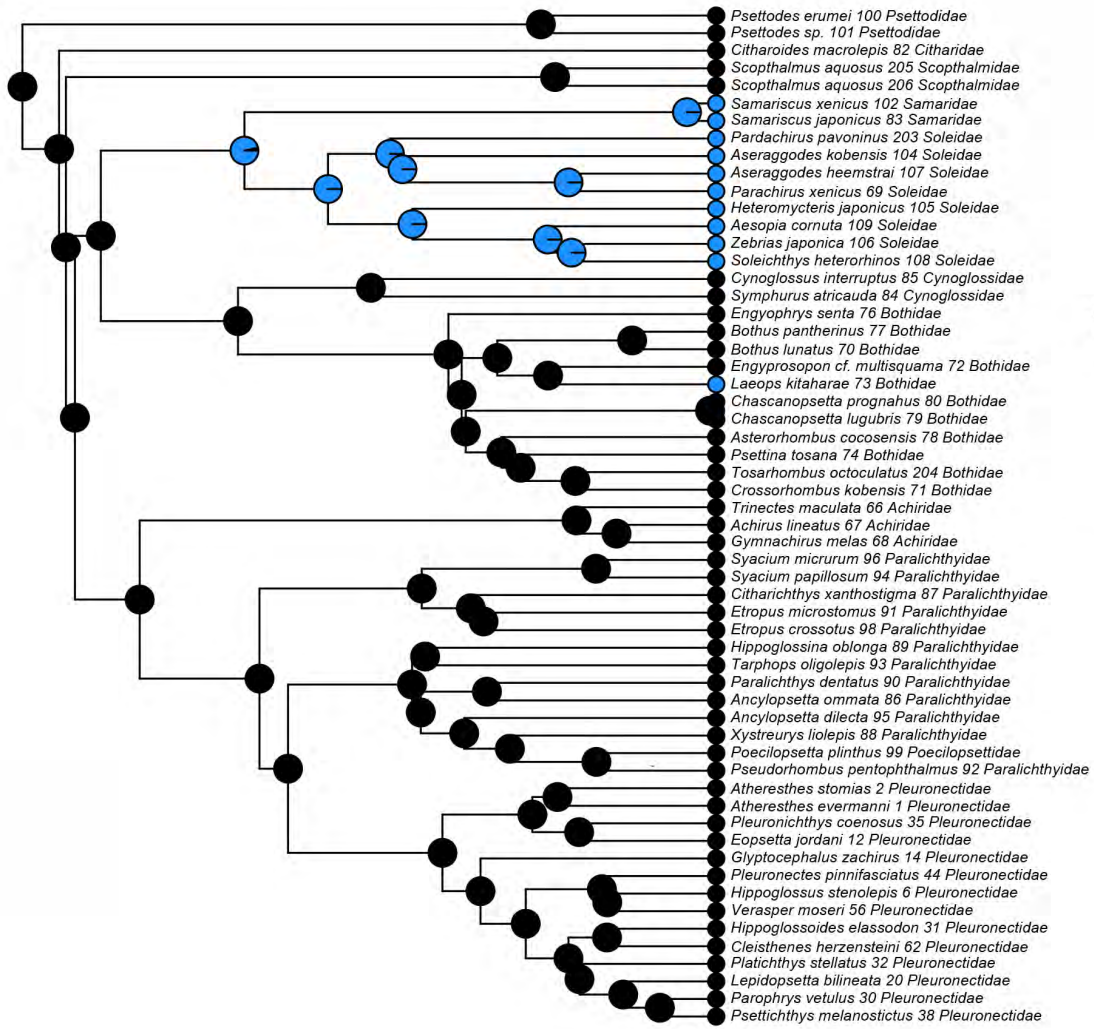
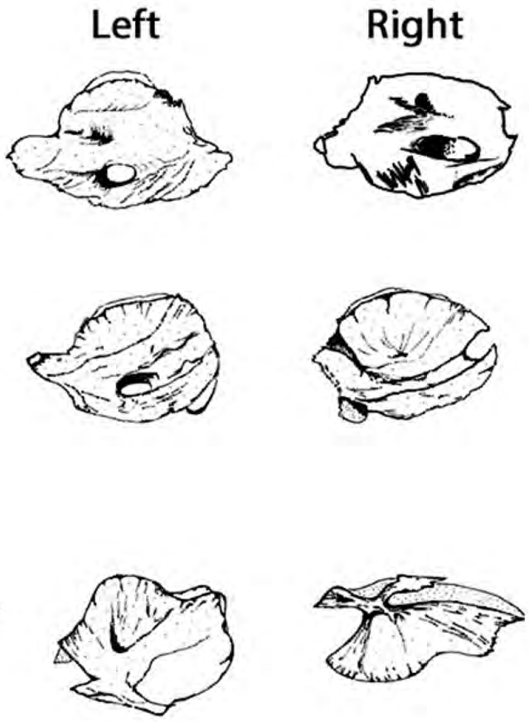

Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Samariscus japonicus 83 Samaridae
Samariscus xenicus 102 Samaridae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Soleichthys heterorhinos 108 Soleidae
Zebrias japonica 106 Soleidae
Aseraggodes kobensis 104 Soleidae
Pardachirus pavoninus 203 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Citharoides macrolepis 82 Citharidae
Scopthalmus aquosus 205 Scopthamlmidae
Scopthalmus aquosus 206 Scopthamlmidae
Trinectes maculata 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 68 Achiridae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Cynoglossus interuptus 84 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Bothus pantherinus 77 Bothidae
Bothus lunatus 70 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta lugubris 79 Bothidae
Chascanopsetta prognahus 80 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Crossorhombus kobensis 71 Bothidae
Tosarhombus octoculatus 204 Bothidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Atheresthes evermanni 1 Pleuronectidae
Atheresthes stomias 2 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae

Table 4-2. Summary of comapartive analyses. D-values, probability of random (no phylogenetic structure) and probability of Brownian motion evolution per character given the ultrametric tree reconstructed in BEAST. Character likelihoods (-lnL) and Mk 1 estimated rates when branch lengths are informative (ML), when they are not (same topology, but all branch lengths assigned to 1.0— the MP case) and the difference in both values per character (1-39 are adult; the others are larval). The differences were calculated as ML-MP; and data are sorted by ascending difference in likelihood and then difference in estimated rate—not by character order. The bolded rows correspond to the ACEs in Figures 4-8 to 4-15.

				With Branch Lengths		Branch Lengths Assigned to 1.0		Differences	
Char	D-value	P(random)	P(BM)	Char Lk	Mk 1Est Rate	Char Lk	Mk 1Est Rate	Char Lk	Mk 1Est Rate
SaT	**-0.076**	**0**	**0.389**	**-28.19073466**	**1.00665215**	**-25.60949907**	**0.07387811**	**-2.58123559**	**0.93277404**
BSp	-0.6813	0	0.98	-8.472040299	0.13429604	-6.40688127	0.00901488	-2.065159029	0.12528116
CSp	-0.8152	0	0.982	-8.472040299	0.13429604	-6.40688127	0.00901488	-2.065159029	0.12528116
Usp	-0.8152	0	0.975	-8.472040299	0.13429604	-6.40688127	0.00901488	-2.065159029	0.12528116
18	**-0.6194775**	**0**	**0.999**	**-16.19110139**	**0.47866347**	**-14.39539988**	**0.02871577**	**-1.79570151**	**0.4499477**
25	**-0.6381765**	**0**	**0.993**	**-16.19110139**	**0.47866348**	**-14.39539988**	**0.02871577**	**-1.79570151**	**0.44994771**
24	-1.166	0	0.96	-7.60866413	0.20022769	-6.39803491	0.00909383	-1.21062922	0.19113386
36	-1.2028	0	0.959	-7.60866413	0.20022769	-6.39803491	0.00909383	-1.21062922	0.19113386
G	-0.5264	0	0.976	-17.5524561	0.53005328	-17.02247381	0.03947139	-0.529982287	0.49058189
29	-1.0844	0	0.99	-6.85045921	0.11134117	-6.40680218	0.009016296	-0.44365703	0.102324874
30	-1.0719	0	0.989	-6.85045921	0.11134117	-6.40680218	0.009016296	-0.44365703	0.102324874
32	-1.0966	0	0.99	-6.85045921	0.11134117	-6.40680218	0.009016296	-0.44365703	0.102324874
33	-1.0599	0	0.982	-6.85045921	0.11134117	-6.40680218	0.009016296	-0.44365703	0.102324874
FrSp	-0.1452	0	0.714	-26.02696008	0.9251272	-25.59513501	0.07422483	-0.43182507	0.85090237
POSp	**-0.1416**	**0**	**0.733**	**-26.02696008**	**0.9251272**	**-25.59513501**	**0.07422483**	**-0.43182507**	**0.85090237**
22	-0.3049946	0	0.905	-25.21770125	0.96117904	-24.83315366	0.07491077	-0.38454759	0.88626827
31	-0.7667335	0	0.977	-10.18619449	0.22846776	-10.02232747	0.01829583	-0.16386702	0.21017193
OtSp	**-0.278**	**0**	**0.863**	**-22.03054756**	**0.77914334**	**-21.92342381**	**0.06096366**	**-0.10712375**	**0.71817968**
14	**-0.7796419**	**0**	**0.979**	**-6.34443816**	**0.10749617**	**-6.4068007**	**0.00901634**	**0.06236254**	**0.09847983**
34	-0.7759	0	0.959	-10.2564082	0.31746702	-10.65173554	0.01888834	0.395327343	0.29857868
16	-0.5312	0	0.964	-16.9882123	0.59177126	-17.63903423	0.04024126	0.650821935	0.55153
3	-2.2974	0	0.938	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
4	-2.3256	0	0.934	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
5	-2.375	0	0.936	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
6	-2.3977	0	0.946	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
7	-2.2176	0	0.928	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
8	-2.351	0	0.955	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
9	-2.3571	0	0.951	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
10	-2.3209	0	0.955	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
11	-2.1599	0	0.928	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
13	-2.1484	0	0.931	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
23	-2.3533	0	0.948	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
35	-2.1663	0	0.941	-5.0075037	0.10584693	-5.722870199	0.00893061	0.715366499	0.09691632
39	-0.8405	0	0.982	-9.94857324	0.22657069	-10.68675582	0.01857641	0.73818258	0.20799428
EDR	**-0.504**	**0**	**0.986**	**-16.88626975**	**0.47553939**	**-17.74323729**	**0.03930814**	**0.85696754**	**0.43623125**
26	**-0.6415**	**0**	**0.995**	**-16.10228226**	**0.51808941**	**-17.03962021**	**0.03931656**	**0.93733795**	**0.47877285**
27	-0.6273	0	0.994	-16.10228226	0.51808942	-17.03962021	0.03931656	0.93733795	0.47877286
28	-0.6178	0	0.991	-16.10228226	0.51808942	-17.03962021	0.03931656	0.93733795	0.47877286
12	-0.3214555	0	0.839	-16.79714406	0.47282483	-17.78475553	0.03885354	0.98761147	0.43397129
38	-0.8234	0	0.976	-12.61360651	0.34304097	-13.74473718	0.02830002	1.13113067	0.31474095
15	-0.8986	0	0.999	-9.47680227	0.21905314	-10.66895621	0.01873814	1.19215394	0.200315
19	-0.9098	0	1	-9.47680227	0.21905314	-10.66895621	0.01873814	1.19215394	0.200315
20	-0.881	0	0.996	-9.47680227	0.21905314	-10.66895621	0.01873814	1.19215394	0.200315
21	-0.9323	0	0.999	-9.47680227	0.21905314	-10.66895621	0.01873814	1.19215394	0.200315
EPR	-0.8277	0	0.957	-8.97911702	0.21528553	-10.6870946	0.01857003	1.707977576	0.1967155
17	-0.6641	0	0.99	-15.28087471	0.46353844	-17.04031532	0.039303	1.75944061	0.42423544
37	-0.4494123	0	0.898	-15.02020223	0.47096925	-17.11269274	0.03862055	2.09249051	0.4323487
BodySp	-0.681	0	0.98	-8.472040299	0.26204051	-10.68741951	0.01856406	2.215379211	0.24347645

Figure 4-8. Ancestral character state estimation (ACE) of the adult character, pterosphenoid present or absent, using ML. Black is 0 (absent) and blue is state 1 (present); the small circles at tips represent observed character states. Illustration at right is an edited version of an illustration in López-Martínez (2010) 2010); the left and right side pterosphenoids of three different species of Paralichthys (top to bottom).


Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Citharoides macrolepis 82 Citharidae
Scopthalmus aquosus 205 Scopthalmidae
Scopthalmus aquosus 206 Scopthalmidae
Samariscus xenicus 102 Samaridae
Samariscus japonicus 83 Samaridae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Zebrias japonica 106 Soleidae
Soleichthys heterorhinos 108 Soleidae
Cynoglossus interruptus 85 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Bothus pantherinus 77 Bothidae
Bothus lunatus 70 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta prognahus 80 Bothidae
Chascanopsetta lugubris 79 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Tosarhombus octoculatus 204 Bothidae
Crossorhombus kobensis 71 Bothidae
Trinectes maculata 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 68 Achiridae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Atheresthes stomias 2 Pleuronectidae
Atheresthes evermanni 1 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae


Left
Right

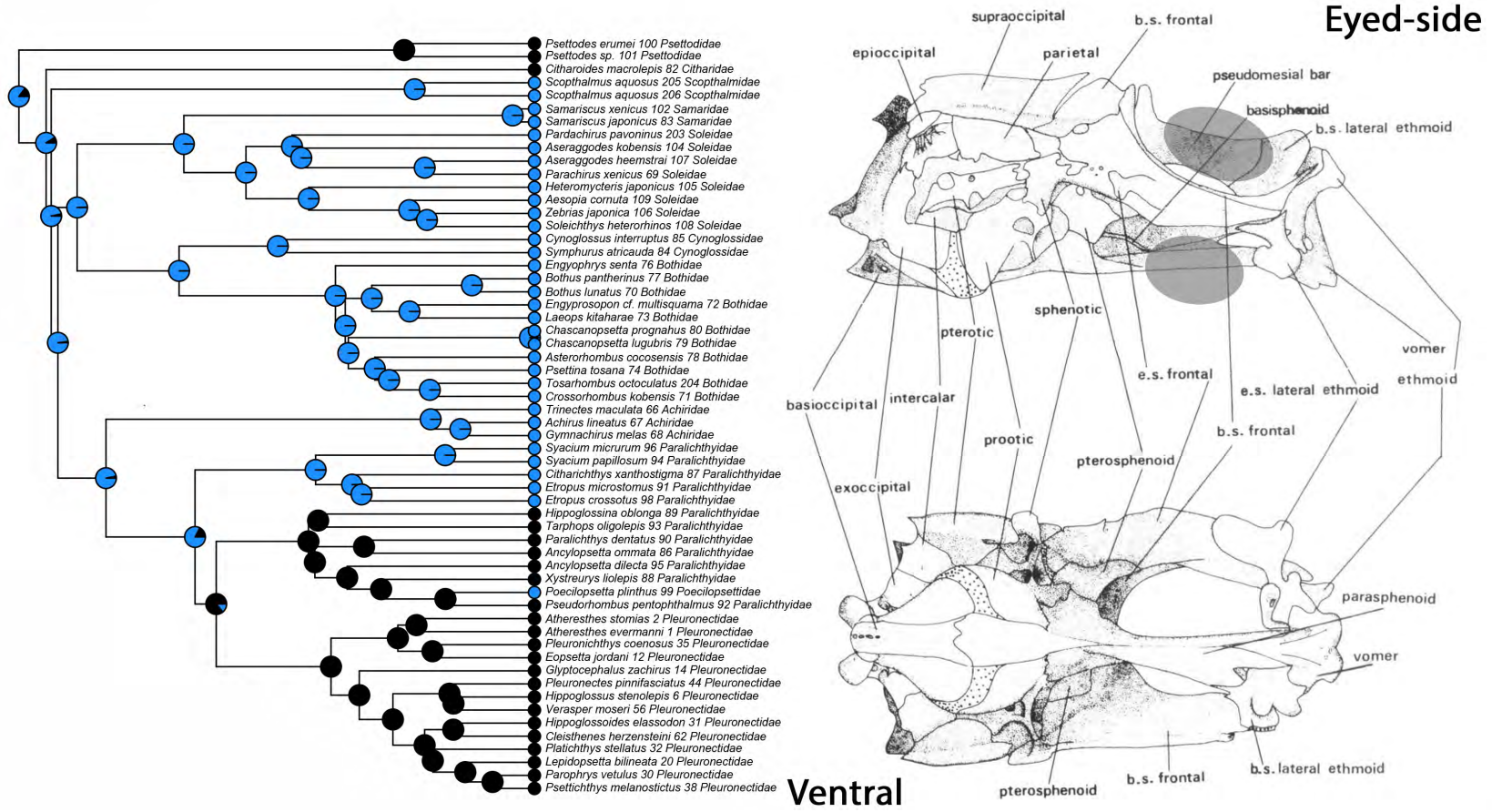


Figure 4-9. Ancestral character state estimation (ACE) of the adult character, eyed-side infraorbitals (0 = present, 1 = reduced to one or two small bones), with an edited illustration of the eyed-side and ventral side of the cranium of Psettodes, reproduced from Fig. 5 in Chapleau (1993). Black is character state 0 and blue is state 1.

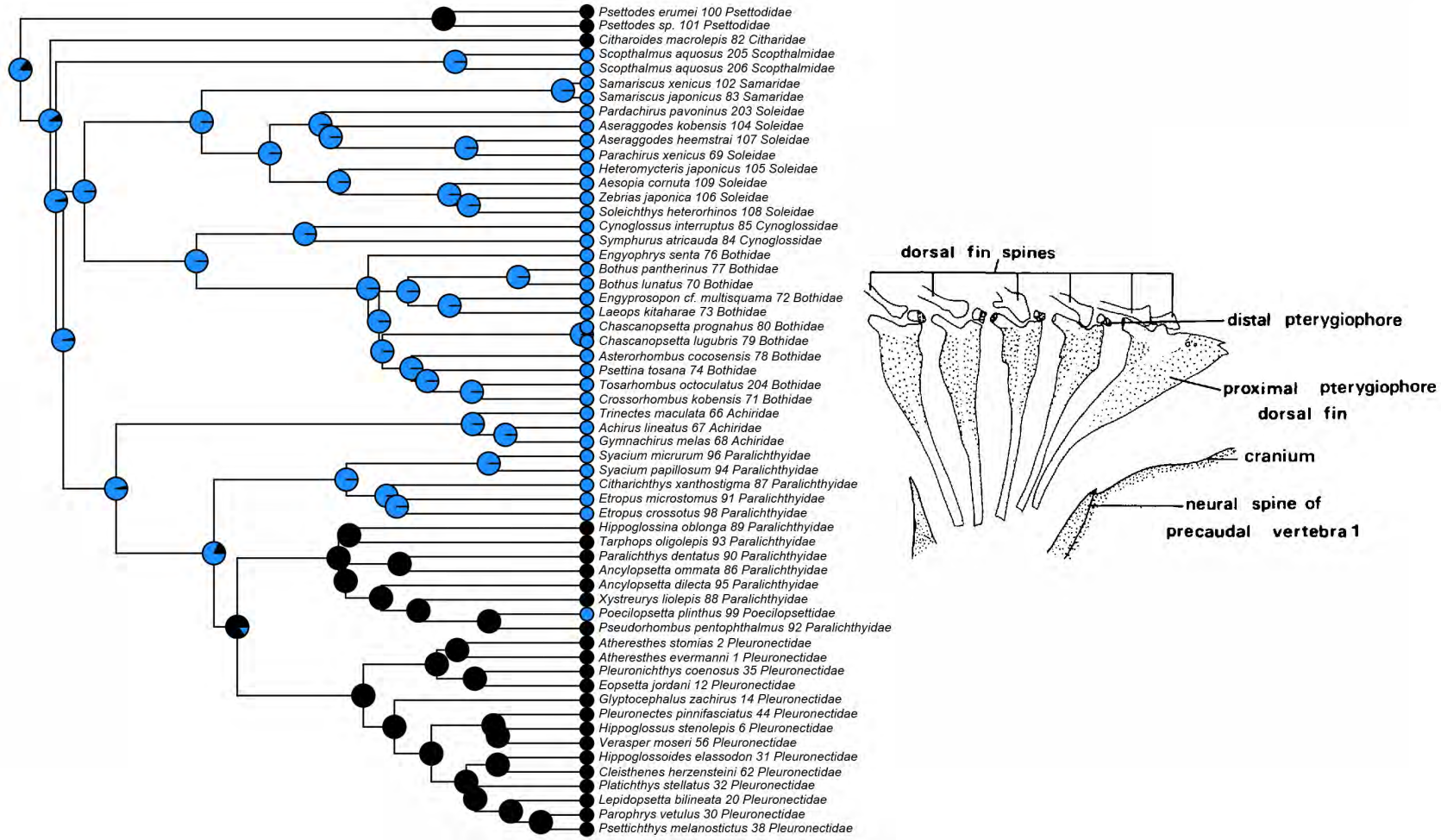


Figure 4-10. Ancestral character state estimation (ACE) of the adult character, neural arch and neural spine of first precaudal vertebra (0 = present, 1 = absence of spine or incomplete or absent arch), with an illustration reproduced from Fig. 3 in Chapleau (1993).Black is character state 0 and blue is state 1.

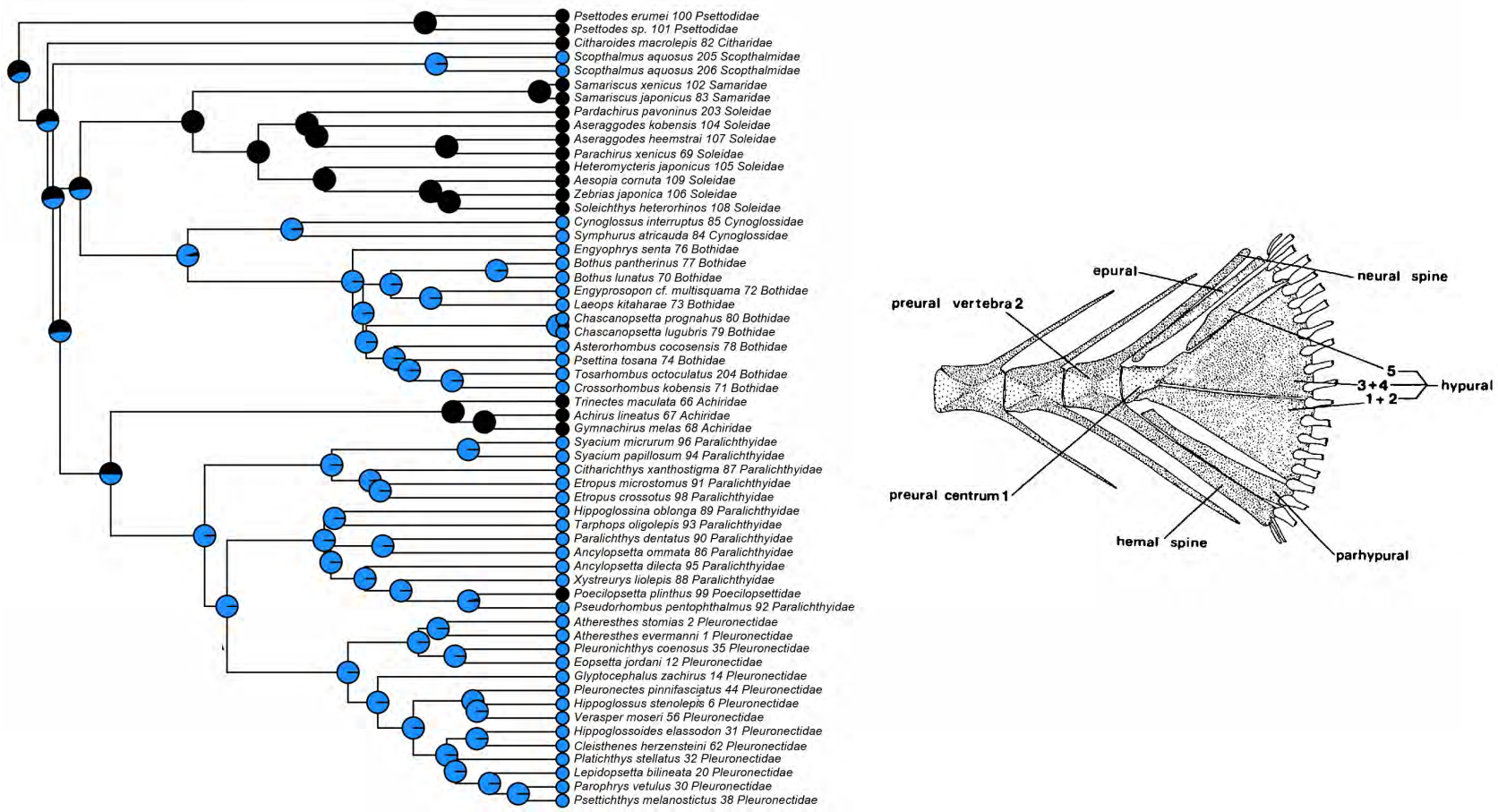


Figure 4-11. Ancestral character state estimation (ACE) of the adult character, fusion of hypural plates (0 = absent, 1 = hypurals 3 and 4 fused plus 1 and 2 fused), with an illustration reproduced from Fig. 6A in Chapleau (1993).

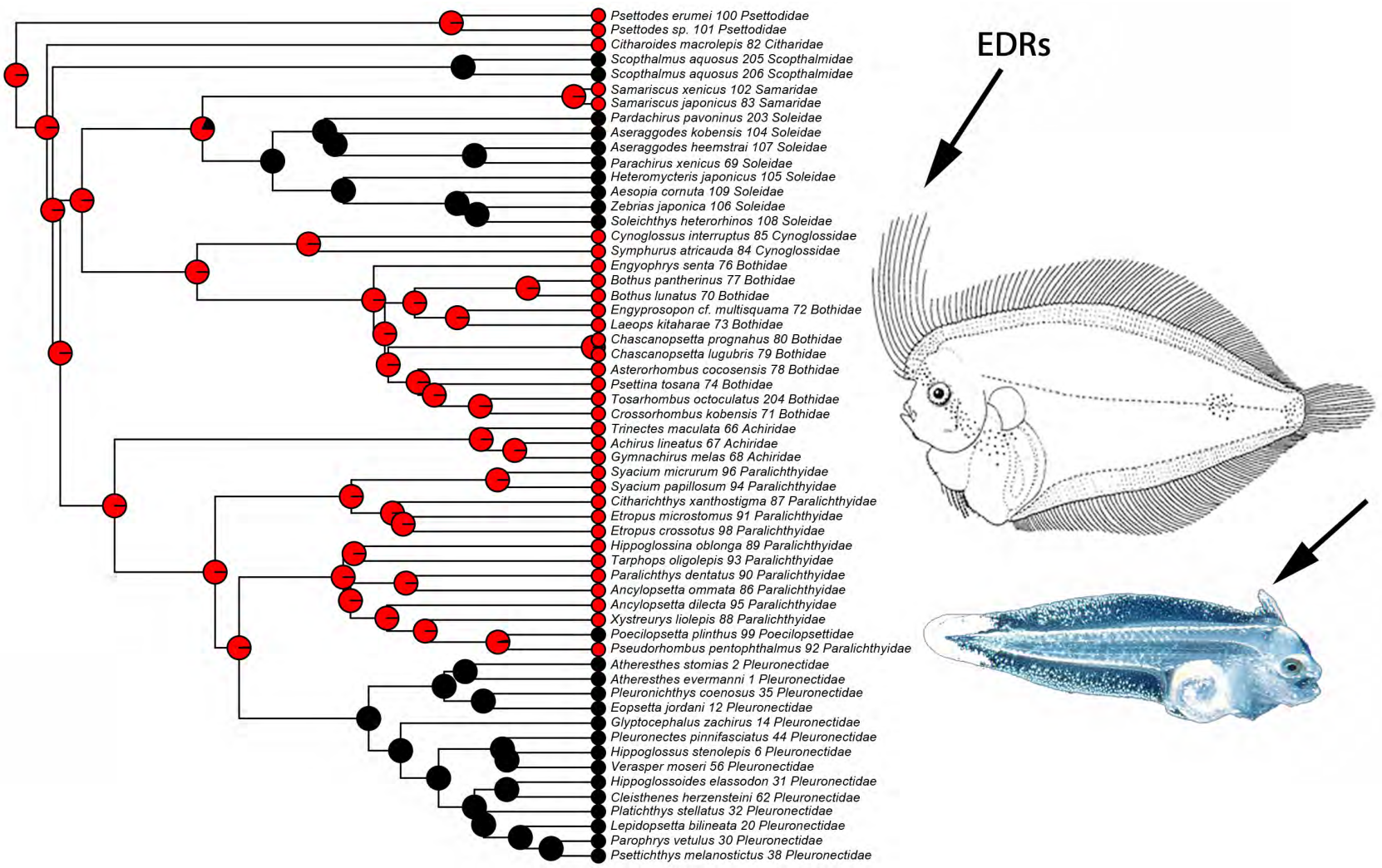


Figure 4-12. Ancestral character state estimation (ACE) of the larval character, elongate dorsal rays (0 = absent, 1 = present). *Parabothus* is top illustration reproduced from Fig. 1 in Tsukamoto et al. (1991). Black is character state 0 and red is state 1.

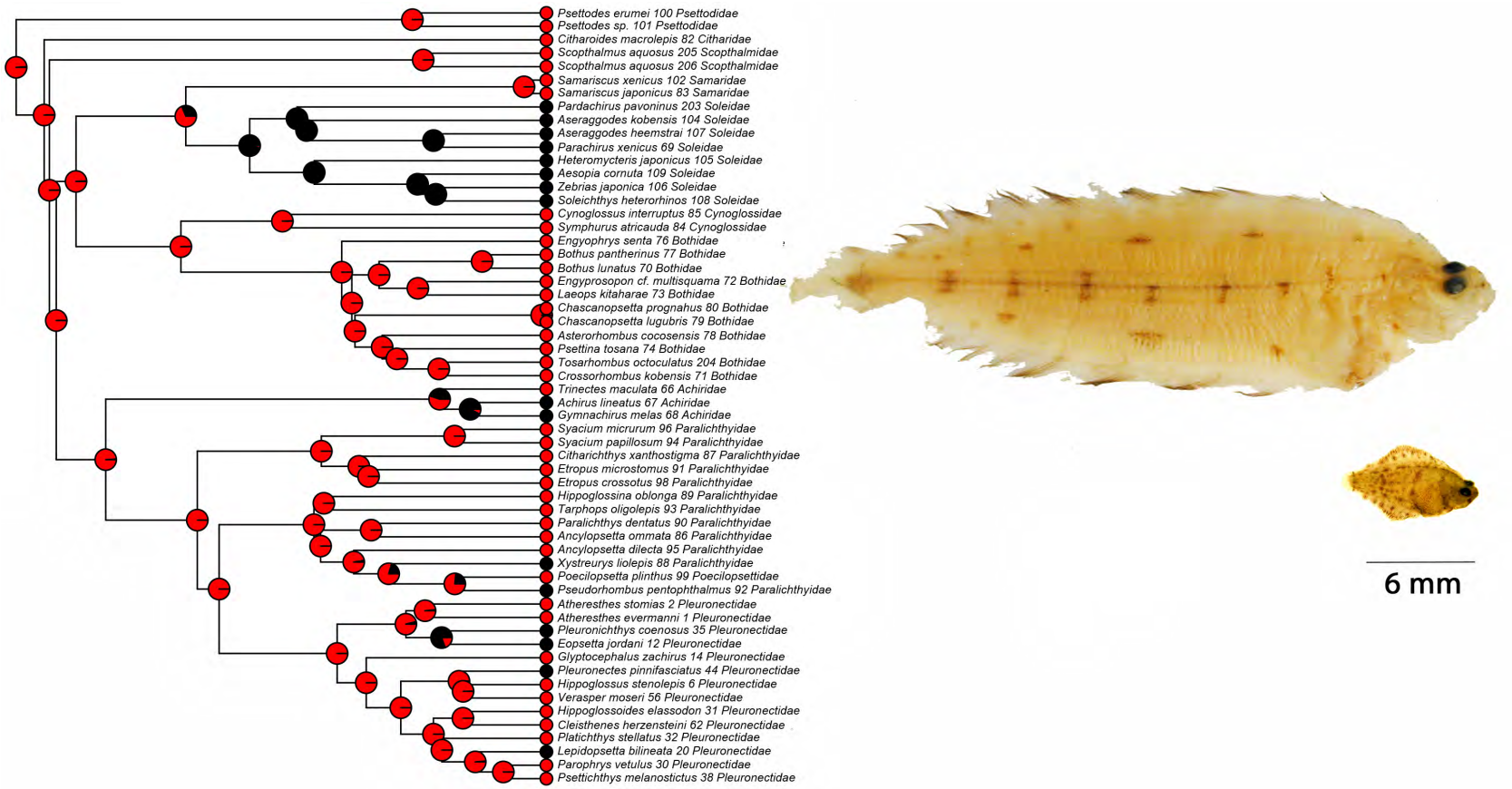
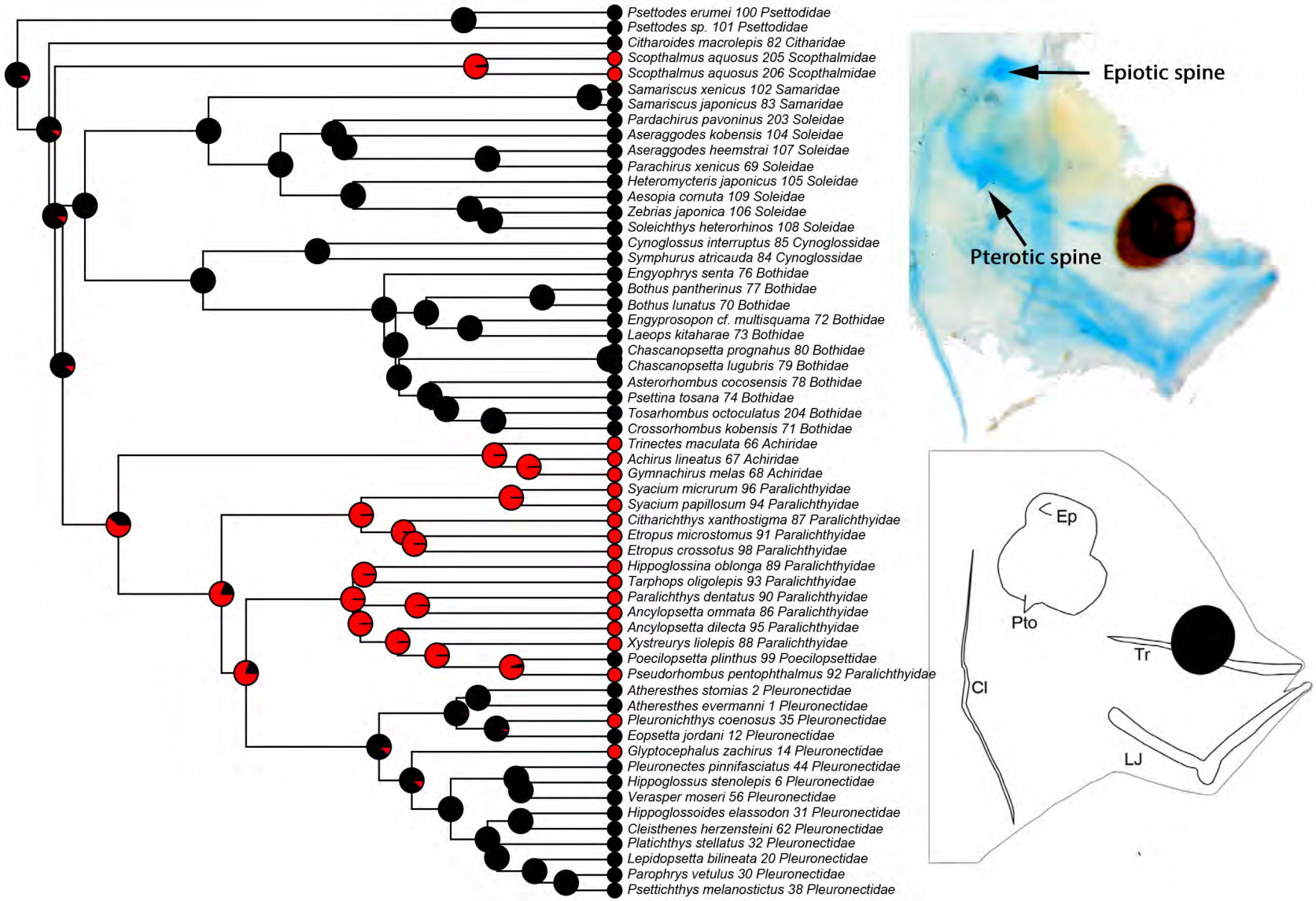


Figure 4-13. Ancestral character state estimation (ACE) of the larval character, size at transformation (0 = ≤ 9mm, 1 = > 9 mm); photo with scale of two similar stage pleuronectids: a large transforming *Glypotocephalus zachirus* (top) and a small *Pleuronichthys* sp. (bottom). Black is character state 0 and red is state 1.

Figure 4-14. Ancestral character state estimation (ACE) of the larval character, otic spination (0 = absent, 1 = present). The same *Embassichthys bathybius* specimen is on top and bottom; illustration is reproduced from Roje (2010), Fig. 5. Black is character state 0 and red is state 1.


Psettodes erumei 100 Psettodidae
Psettodes sp. 101 Psettodidae
Citharoides macrolepis 82 Citharidae
Scopthalmus aquosus 205 Scopthalmidae
Scopthalmus aquosus 206 Scopthalmidae
Samariscus xenicus 102 Samaridae
Samariscus japonicus 83 Samaridae
Pardachirus pavoninus 203 Soleidae
Aseraggodes kobensis 104 Soleidae
Aseraggodes heemstrai 107 Soleidae
Parachirus xenicus 69 Soleidae
Heteromycteris japonicus 105 Soleidae
Aesopia cornuta 109 Soleidae
Zebrias japonica 106 Soleidae
Soleichthys heterorhinos 108 Soleidae
Cynoglossus interruptus 85 Cynoglossidae
Symphurus atricauda 84 Cynoglossidae
Engyophrys senta 76 Bothidae
Bothus pantherinus 77 Bothidae
Bothus lunatus 70 Bothidae
Engyprosopon cf. multisquama 72 Bothidae
Laeops kitaharae 73 Bothidae
Chascanopsetta prognahus 80 Bothidae
Chascanopsetta lugubris 79 Bothidae
Asterorhombus cocosensis 78 Bothidae
Psettina tosana 74 Bothidae
Tosarhombus octoculatus 204 Bothidae
Crossorhombus kobensis 71 Bothidae
Trinectes maculata 66 Achiridae
Achirus lineatus 67 Achiridae
Gymnachirus melas 68 Achiridae
Syacium micrurum 96 Paralichthyidae
Syacium papillosum 94 Paralichthyidae
Citharichthys xanthostigma 87 Paralichthyidae
Etropus microstomus 91 Paralichthyidae
Etropus crossotus 98 Paralichthyidae
Hippoglossina oblonga 89 Paralichthyidae
Tarphops oligolepis 93 Paralichthyidae
Paralichthys dentatus 90 Paralichthyidae
Ancylopsetta ommata 86 Paralichthyidae
Ancylopsetta dilecta 95 Paralichthyidae
Xystreurys liolepis 88 Paralichthyidae
Poecilopsetta plinthus 99 Poecilopsettidae
Pseudorhombus pentophthalmus 92 Paralichthyidae
Atheresthes stomias 2 Pleuronectidae
Atheresthes evermanni 1 Pleuronectidae
Pleuronichthys coenosus 35 Pleuronectidae
Eopsetta jordani 12 Pleuronectidae
Glyptocephalus zachirus 14 Pleuronectidae
Pleuronectes pinnifasciatus 44 Pleuronectidae
Hippoglossus stenolepis 6 Pleuronectidae
Verasper moseri 56 Pleuronectidae
Hippoglossoides elassodon 31 Pleuronectidae
Cleisthenes herzensteini 62 Pleuronectidae
Platichthys stellatus 32 Pleuronectidae
Lepidopsetta bilineata 20 Pleuronectidae
Parophrys vetulus 30 Pleuronectidae
Psettichthys melanostictus 38 Pleuronectidae
Epiotic spine
Pterotic spine
Ep
Pto
Cl
Tr
LJ

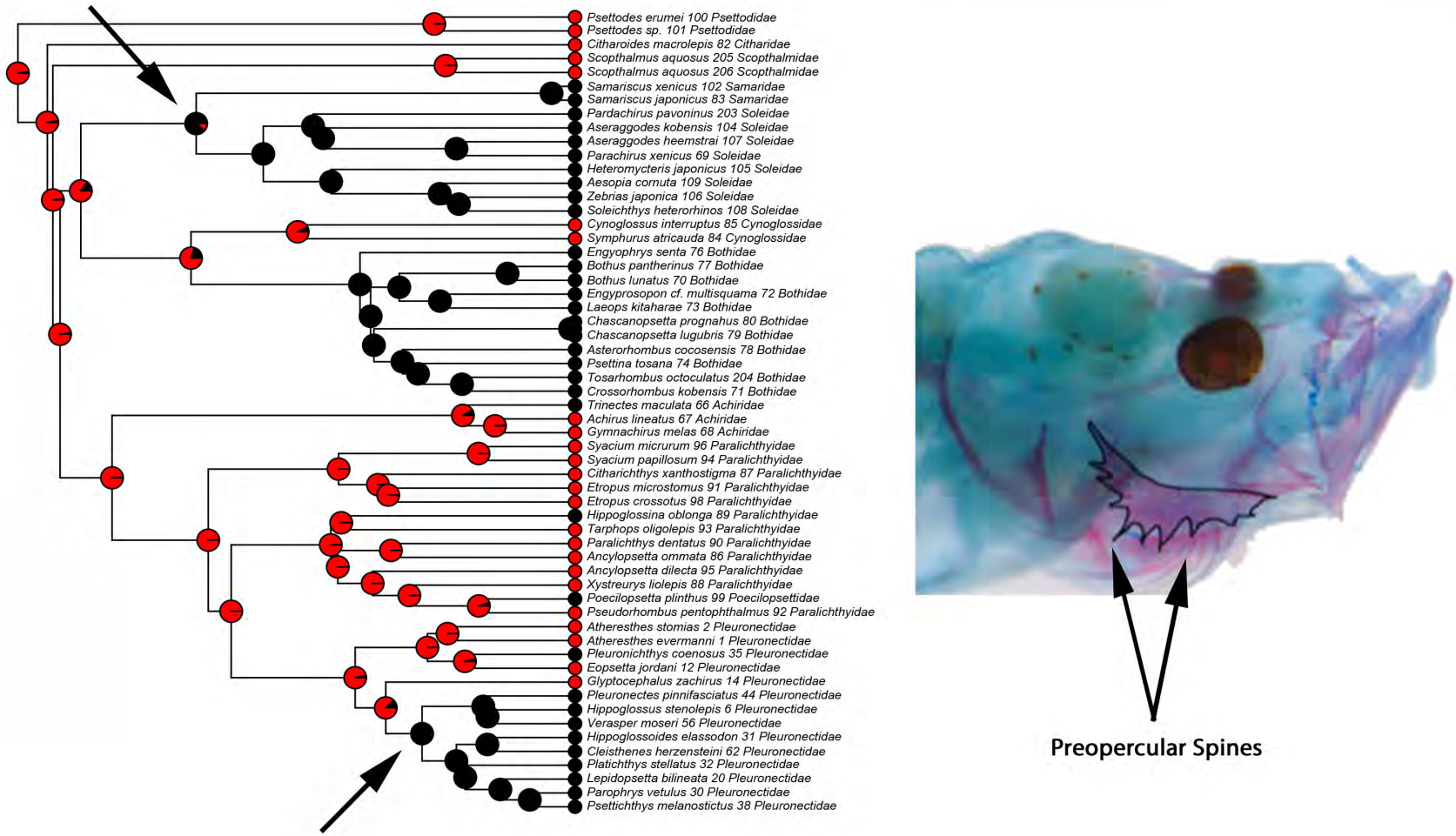


Figure 4-15. Ancestral character state estimation (ACE) of the larval character, preopercular spination (0 = absent, 1 = present). To the right of the ACE is the right side of a cleared and stained *Eopsetta jordani* larva with preopercular spines. Black is character state 0 and red is state 1.

changed across the topology, with a high percentage of the likelihood going to the new state. Changes occurred at the node uniting the Samaridae + Soleidae for eight adult and three larval characters. Cynoglossidae + Bothidae was also supported apomorphically by two adult characters, as were the *Paralichthys* & *Pseudorhombus* group + Pleuronectidae clade and the larger clade with all of "Paralichthyidae" + Pleuronectidae. Monophyly of the *Cyclopsetta* group was supported by the change of state in one larval character: EPR. Inferred character states and distributions are summarized in Table 4-3.

Table 4-3. Summary of characters that resulted in a change of state with ≥75% of the likelihood at that ancestral node.

Clade	**Character**	
	Adult	Larval
Samaridae + Soleidae	14,15,17,19,20,26,27,28	POSp,FrSp,G
Cynoglossidae + Bothidae	26,28	
Cyclopsetta group monophyly		EPR
(*Paralichthys* & *Pseudorhombus* groups) + Pleuronectidae	18,25	
Aciridae +Pleuronectidae + "Paralichthyidae"	27,28	

Overall, the marginal probability (reported as a proportional (–lnL) likelihood) from the ML analyses of the ACEs, as expected based on the D-values, was lowest for the larval characters, with SaT having the lowest likelihood from the ML reconstruction. SaT also had the highest estimated value of the rate parameter and the largest difference between the values from reconstructions with and without branch length information, favoring the no branch length (MP) reconstruction. As Schluter et al. (1997) and others have pointed out, however, the accuracy of ML ACE is low unless the rate of change is low. The condition of low rates appears to be violated by SaT (see Table 4-2), as the likelihood of reconstruction without branch length information is higher (less negative),

but the difference in rate estimates is huge. In fact, SaT shows the lowest likelihoods overall and the greatest differences in rate estimates. In general, this is the case for characters where the –lnL is greater for the MP reconstruction (rows above the black line in Table 4-2).

DISCUSSION

4.1 PHYLOGENETIC RELATIONSHIPS OF PLEURONECTIFORMES AND THE TAXONOMIC STATUS OF PARALICHTHYIDAE

Analysis of the ML phylogeny of the partitioned by gene Con dataset recovered a monophyletic Pleuronectiformes with a very high bootstrap proportion of 92 (Fig. 4-3). This is a novel result, as it is the first time monophyly has been recovered without the removal of data (Betancur et al., 2013b) and with such high statistical support at that node. The Con tree (inferred with the 90 outgroups for comparison) recovered the two major suborders defined by Chapleau (1993): Psettoidei (*Psettodes* spp.) sister to Pleuronectoidei (the remaining flatfishes). Although they were unable to provide strong support for monophyly, these results are consistent with Betancur-R et al. (2013b) as well as Berendzen and Dimmick (2002).

Analysis of the sequence data recovered the only included citharid (*Citharoides macrolepis*) as sister to the remaining pleuronectoids, which is consistent with all the studies described here. There is good support for monophyly of Citharidae (Hoshino, 2001b), and that it is sister to the remaining flatfishes. The next major pleuronectoid lineage recovered in the ML Con tree is Scophthalmidae. This reconstruction is a novel finding, but the statistical support at that node is only moderate (BS = 77) and its position here may be may be an artifact due to low taxon sampling across families (no

rhombosoleids or achiropsettids were included). Much like *Psettodes* when compared to many percomorph outgroups, scophthalmids tend to exhibit varying placements in the pleuronectoid tree and seem highly sensitive to alternative datasets. Most often, Scophthalmidae is recovered somewhere in the middle of the tree (as they were upon analyses of the morphological matrix), a result not entirely inconsistent with the results of the analysis of the sequence data, as support in that part of the tree was low. Regardless, the exact position of Scophthalmidae remains ambiguous, and like *Psettodes*, may be difficult to recover even when more data are analyzed.

The data analyzed here do not support a highly asymmetrical clade, with tonguefishes (Cynoglossidae) sister to the soles (Soleidae). This is contrasted by Betancur-R et al. (2013b) who recover Cynoglossidae + Soleidae with very high statistical support, although not in the most derived position as Chapleau (1993) proposed. That reconstruction, both the sister relationship and their placement within the order, is most likely due to Chapleau's (1993) treatment of the characters as ordered and polarized during coding and analysis. When those characters are combined with the larval ones here and simple Wagner parsimony is applied, cynoglossids and soleids are not recovered together, at least not in the majority-rule consensus.

If the soles and cynoglossids are not sister than who are their respective sister groups? Analysis of the four genes sequenced here recovers Soleidae sister to Samaridae (Fig. 4-3) with high support. Berendzen and Dimmick (2002), whose results were inconclusive regarding the placement of Cynoglossidae, also recovered a highly supported Soleidae + Samaridae. While Chapleau (1993) did not recover this group, the ACEs of eight of his adult characters provide support for Soleidae + Samaridae and two

others support a Cynoglossidae + Bothidae clade, the latter of which is also recovered with high support in the ML Con analysis. Additionally, three larval characters support Samaridae + Soleidae (Table 4-3) even though they were not recovered as sister (but are closely related) in the majority of the six MP trees.

In contrast to the relationships discussed thus far, the inability to recover Paralichthyidae as monophyletic is entirely consistent across all phylogenetic studies of the order. The ML tree inferred from the Con dataset (Fig. 4-4) as well as the results presented by Berendzen and Dimmick (2002) and Betancur-R et al. (2013b), all indicate that Paralichthyidae is not a natural group. Monophyly of the *Cyclopsetta* group is consistently recovered with multiple independent datasets and always with high support. Although Chapleau (1993) suspected this was the case, his choice to treat Paralichthyidae as monophyletic (and as one OTU) when coding his characters essentially ignored that evidence. When his matrix, with the observed states for the *Cyclopsetta* group coded as such, along with the additional larval characters (EPR, elongate pelvic rays in particular), is analyzed it is clear that morphological characters do provide support for the monophyly of this group (see Figs. 4-5 and 4-6).

Because analysis of the four genes sequenced for this study resulted in high statistical support for the monophyly of two paralichthyid groups, as did Betancur-R et al. (2013b), and the *Cyclopsetta* group is additionally supported by 12S, 16S (Berendzen and Dimmick, 2002) and both larval and adult characters, it is my recommendation that the *Cyclopsetta* group be recognized as a separate family, Cyclopsettidae, that is not sister to "Paralichthyidae" (*Paralichthys* + *Pseudorhombus* groups). Even though the type genus, *Cyclopsetta*, was not included here or in the molecular studies described above (the other

three genera were), its inclusion in the family is confirmed by Hensley and Ahlstrom (1984), Chapleau (1993) and Khidir et al. (2005).

The position of the remaining paralichthyids within Pleuronectoidei, however, has not been consistent across different studies. Both Berendzen and Dimmick (2002) and Betancur-R et al. (2013b) recovered these taxa as closely related to a monophyletic Pleuronectidae, but support for their monophyly has been lacking. A monophyletic *Paralichthys* + *Pseudorhombus* group sister to Pleuronectidae was recovered here (Fig 4-4.) and with very high support. This high support and topology is consistent with the favored ML phylogeny of Betancur-R et al. (2013b; Fig. 4-3 here) who had greater taxonomic coverage and much more sequence data than Berendzen and Dimmick (2002). Additional support for their close relationship is provided by two adult characters (Table 4-3), however, neither analysis of the entire 50 character matrix resulted in a (consensus) MP tree or ML tree with a clade comprised only of the remaining paralichthyids sister to Pleuronectidae. This, combined with low taxonomic sampling of the seven genera (approximately 70 species) in all molecular studies carried out thus far requires that the status of this family as natural group remain provisional pending phylogenetic studies that identify additional morphological synapomorphies and/or analyze sequence data representing greater taxonomic diversity.

4.2 Larval Characters as Independent Sources of Data

The results of all pairwise tests for non-independence lead to the rejection of the null (character set independence) suggesting that early life history does not provide an independent set of data and that larval morphology should not be treated as having evolved independently from adult morphology as has been suggested. And this makes

sense, since observed characters and their states (including molecular characters) are ultimately constrained by phylogeny. Alternatively, there are other possible explanations for this result. First, the tendency to code for characters that are parsimony informative (in particular, avoiding autapomorphy) can bias the results. As long as a character is binary and variable it could be analyzed using Pagel's (1994) test modified by Madison and Milford (2006), which would result in a more comprehensive analysis. The exclusion of autapomorphic characters may not explain the decisive results entirely, but their inclusion is a benefit of probabilistic methods that was not exploited here. The second, related, potential alternative explanation is the coding of characters by family. This practice most certainly biased the results, as is evident by the effect recoding of characters for the *Cyclopsetta* group had on the analyses. This was also problematic in the discussion of potentially informative larval characters by Ahlstrom (1984) and Hensley and Ahlstrom (1984) in that they too tended to list and code characters at the family level, albeit not as restrictively as Chapleau (1993). As with ignoring autapomorphy, this biases both the sampling of characters for comparison, as well as the coding.

Third, the requirement to have binary characters for this test makes it difficult to adequately address the confounding issue of multiple developmental stages. One solution would be to code characters by discrete stage and that would certainly increase the amount of larval data, but it may not represent biology since characters can be retained for different periods of time, or appear at slightly different times during development while being truly homologous. Although, assessing homology is dependent on phylogeny as well (de Pinna, 1991) and until viable alternatives to compiling these

matrices are developed and vetted, making homology statements prior to analysis will remain a necessary evil. If enough data is included and analyzed given a complete, fully resolved, tree this problem with stage could be overcome.

Finally, the number of larval characters (11) versus adult (39) provided only 55 unique larval/larval comparisons as opposed to 741 for adult/adult and 429 for the larval/adult tests. Had there been P-values large enough to justify rejection of the null, this asymmetry could make it difficult to substantiate whether life history is, or is not, correlated with phylogenetic signal since only 4% of the tests describe the ability of larval characters to predict the goodness of fit of other larval characters to the independent and dependent models.

Although these results suggest the two types of morphologies should not be treated as having evolved independently, it would be valuable to test this hypothesis for a group where all, or most species, have well described larval stages, along with adult, and a well-supported species-level phylogeny, with enough molecular data to accurately estimate branch lengths. Morphological and/or taxonomic diversity should not be a limiting factor when testing whether life history constrains morphology, but low diversity within a group will likely limit the number of characters for comparison. In addition to focusing on smaller groups, analyzing these characters using more flexible methods, like sensitivity analysis, may also prove insightful. Especially because Pagel's (1994) test, while being robust, is restrictive in that it requires that characters are discrete and binary and there be no ambiguous coding: all cases that most data violate.

4.3 Phylogenic Signal and Ancestral Character Estimation

Interpreting some of Chapleau's (1993) adult characters in light of this new phylogeny does show strong signal where previously there was a high degree of homoplasy, providing additional support for some nodes in the Con tree. Additionally, the recoding of some of those adult characters and the results of MP and ML analyses of the entire morphological dataset highlight the importance of accounting for as much observed variation in character states as possible and for including larval characters in morphological studies of marine teleosts, since resolution was greatly improved compared to Chapleau's (1993) tree.

Although these data maintain strong phylogenetic signal overall (all had negative D-values), the results of the ACEs highlight the necessity to estimate rates of change in order to resolve ancestral nodes with accuracy. When rates are low, the accuracy of ML reconstruction is high, but knowing when this condition is violated can be difficult. The high rate and large difference in estimation of rates between the ML and MP (where change is minimized) conditions indicate that SaT (size at transformation) and possibly another larval character, POSp (preopercular spines), may be violating the low rates condition. It would be valuable to carry out likelihood ratio tests comparing this simple model to more complicated ones to determine the goodness of fit to the data.

Regardless of the accuracy of these ACEs over the entire tree, however, there are still high proportional likelihoods supporting synapomorphies on the Con tree. In particular, the ACE of POSp recovers a state change with high proportional likelihood at the node uniting Samaridae and Soleidae as well as a clade within Pleuronectidae. Unfortunately, this latter case may be misleading as this character is present in taxa that were not included here and the early life history of *Clidoderma* is unknown (Roje, 2010);

it was coded as absent for *Clidoderma* because all characters had to be coded unambiguously and that was the most common state for the pleuronectids sampled.

4.4 Larval Morphology and Sister Group Hypotheses

Synapomorphy providing support for relationships within the order is exciting, but so to is the pleisiomorphy seen in the ACE of some larval characters, particularly at the root. No derived characters of adult morphology unite Pleuronectiformes to any other monophyletic group and molecular data has not been able to recover a sister group to flatfishes. Larval characters, however, have never been proposed as providing resolution, but similarities in the larvae of pleuronectiforms to some adult and juvenile carangoids and adult *Nematistius*—both putative sister groups—imply they may provide the evidence needed.

Elongate first dorsal rays are common among flatfish larvae (Fig. 4-12) and are even retained in some adults, most notably the cockatoo flounder, *Samariscus cristatus*. This morphology is similar to the "cock's comb" dorsal fin of adult *Nematistius* as well as some juvenile carangoids. This character is also present in *Psettodes* and its presence is recovered as pleisiomorphic for the order based on the ACE. Since both Nematistiidae and the much larger Carangoidei remain good candidates for the sister group to Pleuronectiformes this similarity in dorsal fin morphology, regardless of which developmental stage(s) it is observed in, is an indication that larval, juvenile and adult characters may indeed provide useful data to support sister group hypotheses. It's also an indication that this character may be neotenic for carangiforms. Unfortunately, however compelling that hypothesis may seem, testing it is contingent on polarizing the character

that requires a good estimate of its distribution and a well-supported phylogeny, both of which we have yet to obtain.

Another carangid, the monotypic *Parastromateus*, like *Nematistiius,* has been difficult to place phylogenetically even leading to its own designation as a monotypic family (Apolectidae) with unknown affinity (Witzell, 1978), however, its current status a member of Carangidae is supported by the results described in Chapter III (See Figs. 3-3 and 3-4 on pages 47-49) and others (Gushiken, 1978; Hilton et al., 2010; Betancur-R et al., 2013b). Also like *Nematistiius, Parastromateus* shares a character with larval flatfishes. In this case, however, it is larval *Parastromateus* that share the presence of spines on the otic capsule with many larval flatfish taxa (Ahlstrom et al, 1984; Hensley and Ahlstrom; Johnson, 1984). This trait is not uncommon across Pleuronectiformes and is probably unique for *Parastromateus* among carangids (Johnson, 1984). It also appears to be present in at least one larval istiophorifom (another putative sister to flatfishes), but its distribution may be even more widespread. Roje (2010) was only ably to identify the presence of otic spines in a larval flatfish, *Emabssichthys bathybius* (Fig. 4-14), described in the literature as lacking the character, upon clearing and staining. And like many other fish larvae *Embassichthys* had only been described until then from specimens that had not been stained or dissected. Although these similarities are promising leads, pending a study focused on determining the true distribution of these characters across Carangimorpha the potential for larval characters to resolve hypotheses regarding the flatfish sister group will remain in doubt.

REFERENCES

Ahlstrom, E. H., K. Amaoka, D. A. Hensley, H. G. Moser, B. Y. Sumida, and others. 1984. “Pleuronectiformes: Development.” *Ontogeny and Systematics of Fishes* 1: 640–70.

Berendzen, Peter B., and Walter Wheaton Dimmick. 2002. “Phylogenetic Relationships of Pleuronectiformes Based on Molecular Evidence.” *Journal Information* 2002 (3).

Bergsten, Johannes. 2005. “A Review of Long-Branch Attraction.” *Cladistics* 21 (2): 163–93.

Betancur-R, Ricardo, Richard E. Broughton, Edward O. Wiley, Kent Carpenter, J. Andres Lopez, Chenhong Li, Nancy I. Holcroft, et al. 2013. “The Tree of Life and a New Classification of Bony Fishes.” *PLoS Currents* 5 (April). doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, Ricardo, Chenhong Li, Thomas A. Munroe, Jesus A. Ballesteros, and Guillermo Ortí. 2013. “Addressing Gene Tree Discordance and Non-Stationarity to Resolve a Multi-Locus Phylogeny of the Flatfishes (Teleostei: Pleuronectiformes).” *Systematic Biology*, syt039.

Chanet, B. 2003. “Interrelationships of Scophthalmid Fishes (Pleuronectiformes: Scophthalmidae). Cybium 27: 275-286. Polymorphic Microsatellites in Turbot Scophthalmus Maximus.” *Anim. Genet* 27: 433–42.

Chapleau, François. 1993. “Pleuronectiform Relationships: A Cladistic Reassessment.” *Bulletin of Marine Science* 52 (1): 516–40.

Chen, Wei-Jen, Céline Bonillo and Guillaume Lecointre. 2003. “Repeatability of Clades as a Criterion of Reliability: A Case Study for Molecular Phylogeny of Acanthomorpha (Teleostei) with Larger Number of Taxa.” *Molecular Phylogenetics and Evolution* 26 (2): 262–88. doi:10.1016/S1055-7903(02)00371-8.

Cohen, Daniel M. 1984. “Ontogeny, Systematics, and Phylogeny.” *Ontogeny and Systematics of Fishes. Amer. Soc. Ichthyol. Herpetol., Spec. Publ* 1: 7–11.

De Pinna, Mário CC. 1991. “Concepts and Tests of Homology in the Cladistic Paradigm.” *Cladistics* 7 (4): 367–94.

Dettai, Agnès, and Guillaume Lecointre. 2005. “Further Support for the Clades Obtained by Multiple Molecular Phylogenies in the Acanthomorph Bush.” *Comptes Rendus Biologies* 328 (7): 674–89. doi:10.1016/j.crvi.2005.04.002.

Drummond, Alexei J., and Andrew Rambaut. 2007. “BEAST: Bayesian Evolutionary Analysis by Sampling Trees.” *BMC Evolutionary Biology* 7 (1): 214.

Felsenstein, Joe. 2009. “PHYLIP (Phylogeny Inference Package) Version 3.7a.” *Distributed by the Author. Department of Genome Sciences, University of Washington, Seattle.*

Felsenstein, Joseph. 1978. “Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading.” *Systematic Biology* 27 (4): 401–10.

Fritz, Susanne A., and Andy Purvis. 2010. “Selectivity in Mammalian Extinction Risk and Threat Types: A New Measure of Phylogenetic Signal Strength in Binary Traits.” *Conservation Biology* 24 (4): 1042–51.

Gushiken, Soko. 1988. “Phylogenetic Relationships of the Perciform Genera of the Family Carangidae.” *Japanese Journal of Ichthyology* 34 (4): 443–61.

Hensley, D. A., and E. H. Ahlstrom. 1984. “Pleuronectiformes: Relationships.” *Ontogeny and Systematics of Fishes*, 670–87.

Hilton, Eric J., G. David Johnson, and William F. Smith-Vaniz. 2010. “Osteology and Systematics of Parastromateus Niger (Perciformes: Carangidae), with Comments on the Carangid Dorsal Gill-Arch Skeleton.” *Copeia* 2010 (2): 312–33.

Hoshino, Koichi. 2001. “Monophyly of the Citharidae (Pleuronectoidei: Pleuronectiformes: Teleostei) with Considerations of Pleuronectoid Phylogeny.” *Ichthyological Research* 48 (4): 391–404.

Johnson, G. D. 1984. “Percoidei: Development and Relationships. In ‘Ontogeny and Systematics of Fishes’.(Eds HG Moser, WJ Richards, DM Cohen, MP Fahay, AW Kendall Jr and SL Richardson.) Pp. 24–7.” *American Society of Ichthyologists and Herpetologists (Special Publication* 1).

Khidir, K. Teresa, François Chapleau, and Claude B. Renaud. 2005. “Phylogenetic Relationships of *Cyclopsetta* and *Syacium* (Pleuronectiformes: Paralichthyidae).” *Taxon* 1 (10): 11–20.

Lewis, Paul O. 2001. “A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data.” *Systematic Biology* 50 (6): 913–25.

Li, B., A. Dettaī, C. Cruaud, A. Couloux, M. Desoutter-Meniger, and G. Lecointre. 2009. “RNF213, a New Nuclear Marker for Acanthomorph Phylogeny.” *Molecular Phylogenetics and Evolution* 50 (2): 345–63.

López-Martínez, Juana, Eloisa Herrera-Valdivia, Jesús Rodríguez-Romero, and Sergio Hernández-Vázquez. 2010. “Peces de La Fauna de Acompañamiento En La Pesca Industrial de Camarón En El Golfo de California, México.” *Revista de Biología Tropical* 58 (3): 925–42.

Maddison, W. P., and D. R. Maddison. 2011. *Mesquite 2.75: A Modular System for Evolutionary Analysis*.

Midford, P., and W. Maddison. 2006. “Correl Package for Mesquite, Version 0.1.” Website Http://mesquiteproject. Org.

Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. 2010. “Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.” In *Gateway Computing Environments Workshop (GCE), 2010*, 1–8. IEEE.

Moser, HGeoffrey H., Paul E. Smith, and Lawrence E.. Eber. 1987. “Larval Fish Assemblages in the California Current Region, 1954–1960, a Period of Dynamic Environmental Change.” *California Cooperative Oceanic Fisheries Investigations Reports* 28: 97–127.

Near, Thomas J., Alex Dornburg, Ron I. Eytan, Benjamin P. Keck, W. Leo Smith, Kristen L. Kuhn, Jon A. Moore, et al. 2013. “Phylogeny and Tempo of Diversification in the Superradiation of Spiny-Rayed Fishes.” *Proceedings of the National Academy of Sciences* 110 (31): 12738–43.

Nelson, Joseph S. 2006. *Fishes of the World*. John Wiley & Sons.

Orme, David. 2013. “The Caper Package: Comparative Analysis of Phylogenetics and Evolution in R.” *R Package Version 0.5* 2. http://star-www.st-andrews.ac.uk/cran/web/packages/caper/vignettes/caper.pdf.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20 (2): 289–90.

Roje, Dawn M. 2010. “Incorporating Molecular Phylogenetics with Larval Morphology While Mitigating the Effects of Substitution Saturation on Phylogeny Estimation: A New Hypothesis of Relationships for the Flatfish Family Pleuronectidae (Percomorpha: Pleuronectiformes).” *Molecular Phylogenetics and Evolution* 56 (2): 586–600. doi:10.1016/j.ympev.2010.04.036.

Schluter, Dolph, Trevor Price, Arne Ø Mooers, and Donald Ludwig. 1997. “Likelihood of Ancestor States in Adaptive Radiation.” *Evolution*, 1699–1711.

Schwartz, Rachel S., and Rachel L. Mueller. 2010. “Branch Length Estimation and Divergence Dating: Estimates of Error in Bayesian and Maximum Likelihood Frameworks.” *BMC Evolutionary Biology* 10 (1): 5. doi:10.1186/1471-2148-10-5.

Smith, William Leo, and Ward C. Wheeler. 2006. “Venom Evolution Widespread in Fishes: A Phylogenetic Road Map for the Bioprospecting of Piscine Venoms.” *Journal of Heredity* 97 (3): 206–17.

Smith, W.L., and M.T. Craig. 2007. “Casting the Percomorph Net Widely: The Importance of Broad Taxonomic Sampling in the Search for the Placement of Serranid and Percid Fishes.” *Copeia* 2007 (1).

Stamatakis, Alexandros. 2006. “RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models.” *Bioinformatics* 22 (21): 2688–90.

Stamatakis, Alexandros, Paul Hoover, and Jacques Rougemont. 2008. “A Rapid Bootstrap Algorithm for the RAxML Web Servers.” *Systematic Biology* 57 (5): 758–71.

Tsukamoto, Youichi, Atsushi Fukui, and Muneo Okiyama. 1991. “Larvae of the Genus Parabothus (Bothidae) Collected from Japan.” *Ichthyological Research* 37 (4): 414–17.

Witzell, W. N. 1978. “Apolectus Niger (family Apolectidae): Synonymy and Systematics.” *Matsya* 3: 72–82.

CHAPTER V

CONCLUSION

Early in this dissertation I identified the following three problems in pleuronectiform phylogenetics: 1) relationships of the major groups within the order remain mostly unresolved, 2) the sister group of flatfishes is unknown and 3) monophyly of the assemblage is weakly supported. Theses data, analyses and results provide overwhelming support for a monophyletic Pleuronectiformes. While the sister group to flatfishes remains resolved, these studies also provide support, both molecular and morphological, for novel clades within the order.

In the first study, Chapter II, I used previously published sequence data from 78 acanthomorph (including flatfishes) taxa for *rho, rnf213, irbp* and *mll*, performed tests for neutrality, and compared neutral versus non-neutral markers for congruence using tree distance metrics and topology testing. I find that while the signal provided by *rho* may be discordant with others, neutrality alone does not predict congruence and therefore should not be used as a justification to omit data.

In the second study, Chapter III, I optimized new molecular markers and sequence them along with *rho* and *rnf213* for 58 flatfishes and 90 putative outgroups to test monophyly, intraordinal relationships and sister group hypotheses. Those sequences along with data from a previous study were analyzed to determine possible causes for gene tree incongruence or phylogenetic error. I discover that the new markers are variable, providing large amounts of data, while being conserved so that alignment is unambiguous. When those

data are combined with the rest and analyzed simultaneously, they provide overwhelming support for a monophyletic Pleuronectiformes. Additionally, I demonstrate that abundant missing data is likely the cause of low resolution in the EToL study, validate the importance of investigating substitution saturation as a cause of error and discuss asymmetrical taxonomic distribution as a cause of low resolution at the base of Carangimorpha.

Finally, in the third study (Chapter IV) I inferred an ultrametric tree, recoded Chapleau's (1993) matrix of characters of adult morphology and combined them with new larval characters to test whether life history is correlated with phylogenic signal. I then investigated the accuracy of ML ancestral character state estimation (ACE) to determine if these morphological characters provide additional support for hypotheses of relationships among major pleuronectiform groups. My results suggest that larval characters should not be treated as a source of independent data, but do provide resolution and additional support for novel relationships within Pleuronectiformes, although they may be in violation of the condition of low rates on ML ACE. Lastly, I show that because larval characters are mostly pleisiomorphic for the order and that larval morphology is similar to that of putative sister groups, these characters are a potential source of evidence needed to resolve the placement of this lineage within Acanthomorpha.

What is still not understood and is a promising line of research is the developmental mechanism underlying metamorphosis. Recent insights provided by genomics and evo-devo studies suggest that complex and novel morphologies can arise from changes in only one or a few regulatory developmental genes and gene switches (Brakefield and Breuker, 1996; Gompel, 2005; and others). Developmental biologists have established that certain genes in

the nodal-lefty-pitx2 pathway control the formation of the left-right axis in vertebrate embryos (Branford et al., 2000; Cheng et al., 2000; Essner et al., 2000; Yoshioika et al.; 1998) and one gene in this pathway, *pitx2*, has been recognized as a key player in flatfish metamorphosis Suzuki, 2009). The loss of re-expression of *pitx2* in the sinistral flounder *Paralichthys olivaceus* leads to reversed (sinistral becomes dextral and vice versa) or bilateral symmetry (Suzuki, 2009). Clearly this regulatory gene(s) has played a role in the evolution away from bilateral symmetry in the adult, but what remains unclear is what changes have taken place at the sequence level, as well as the selection pressures these genes have been under in the evolution from a bilaterally symmetrical morphology to the novel asymmetrical body plan of flatfishes.

In addition to research focused on molecular mechanisms that generate this bizarre morphology, of great interest is the evolution of novel morphological structures associated with a benthic lifestyle. All flatfishes are lie-in-wait predators (either facultative or obligate) and bury themselves to lower their profile and it is presumed that the function of the recessus orbitalis (RO, an accessory organ of the eye) is to lift the eyes above the plane of the body while the fish is buried so that vision remains unobstructed. Chapleau (1993) even deemed the presence of this structure a synapomorphy uniting the group; however, he did not determine the actual distribution of the RO within Pleuronectiformes, most notably for *Psettodes*. He cited Holt's (1894) and Bishop's (1900) studies focusing on the RO, but even they had only examined the organ in a few taxa, none of which was *Psettodes*. To date, their taxonomically restricted studies are the only morphological studies of the RO.

As a result, many critical questions remain unanswered. Is the RO present in all

flatfish families? Is it morphologically variable within the order? How did this structure evolve, and is it present, for example, in reduced form, in groups closely related to flatfishes? Confounding these issues is the fact that fossil flatfishes that exhibit an intermediate morphology have been described (Friedman, 2008). Since those species do not have a blind side as adults, incomplete migration is difficult to portray as an adaptation to a benthic lifestyle, as vision from that eye would be obstructed when that side made contact with the sea floor. If asymmetry did not evolve for this reason, it is possible that the RO's original function is not related to the function that has been hypothesized for extant flatfishes. Our ability to answer these questions, however, is dependent on rigorous comparative analyses—analyses that are now possible given the great strides in pleuronectiform phylogenetics made in previous studies and those described here.

REFERENCES

Bejder, L., and B.K. Hall. 2002. “Limbs in Whales and Limblessness in Other Vertebrates: Mechanisms of Evolutionary and Developmental Transformation and Loss.” *Evolution and Development* 4 (6): 445–58.

Bishop, N. 1900. *Journal of Anatomy and Physiology*. Griffin.

Brakefield, P.M., and C.J. Breuker. 1996. “The Genetical Basis of Fluctuating Asymmetry for Developmentally Integrated Traits in a Butterfly Eyespot Pattern.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 263 (1376): 1557.

Branford, W.W., J.J. Essner, and H.J. Yost. 2000. “Regulation of Gut and Heart Left-Right Asymmetry by Context-Dependent Interactions between Xenopus Lefty and BMP4 Signaling.” *Developmental Biology* 223 (2): 291–306.

Chapleau, François. 1993. “Pleuronectiform Relationships: A Cladistic Reassessment.” *Bulletin of Marine Science* 52 (1): 516–40.

Cheng, AM, B. Thisse, C. Thisse, and CV Wright. 2000. “The Lefty-Related Factor Xatv Acts as a Feedback Inhibitor of Nodal Signaling in Mesoderm Induction and LR Axis Development in Xenopus.” *Development* 127 (5): 1049.

Essner, J.J., W.W. Branford, J. Zhang, and H.J. Yost. 2000. “Mesendoderm and Left-Right Brain, Heart and Gut Development Are Differentially Regulated by pitx2 Isoforms.” *Development* 127 (5): 1081.

Friedman, Matt. 2008. “The Evolutionary Origin of Flatfish Asymmetry.” *Nature* 454 (7201): 209–12.

Gompel, Nicolas, Benjamin Prud’homme, Patricia J. Wittkopp, Victoria A. Kassner, and Sean B. Carroll. 2005. “Chance Caught on the Wing: Cis-Regulatory Evolution and the Origin of Pigment Patterns in Drosophila.” *Nature* 433 (7025): 481–87.

Holt, EWL. 1894. “Studies in Teleostean Morphology.” In *Proc. Zool. Soc. of London*.

Suzuki, T., Y. Washio, M. Aritaki, Y. Fujinami, D. Shimizu, S. Uji, and H. Hashimoto. 2009. “Metamorphic pitx2 Expression in the Left Habenula Correlated with Lateralization of Eye-Sidedness in Flounder.” *Development, Growth & Differentiation* 51 (9): 797–808.

Yoshioka, H., C. Meno, K. Koshiba, M. Sugihara, H. Itoh, Y. Ishimaru, T. Inoue, et al. 1998. "Pitx2, a Bicoid-Type Homeobox Gene, Is Involved in a Lefty-Signaling Pathway in Determination of Left-Right Asymmetry." *Cell* 94 (3): 299–305.