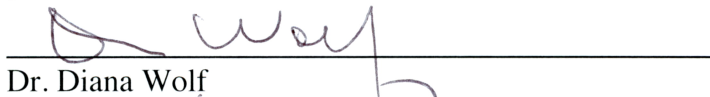
POLYPLOIDY, BASE COMPOSITION BIAS, AND INCOMPLETE LINEAGE SORTING IN FISH PHYLOGENETICS

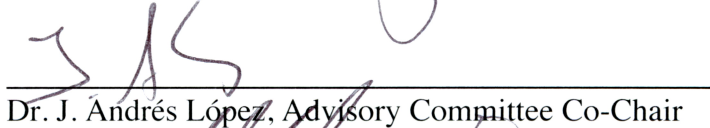
By

Matthew A. Campbell

RECOMMENDED:

Dr. John Rhodes

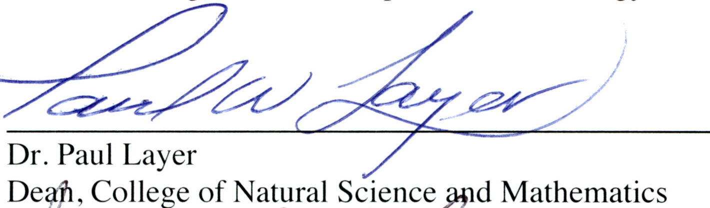

Dr. Diana Wolf

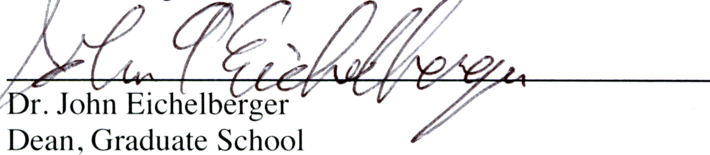

Dr. J. Andrés López, Advisory Committee Co-Chair

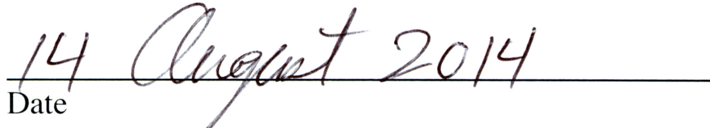
Dr. Naoki Takebayashi, Advisory Committee Co-Chair

Dr. Diane Wagner, Chair, Department of Biology and Wildlife

APPROVED:


Dr. Paul Layer
Dean, College of Natural Science and Mathematics


Dr. John Eichelberger
Dean, Graduate School



Date

POLYPLOIDY, BASE COMPOSITION BIAS, AND INCOMPLETE LINEAGE SORTING IN FISH PHYLOGENETICS

A

DISSERTATION

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Matthew A. Campbell, B.A., M.Sc.

August 2014

Abstract

Understanding the evolutionary relationships between organisms is of fundamental importance in biology. Originally based on overall similarity in morphological traits, depiction of evolutionary relationships is now often pursued by constructing trees based on molecular data- molecular phylogenetics. Molecular phylogenetic inference uses variation in molecular data in a variety of frameworks to produce hypothetical relationships between organisms. As with many practices making use of biological data, the inherent noise and complexity challenges phylogeneticists. In this dissertation, I examine three empirical datasets while addressing three possible issues in phylogenetic inference: polyploidy, base composition bias and incomplete lineage sorting. Polyploidy leads to incorrect genes (paralogs) being analyzed, since it is often impossible to distinguish between gene copies generated as a result of polyploidization. My analysis indicates that incorrect assumptions of orthology have led to incorrect conclusions being drawn from phylogenetic studies including the polyploid salmons (Salmoniformes). Results indicate that pikes (Esociformes) and the polyploid salmons are not only sister taxa, but that the graylings (Thymallinae) and whitefishes (Coregoninae) are most closely related to each other. Base composition bias misleads inference through the overall similarity between sequences being a result of changes in base composition, not shared evolutionary history. Incomplete lineage sorting refers to the fact that the reconstructed relationships of different genes do not agree. Genetic variants may persist through speciation events and are not completely "sorted" between lineages, and require a methodology to reconcile the different genealogies. In two chapters I focused on base composition bias and incomplete lineage sorting in a detailed study of flatfish (Pleuronectiformes) origins. A major issue in fish

phylogenetics is the question of whether flatfish are monophyletic with poor support from both morphological and molecular data. Often it appears that cranial asymmetry is the only characteristic uniting the group. I found very little evidence for a single evolutionary origin of the extant flatfishes. Base composition bias appears not to be a major contributor to flatfish non-monophyly; however incomplete lineage sorting likely results in the inability to generate robust statistical support for inferred relationships of flatfishes and relatives. Results of my work indicate that more care should be exercised in phylogenetics in determining orthology of genes. I also find that not acknowledging the presence of paralogs does indeed mislead analyses. With increased data availability and computational capabilities, non-neutral models of nucleotide evolution should be developed and included in further studies. Presenting the heterogeneity of datasets and actively accounting for incomplete lineage sorting will definitively improve the field of phylogenetics as well.

Table of Contents

Page

Signature Page i

Title Page iii

Abstract v

Table of Contents vii

List of Figures ix

List of Tables xi

Acknowledgements xiii

Chapter 1 Introduction 1

- Phylogenetics 1
- Dissertation Research and Organization 5
- Authorship Statement 7
- References 8

Chapter 2 Pike and Salmon as Sister Taxa: Detailed Intraclade Resolution and Divergence Time Estimation of Esociformes + Salmoniformes Based on Whole Mitochondrial Genome Sequences 11

- Abstract 11
- Introduction 12
- Materials and Methods 13
- Results 18
- Discussion 20
- Conclusion 24
- Acknowledgements 25
- References 33

Chapter 3 Are Flatfishes (Pleuronectiformes) Monophyletic? 41

Abstract 41
Introduction 43
Materials and Methods 49
Results 55
Discussion 58
Acknowledgements 63
References 74

Chapter 4 Mitochondrial Genomic Investigation of Flatfish (Pleuronectiformes)

Monophyly 87

Abstract 87
Introduction 88
Materials and Methods 90
Results 92
Discussion 93
Conclusions 96
Acknowledgements 96
References 104

Chapter 5 Conclusion 111

Conclusion 112
References 121

List of Figures

Page

Chapter 2

Figure 2.1: Maximum-likelihood (ML) phylogenetic tree of 93 actinopterygiian taxa 26

Figure 2.2A: Fossil calibrated phylogeny of Salmoniformes + Esociformes generated under a Bayesian relaxed clock model in BEAST with salmonid calibration..................27

Figure 2.2B: Fossil calibrated phylogeny of Salmoniformes + Esociformes generated under a Bayesian relaxed clock model in BEAST without salmonid calibration............28

Chapter 3

Figure 3.1: A maximum-likelihood (ML) tree generated under a GTR+I+Γ model of sequence evolution in RAxML, depicting phylogenetic positions of the flatfishes (Pleuronectiformes) (taxa within the white rectangle boxes) in relation to other percomorph taxa.............64

Figure 3.2: Timetree based on a Bayesian relaxed clock calibrated by fossils and distributions described in Supplementary Table S3.4...65

Chapter 4

Figure 4.1: A maximum-likelihood (ML) tree generated in RAxML version 8.0.0 under a GTR+Γ model of nucleotide evolution...98

List of Tables

Page

Chapter 2

Table 2.1A: Taxa included in study and corresponding GenBank accession numbers 29

Table 2.1B: Taxa included in study and corresponding GenBank accession numbers 30

Table 2.2: Fossil calibrations used in divergence time estimation 31

Table 2.3: Posterior characteristics of selected nodes from a simultaneous Bayesian divergence time and tree search conducted in BEAST 32

Chapter 3

Table 3.1: Summary of key relationships and support values for phylogenetic analyses of different data coding schemes and taxon composition 66

Table S3.1A: Non-pleuronectiform taxa included in this study, the corresponding accession to the tissue (if any), and corresponding GenBank accession numbers for gene sequences 67

Table S3.1B: Pleuronectiform taxa included in this study, the corresponding accession to the tissue (if any), and corresponding GenBank accession numbers for gene sequences 68

Table S3.2: Primers used in this study and their sources 69

Table S3.3A: Taxa that failed at least one X^2 test of base compostion at one gene...70

Table S3.3B: Taxa that failed at least one X^2 test of base compostion at one gene...71

Table S3.3C: Taxa that failed at least one X^2 test of base compostion at one gene...72

Table S3.4: Prior characteristics of calibration points used in divergence time estimation, the taxa whose most recent common ancestor (MRCA) is dated by the calibration point, and the source of the calibration point 73

Chapter 4

Table 4.1A: Pleuronectiform mitogenomes included in this study...........................99

Table 4.1B: Non-pleuronectiform clade L mitogenomes included in this study.......100

Table 4.1C: Non-clade L mitogenomes included in this study.................................101

Table 4.2A: Summary of fifteen of the twenty-three analyses conducted for this study 102

Table 4.2B: Summary of eight of the twenty-three analyes conducted for this study 103

Chapter 5

Table 5.1: Selected references for phylogenetic inference methods and number of citations for each reference from Web of Science 120

Acknowledgements

Support and time from my graduate committee Dr. John Rhodes, Dr. Diana Wolf, Dr. Naoki Takebayashi, and Dr. J. Andrés López were very helpful in the completion of this dissertation. Dr. G. Javier Fochesatto kindly served as the outside examiner for both my comprehensive exam and dissertation defense, and provided many useful comments on the dissertation. My wife Sienna Campbell deserves great credit in bringing this dissertation to a successful end. Financial support provided by Dr. Kevin McCracken through UA Life Science Informatics was critical in making this dissertation happen. I also received support from the Department of Biology and Wildlife, the Institute of Arctic Biology, the NSF East Asia and Pacific Summer Institutes and UA INBRE.

CHAPTER 1: INTRODUCTION

Phylogenetics

How organisms are related to each other is a question commonly asked across biology. Overall similarity in appearance, as used traditionally and reflected in our common name systems (*e.g*. Pacific halibut, Atlantic halibut), is not sufficient for scientific study. Methods of inferring evolutionary relationships have developed in complexity in order to increase accuracy, for instance modern classification indicates California halibut and spotted halibut are not technically halibuts. Modern phylogenetic approaches have determined more appropriate relationships which are reflected in current taxonomy. In this dissertation I employ methods designed to reconstruct the evolutionary relationships between organisms, a field broadly known as phylogenetics.

Throughout this dissertation, phylogenetic trees are generated with DNA sequence data through explicit modeling of nucleotide evolution with maximum likelihood and Bayesian inference models. Sequences used in phylogenetic construction are the same between organisms. They are homologous (i.e., identical in origin) but have diverged due to speciation to become orthologous. Each site of the sequence is commonly treated to be independent of other sites. Modeling DNA evolution is largely based on a general Markov (GM) model. The GM model is heavily parameterized. Three assumptions are made to simplify the GM model; these are stationarity, homogeneity and reversibility. Stationarity is the assumption that the composition of DNA nucleotides (A, G, C, T) will be fixed throughout the tree. Homogeneity defines the probability of change between nucleotides (e.g. A->G) to be the same; often this property is global and is constant across the whole tree. Reversibility is an hypothesis by which the

probability of a base changing to another base and vice versa is the same (A->G occurs at the same probability as G->A). These sets of assumptions greatly reduce the number of parameters required to be estimated.

The DNA based phylogenetic construction methods are not free of flaws and drawbacks. Polyploidy, base composition bias, and incomplete lineage sorting are three major issues that can hinder the accuracy of phylogenetic inference. These issues are addressed throughout the chapters of this dissertation but in particular each chapter focuses on addressing one of these three issues based on an empirical dataset.

Polyploidy refers to the duplication of the entire complement of chromosomes. These whole genome duplications (WGD) were undoubtedly important in the development of vertebrate traits, occurring twice in the common ancestor of all vertebrates and a third time in the common ancestor of all teleost fishes (Sato and Nishida 2010). Teleost fishes represent 99% of the fish species diversity, and among some lineages a fourth WGD is apparent, such as salmons (Salmoniformes: salmons, trouts, charrs, whitefishes, graylings, etc.) (Santini et al. 2009; Sato and Nishida 2010). For phylogenetic estimation, WGDs are problematic resulting in multiple copies of nuclear loci and uncertainty in determining which copies should be compared between organisms. The assumption of orthology is an underpinning of phylogenetic analyses, sequences that are being compared should be the same between organisms. In the event of allopolyploidization two species (1 and 2) hybridize to form a third (3). Species 3 contains the entire genomes of both its parents, for gene A it contains one copy from species 1 and a second copy from species 2. If a single copy of gene A is sequenced from species 3, phylogenetic analyses would place it as more closely related to either species 1 or species 2 depending on the origin of the gene copy. When the allopolyploid lineage speciates (species 3), without

sequencing both gene copies or unambiguously all gene copies of the same parental origin from the descendant polyploids an incorrect phylogeny will be inferred with the origin of each gene copy (e.g. gene A from species 1) determining the relationships observed. Descendants of species 3 will be found to more closely related to species 1 or species 2 than to each other.

The results of incorporating non-orthologous sequences into a phylogenetic inference are varied. When comparing divergences prior to a duplication, it is not so problematic. However, when comparing divergences after a duplication, it is a certain way to generate a flawed hypothesis.

Base composition bias is another issue in phylogenetic inference. Base composition bias affects phylogenetic inference by the placement of non-related taxa next to each other in phylogenetic trees driven by overall similarity in DNA sequence composition. The frequencies of the four nucleotide bases (A, G, C, T) in DNA is restricted into a large space with a large amount of variation within genomes and between organisms (Mooers and Holmes 2000). The assumptions of stationarity, reversibility and homogeneity are affected by base composition bias. Such simplifications assume that the equilibrium frequencies of nucleotides (A, G, C, T) are constant throughout the inferred phylogeny and that the probability of change is constant as well. Evidence has been established that nucleotide frequencies non-randomly change between lineages (Akashi et al. 1998; Eyre-Walker 1999; Galtier and Gouy 1995; Mooers and Holmes 2000). Similarity in the frequencies of nucleotides can mislead inference to identify close relationships between taxa which do reflect the true evolutionary history of the taxa (Delsuc et al. 2005; Foster and Hickey 1999; Phillips et al. 2004; Steel et al. 1993).

Compositional biases particular to certain groups are also traits that may be informative in phylogenetics (see Chapter 3), and determining if an inference is misled or not by compositional similarity is a challenge. The practicality of more heavily parameterized models (e.g. Jayaswal et al. 2011; Jayaswal et al. 2005; Jayaswal et al. 2007) remains to be tested. However, there are computer programs which allow standard phylogenetic assumptions to be relaxed such as nhPhyML (Boussau and Gouy 2006; Galtier and Gouy 1998) and p4 (Foster 2004). In Chapter 3 the effects of base composition bias in influencing the results of phylogenetic inference is investigated.

The third major issue in phylogenetic inference detailed in this dissertation is incomplete lineage sorting (ILS). ILS can be defined as multiple gene lineages persisting through speciation events, which can be problematic when it leads to incongruence between trees generated between different loci in the genome, "gene trees." Resolving the separate gene trees into a single species tree that represents the evolutionary history of the organisms in question is a goal in phylogenetics (e. g. Ané et al. 2007; Cranston 2010; Maddison 1997; Pamilo and Nei 1988). The effects of ILS on phylogenetic inference are expected to be greatest when time between speciation events is small and population sizes are large (Pamilo and Nei 1988). Not all datasets have ILS or ILS to such a degree that it hampers inference. However, ILS is most likely to be problematic in cases such as the radiation of carangimorph fishes with both (1) rapid speciation events and (2) large population sizes (Campbell et al. 2014).

Dissertation Research and Organization

The second, third, and fourth chapters of this dissertation address major issues in phylogenetics in three categories: polyploidization, base composition bias, and incomplete lineage sorting. A fifth concluding chapter summarizes the overall findings of the dissertation and places them in context of current phylogenetic resarch.

Chapter 2: "Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences" utilizes whole mitochondrial genome sequences to investigate the following: Is there support for pike and salmon as sister taxa? What is their relationship to other basal euteleost fishes? How old are pike, salmon, and pike+salmon? And, what are the relationships between major lineages of pikes and salmons? Mitochondrial genomes are extremely suitable for the particular investigation since salmon underwent a whole genome duplication in the past resulting in difficulty in assigning orthology with nuclear sequence data and sequenced RNA. Mitochondrial genomes, however are single copy in nature and maternally inherited having a smaller effective population size. The properties of mitochondrial genomes lends themselves towards (1) few problems with orthology, and (2) resolving incomplete lineage sorting associated with a rapid radiation. Consequently the radiation of salmonid subfamilies should be accurately resolved with mitochondrial genome data with respect to orthology although other isseues (i.e. saturation) may affect inference. Mitochondrial genomes were analyzed maximum likelihood and Bayesian phylogenetic frameworks. Fossil calibrated divergence time estimates were conducted in BEAST using a relaxed clock method.

Chapter 3: “Are the flatfishes (Pleuronectiformes) monophyletic?” uses multiple nuclear DNA sequences to identify identify if flatfish represent parallel evolution of a body plan or are descended from a single common ancestor (monophyletic). I used data from six nuclear protein coding genes in numerous phylogenetic analyses. Of particular concern is that the base composition bias typical of some flatfishes is affecting the results of phylogenetic analyses due to model violations. A careful and varied methodological approach was implemented to address base composition bias, the influence of missing data and phylogenetic model choice. Analysis methods included maximum likelihood (ML) in RAxML, and several Bayesian based programs. Phylobayes allowed the implementation of the GTR-CAT model. P4 allowed non-standard phylogenetic model implementation. A fossil calibrated timetree was produced with was the first of it’s kind for Plueronectiformes.

Chapter 4: “Mitochondrial evidence for the evolutionary origins of flatfishes (Pleuronectiformes).” Given that the ability of nuclear gene datasets may be unable to resolve the ILS present in the carangimorph fish radiation (Campbell et al. 2014), an approach using whole mitochondrial genomes to investigate the evolutionary affinity of flatfishes and whether they form a monophyletic assemblage was undertaken. Using newly determined flatfish mitochondrial genomes twenty-three analyses in a ML framework were conducted to evaluate the strength of mitochondrial genome support for flatfish monophyly. As a maternally inherited and haploid data source, the population size of mitochondrial genomes is much smaller than that of nuclear genomes. Consequently, the effects of ILS should be smaller.

Chapter 5: “Concluding Chapter”, summarizes the main findings of the thesis work both in specific detail, but also in a larger picture where implications for phylogenetics have been noted.

Authorship Statement

In all of the dissertation chapters, I am first author and contributed most to each manuscript, including the development of study ideas, their refinement, and their implementation. In Chapter 2, I performed sample collection, DNA extraction, amplification and sequencing. I aligned the DNA sequences, conducted all analyses and generated all figures and tables, and created and refined the manuscript with editorial input from co-authors. In Chapter 3, I amplified and sequenced key lineages of flatfish. I performed all analyses, made tables, and generated drafts of figures, which were refined by my co-authors. I edited and refined the manuscript with editorial input from co-authors. In Chapter 4, I compiled all relevant archived sequences, aligned these and all the newly determined sequences. I conducted all analyses, generated the figures and tables for this chapter, and created and refined the manuscripts with editorial input from co-authors.

References

Akashi, H., R. Kliman, and A. Eyre-Walker. 1998. "Mutation Pressure, Natural Selection, and the Evolution of Base Composition in Drosophila." *Genetica* 102-103 (0): 49–60. doi:10.1023/A:1017078607465.

Ané, C., B. T. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2007. "Bayesian Estimation of Concordance among Gene Trees." *Molecular Biology and Evolution* 24 (2): 412–26. doi:10.1093/molbev/msl170.

Boussau, B., and M. Gouy. 2006. "Efficient Likelihood Computations with Nonreversible Models of Evolution." *Systematic Biology* 55 (5): 756–68.

Campbell, M. A., W.-J. Chen, and J. A. López. 2014. "Molecular Data Do Not Provide Unambiguous Support for the Monophyly of Flatfishes (Pleuronectiformes): A Reply to Betancur-R and Ortí." *Molecular Phylogenetics and Evolution* 75 (0): 149–53. doi:10.1016/j.ympev.2014.02.011.

Cranston, K. 2010. "Summarizing Gene Tree Incongruence at Multiple Phylogenetic Depths." In *Estimating Species Trees: Practical and Theoretical Aspects*, edited by L. L. Knowles and L. S. Kubatko, 129–42. Hoboken, New Jersey: John Wiley & Sons, Inc.

Delsuc, F., H. Brinkmann, and H. Philippe. 2005. "Phylogenomics and the Reconstruction of the Tree of Life." *Nature Reviews in Genetics* 6 (5): 361–75. doi:10.1038/nrg1603.

Eyre-Walker, A. 1999. "Evidence of Selection on Silent Site Base Composition in Mammals: Potential Implications for the Evolution of Isochores and Junk DNA." *Genetics* 152 (2): 675 –683.

Felsenstein, J. 2004. *Inferrring Phylogenies*. Sunderland: Sinauer Associates, Inc.

Foster, P. G. 2004. "Modeling Compositional Heterogeneity." *Systematic Biology* 53 (3): 485–95.

Foster, P. G., and D. A. Hickey. 1999. "Compositional Bias May Affect Both DNA-Based and Protein-Based Phylogenetic Reconstructions." *Journal of Molecular Evolution* 48 (3): 284–90. doi:10.1007/PL00006471.

Galtier, N., and M. Gouy. 1995. "Inferring Phylogenies from DNA Sequences of Unequal Base Compositions." *Proceedings of the National Academy of Sciences* 92 (24): 11317 – 11321.

———. 1998. "Inferring Pattern and Process: Maximum-Likelihood Implementation of a Nonhomogeneous Model of DNA Sequence Evolution for Phylogenetic Analysis." *Molecular Biology and Evolution* 15 (7): 871 –879.

Jayaswal, V., F. Ababneh, L. S. Jermiin, and J. Robinson. 2011. "Reducing Model Complexity of the General Markov Model of Evolution." *Molecular Biology and Evolution*, May. doi:10.1093/molbev/msr128. http://mbe.oxfordjournals.org/content/early/2011/05/17/molbev.msr128.abstract.

Jayaswal, V., L. S. Jermiin, and J. Robinson. 2005. "Estimation of Phylogeny Using a General Markov Model." *Evolutionary Bioinformatics Online* 1: 62–80.

Jayaswal, V., J. Robinson, and L. Jermiin. 2007. "Estimation of Phylogeny and Invariant Sites under the General Markov Model of Nucleotide Sequence Evolution." *Systematic Biology* 56 (2): 155 –162. doi:10.1080/10635150701247921.

Maddison, W. P. 1997. “Gene Trees in Species Trees.” *Systematic Biology* 46 (3): 523–36. doi:10.1093/sysbio/46.3.523.

Mooers, A. O., and E. C. Holmes. 2000. “The Evolution of Base Composition and Phylogenetic Inference.” *Trends in Ecology & Evolution* 15 (9): 365–69. doi:doi: DOI: 10.1016/S0169-5347(00)01934-0.

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

Phillips, M. J., F. Delsuc, and D. Penny. 2004. “Genome-Scale Phylogeny and the Detection of Systematic Biases.” *Molecular Biology and Evolution* 21 (7): 1455 –1458. doi:10.1093/molbev/msh137.

Santini, F., L. Harmon, G. Carnevale, and M. Alfaro. 2009. “Did Genome Duplication Drive the Origin of Teleosts? A Comparative Study of Diversification in Ray-Finned Fishes.” *BMC Evolutionary Biology* 9 (1): 194.

Sato, Y., and M. Nishida. 2010. “Teleost Fish with Specific Genome Duplication as Unique Models of Vertebrate Evolution.” *Environmental Biology of Fishes* 88 (2): 169–88. doi:10.1007/s10641-010-9628-7.

Steel, M. A., P. J. Lockhart, and D. Penny. 1993. “Confidence in Evolutionary Trees from Biological Sequence Data.” *Nature* 364 (6436): 440–42. doi:10.1038/364440a0.

Chapter 2: Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes+Salmoniformes based on whole mitochondrial genome sequences.[1]

Abstract

The increasing number of taxa and loci in molecular phylogenetic studies of basal euteleosts has brought stability in a controversial area. A key emerging aspect to these studies is a sister Esociformes (pike) and Salmoniformes (salmon) relationship. We evaluate mitochondrial genome support for a sister Esociformes and Salmoniformes hypothesis by surveying many potential outgroups for these taxa, employing multiple phylogenetic approaches, and utilizing a thorough sampling scheme. Secondly, we conduct a simultaneous divergence time estimation and phylogenetic inference in a Bayesian framework with fossil calibrations focusing on relationships within Esociformes + Salmoniformes. Our dataset supports a sister relationship between Esociformes and Salmoniformes; however the nearest relatives of Esociformes + Salmoniformes are inconsistent amongst analyses. Within the order Esociformes, we advocate for a single family, Esocidae. Subfamily relationships within Salmonidae are poorly supported as Salmoninae sister to Thymallinae + Coregoninae.

[1]Campbell, M.A, Lopez, J.A.L., Sado T., and Miya M. 2013. Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes+Salmoniformes based on whole mitochondrial genome sequences. Gene 530:57–65.

Introduction

A consensus on the evolutionary relationships among basal euteleost lineages is emerging as a result of increasing numbers of both taxa and loci represented in molecular data sets. Results from these studies are beginning to identify stable patterns of relationships between a set of lineages whose affinities have been controversial area since the inception of Euteleostei (Greenwood et al., 1966). Protacanthopterygian (Rosen, 1974) relationships have been examined in multiple phylogenetic studies relying on evidence from morphological and molecular traits (Begle, 1992, 1991; Diogo et al., 2008; Fink and Weitzman, 1982; Fink, 1984; Ishiguro et al., 2003; Johnson and Patterson, 1996; Lauder and Liem, 1983; López et al., 2004; Patterson, 1994; Rosen, 1982; Sanford, 1990; Williams, 1987). And, while a sister group relationship between Salmoniformes and Esociformes is broadly supported by analyses based on the suspensorium and associated musculature (Williams, 1987; Wilson and Williams, 2010), mitochondrial genome data (Ishiguro et al., 2003; Li et al., 2010), nuclear sequence data (López et al., 2004; Near et al., 2012; Santini et al., 2009), and combined nuclear and mitochondrial data (Burridge et al., 2012; López et al., 2004), the placement of the Esociformes + Salmoniformes clade among basal euteleost lineages remains problematic.

Mitochondrial genome (mitogenome) sequences from 33 teleost species provide evidence for a sister group relationship between esociforms and salmoniforms; however these two lineages were only represented with one species each in that analysis (Ishiguro et al. 2003). In this study, we expand the sampling of protacanthopterygiians to 93 species with the addition of five newly determined mitogenome sequences and a targeted selection of previously published sequences designed to help test existing ideas on basal euteleost relationships. Specifically, we determined

mitogenome sequences from two salmoniform and three esociform species. Increased taxon sampling is known to improve phylogenetic inference (Hedtke et al., 2006; Hillis, 1998; Hillis et al., 2003; Pollock et al., 2002), and to enhance the ability to infer macroevolutionary processes from a phylogenetic tree (Heath et al., 2008).

Our goals are to test possible placements of the Esociformes + Salmoniformes clade among basal euteleost lineages and to generate a hypothesis of intra-ordinal relationships within the Esociformes and Salmoniformes. Within esociforms we test whether the family Umbridae (Nelson, 2006) is a monophyletic group containing the genera *Umbra*, *Novumbra*, and *Dallia*; and within salmoniforms we examine alternative arrangements of the relationships between the three salmonid subfamilies and among the genera of Salmoninae. Finally, we also estimate timing of major cladogenetic events in the history of the esociform + salmoniform group. We use a maximum likelihood (ML) framework to infer a mitochondrial genome phylogeny for the 93 taxa considered here and a Bayesian-based joint tree inference and divergence time estimation procedure on a 35 species taxonomic subset to focus on the intra-ordinal history of the esociform + salmoniform clade.

Materials and Methods

Taxonomic Sampling

Sampling for novel mitogenome sequence determinations targeted unrepresented lineages within Salmoniformes and Esociformes (Table 1). Species were selected to divide long branches to reduce possible long branch generated artifacts in the phylogenetic inference (Hillis,1998). We newly determined five mitogenomes for this study: *Novumbra hubbsi*, *Umbra pygmaea*, and

Esox niger (Esociformes) and *Prosopium cylindraceum* and *Parahucho perryi* (Salmoniformes). The newly determined mitogenome sequences are available on Genbank as accessions AP013046-AP013050. Additional mitogenome sequences were obtained from GenBank guided by the goal of testing the placement of Salmoniformes and Esociformes among basal euteleost lineages.

DNA Extraction, PCR Amplification and Sequencing

We extracted DNA from ethanol-preserved fin clips using Qiagen DNEasy or QIamp tissue kits following manufacturer instructions. Mitogenome sequences were determined using a combination of long and short PCR amplifications (Miya & Nishida 1999). Briefly, whole mitogenomes of target organisms were first amplified using long PCR (Cheng et al., 1994). Long PCR amplicons were diluted in TE buffer and used as templates for a series of short PCRs that produced a set of overlapping fragments covering the mitochondrial genome. Short PCR products were purified using the ExoSAP protocol and sequenced with ABI Big-Dye v1.1 chemistry on an ABI 3130XL automated sequencer.

DNA Sequence Assembly and Alignment

DNA sequences were examined and edited using EditView version 1.0.1, AutoAssembler version 2.1 and DNASIS ver. 3.2. Existing mitogenome sequences were retrieved from GenBank (Benson et al., 2005). Protein coding and RNA loci were extracted from GenBank flatfiles with GenBankStrip.pl versions 2.0 (Bininda-Emonds, 2005). Two separate alignments were generated. An alignment with 93 species including thirteen salmoniform and five esociform representatives was generated to estimated the phylogenetic placement of Esociformes

and Salmoniformes among basal Euteleost lineages. To generate this alignment, protein-coding genes were each imported into MacClade version 4.08 (Maddison and Maddison, 2000) and translated to amino acids. The amino acid sequences were aligned using MAFFT version 6.814 (Katoh and Toh, 2008; Katoh et al., 2002) then merged with nucleotide sequence files in MacClade and gaps removed to produce a statistically consistent alignment. The mitochondrial gene NADH-ubiquinone oxidoreductase chain 6 (ND6) was excluded due to heterogeneous base composition. 12S and 16s ribosomal RNA (rRNA) sequences were aligned using ProAlign version 5.3 (Löytynoja and Milinkovitch, 2003) with a 70% posterior probability limit on site homology. Additional gaps were removed by hand from the rRNA alignments, which were subsequently concatenated. Transfer RNA (tRNA) sequences were individually aligned with MUSCLE version 3.8.31 (Edgar, 2004a, 2004b), then imported into Mesquite version 2.71(Maddison and Maddison, 2009) and edited by hand.

A second alignment for evaluating intraordinal relationships and divergence times was generated by excluding some outgroup taxa and increasing Esociformes+Salmoniformes representation. The reduced alignment consisting of five esociform, seventeen salmoniform and twelve euteleost outgroups (34 taxa) was generated following the alignment procedure described above.

Phylogenetic Placement of Esociformes + Salmoniformes

Phylogenetic placement of Salmoniformes and Esociformes was estimated by maximum likelihood (ML) search implemented in RAxML version 7.3.0 (Stamatakis, 2006). The general time reversible model (GTR) with a four-category gamma distributed rate variation among sites (Γ) model of DNA evolution was used. 1,000 bootstrap replicates were used to evaluate the

support for different aspects of the optimal topology. In this analysis, third codon position sites were recoded as purines and pyrimidines (RY) to reduce the potential effect of substitution saturation on phylogenetic inference. This coding scheme is noted as $1_n2_n3_{RY}R_nT_n$, where subscripts indicate RY or nucleotide (n) coding for each category of sites, numbers denote codon positions for sites within protein-coding regions, R refers to ribosomal RNA coding sites and T indicates transfer RNA coding sites. To characterize the effect of variations in mutation rate among sites, the CAT-GTR model (Lartillot and Philippe, 2004) as implemented in PhyloBayes version 3.3b (Lartillot et al., 2009) was used on the 93-taxon alignment with three coding schemes ($1_n2_nR_nT_n$, $1_n2_n3_nR_nT_n$, $1_n2_n3_{RY}R_nT_n$).

Simultaneous Bayesian Phylogenetic Inference and Divergence Time Estimation

We performed Bayesian phylogenetic inference and divergence time estimation on the 35-taxon dataset with five data partitions ($1_n2_n3_{RY}R_nT_n$), and a Bayesian relaxed clock with uncorrelated lognormal rate heterogeneity as implemented in BEAST version 1.7.4 (Drummond et al., 2012, 2006). An input tree was generated from a partitioned alignment using the HKY+Γ model of sequence evolution with a proportion of invariant sites. We calibrated the root of the tree using the known appearance of euteleost and ostariophysan fish in the fossil record at a minimum of 149.85 million years ago (Ma). Strong evidence exists to constrain this node at 165.2 Ma (Benton et al., 2009). A strict molecular clock was used to generate the input tree with a Markov chain Monte Carlo (MCMC) chain of 50 million generations sampled every 5,000 generations. We applied a 10% burnin and used Tracer v 1.5 to examine MCMC output and quality of parameter sampling (Drummond et al., 2012). Subsequently the input tree was used to initialize the divergence time analysis. We used lognormal fossil constraint distributions which

produce more conservative estimates of divergence times due to the underlying assumption that the fossil record can inform maximum and minimum divergences of some clades in the analysis (Lavoué et al., 2011).

For each calibration point, a fossil record was used as a hard minimum bound, with upper bounds considered and applied on a case by case basis (Table 2). Fossil aulopiforms provide well supported constraints with both stem and crown representation, constraining the age of this node to between 96 and 128 Ma (Benton, 1993; Kriwet, 2003; Santini et al., 2009). Based on age of crown representatives, the origin of Acanthomorpha and Beryciformes was constrained to between 70 and 99 Ma, respectively (Benton, 1993; Dirk, 2004).

The following fossil calibrations specific to the Esociformes and Salmoniformes were used: (1) *Esteesox*, a stem esociform from the late Cretaceous (Wilson et al., 1992) as the minimum age of Esociformes at 85 Ma; and (2) *Esox kronneri,* the first record of the subgenus *Kenoza* from the late early Eocene (Grande, 1999) as a minimum bound for the divergence between *E. lucius* and *E. niger* at 42 Ma. The genus *Novumbra* was present by the Oligocene (Cavender, 1969) however, because this first appearance is much more recent than the evidence for *Kenoza*, it was not used as a minimum bound for the divergence of *Novumbra* from *Esox*. The taxonomic affinities of older fossils associated with Umbridae such as *Boltyshia* from the Ypresian (Benton, 1993; Syŝevskaâ and Daniltšenko, 1975) remain poorly resolved (Nelson, 2006). Due to that uncertainty, those records are not included in this analysis.

The earliest definitive fossil evidence of a salmoniform comes from fossils of *Eosalmo driftwoodensis* from middle Eocene lacustrine deposits (Wilson, 1977). *Eosalmo* is considered a stem salmonin (Wilson and Li, 1999; Wilson and Williams, 1992). We constrained the

minimum date of the origin of Salmonidae at 51.8 Ma (Greenwood et al., 2005; Near et al., 2012). Alternate placements for this fossil exist, such as dating the most recent common ancestor of Coregoninae and Salmoninae (Crête-Lafrenière et al., 2012). Therefore we examined effects of the *Eosalmo* calibration were examined through an alternative analysis with this calibration point omitted.

For the four data partitions ($1_n2_nR_nT_n$) we used the GTR+Γ+I model of nucleotide evolution. Three independent runs of 100 million generations sampled every 5,000 generations were generated. After verifying adequate sampling (ESS > 200) and convergence with Tracer, we applied a 10% burnin and combined the tree files with LogCombiner. Finally, we used TreeAnnotator to calculate a maximum clade credibility tree, mean values of divergence times, posterior probabilities, and bounds for the 95% highest posterior density (HPD) interval.

Results

Sequencing

We sequenced complete or nearly complete mitochondrial genomes of *Prosopium cylindraceum*, *Parahucho perryi*, *Novumbra hubbsi*, *Umbra krameri,* and *Esox niger*. The mitochondrial control regions contained repeating motifs and were not sequenced completely in some taxa. Gene content and order in the newly determined mitochondrial genomes follow the standard arrangement found in most vertebrates.

Esociform and Salmoniform Phylogenetic Relationships

The Esociformes+Salmoniformes clade is supported in the ML topology using the $1_n2_nR_nT_n$ and $1_n2_n3_{RY}R_nT_n$ codings with boostrap values of 99 and 100 (Figure 2.1). Support for

Lepidogalaxias salamandroides as the most basal Euteleost is supported with a bootstrap value of 93 and 98 using $1_n2_nR_nT_n$ and $1_n2_n3_{RY}R_nT_n$ codings, respectively. Among esociforms, *Umbra* is sister group to a clade formed by the remaining three esociform genera, and *Novumbra* and *Esox* are sister lineages. Among salmoniforms, there is weak support for a sister relationship between Coregoninae and Thymallinae under the $1_n2_nR_nT_n$ coding scheme (35% bootstrap). In contrast, with the $1_n2_n3_{RY}R_nT_n$ scheme, the Thymallinae + Salmoninae clade is strongly supported (100% bootstrap). Convergence occurred in PhyloBayes using CAT-GTR only when third codon position sites were excluded ($1_n2_nR_nT_n$), and not under any coding schemes that included those sites. In the PhyloBayes analysis, a posterior probability of 0.99 is assigned to the Esociformes+Salmoniformes clade. The topology: (*Lepidogalaxias salamandroides*, ((Esociformes+Salmoniformes), (remaining euteleosts)) was supported by this analysis. Strong support for this branching pattern is observed with a posterior probability of 0.96 for the placement of *Lepidogalaxias salamandroides*, 1.00 for support of Esociformes+Salmoniformes, and 0.99 for the Esociformes+Salmoniformes as sister clade to all other euteleosts.

Intraordinal Relationships and Divergence Time Estimation

The divergence time estimation analysis based on the 35 species alignment with the *Eosalmo* calibration point included yields a divergence time for the Esociformes+Salmoniformes from other euteleost lineages of 124.99 Ma (Fig 2.2a, Table 2.3). The divergence between Esociformes and Salmoniformes is estimated to be 113.02 Ma. As in all other analyses, the Esociformes+Salmoniformes clade is strongly supported (1.00 posterior probability). The mean divergence estimate between *Umbra* and the *Esox* + *Novumbra* + *Dallia* clade is 88.61 Ma. Monophyly of both esociforms (1.00 posterior probability) and the *Esox* + *Novumbra* + *Dallia*

clade are strongly supported (1.00 posterior probability). Major salmonid lineages originate within the last 55.19 million years, with a sister Thymallinae and Coregoninae relationship strongly supported (1.00 posterior probability). The estimated divergence between Coregoninae and Thymallinae is 47.42 Ma. The age of Salmoninae is estimated to be 33.87 Ma.

Removing the *Eosalmo* calibration point produced a divergence time of Salmoniformes+Esociformes from other euteleost lineages of 120.09 Ma and a divergence between Esociformes + Salmoniformes of 106.03 Ma (Fig 2.2b, Table 2.3). The mean estimated ages for time to most recent common ancestor of salmonids is 40.28 Ma. Thymallinae and Coregoninae are strongly supported as sister taxa (1.00 posterior probability) with a mean estimated divergence time of 34.59 Ma. The origin of Salmoninae is estimated to be 27.72 Ma.

Discussion

Phylogenetic Placement of the Esociformes+Salmoniformes

Results of both full and reduced taxon set analyses reported here further strengthen the case for a sister group relationship between esociforms and salmoniforms (López et al., 2004, 2000). All our analyses invariably support a sister relationship of Esociformes and Salmoniformes. Among the euteleosts, the placement of *Lepidogalaxias* as the sister group of all other euteleost is in agreement with mitogenomic (Li et al., 2010), combined nuclear and mitochondrial data (Burridge et al., 2012), and with multilocus nuclear data (Near et al. 2012). We recover five clades of Euteleosts (excluding *Lepidogalaxias*) with high support: Esociformes+Salmoniformes, Argentiformes, Osmeriformes+Stomiiformes, Galaxiids, and the neoteleosts. Relationships among these five clades is unstable in our analyses, and consequently

so is the sister group of the Esociformes+Salmoniformes. The sister of Esociformes+Salmoniformes is inferred to be all remaining euteleost fishes (less *Lepidogalaxias*) in this study with a 93 taxa $1_n2_nR_nT_n$ data scheme analyzed under both ML and Bayesian frameworks. A similar relationship was demonstrated by Burridge et al. (2012). However, under ML and using a $1_n2_n3_{RY}R_nT_n$ coding scheme for that same taxon set results in Esociformes+Salmoniformes sister to a clade of Osmeriformes+Stomiiformes and Argentiformes. In the simultaneous Bayesian divergence time estimation and phylogenetic inference of a 34-taxon $1_n2_nR_nT_n$ alignment, Esociformes+Salmoniformes is sister to the Argentiformes without strong support (posterior probability of 0.65 or 0.85). Stronger support for a sister relationship of Argentiformes to the Esociformes+Salmoniformes was found by Li et al. (2010) and Near et al. (2012).

Relationships within Esociformes and Salmoniformes

Among esociforms, all our analyses support the (*Umbra*, (*Dallia*, (*Novumbra*, *Esox*))) topology with a monophyletic *Esox* previously advanced based on molecular evidence (Burridge et al., 2012; Grande et al., 2004; López et al., 2004). This hypothesis is incongruent with the morphology based hypothesis (e.g. Wilson and Veilleux, 1982) that serves as the basis of currently accepted classification schemes for esociform taxa, but is in agreement with the morphological hypothesis of Wilson and Williams (2010). A classification congruent with relationships based on more recent morphological and molecular evidence would require alteration of the generic composition of the families Esocidae and Umbridae. We propose the redefinition of the Esocidae to be coextensive with the order Esociformes and abandonment of the Umbridae. If taxonomic classification is to reflect best understanding of phylogenetic

relationships, no compelling argument remains to preserve current usage of the two esociform families.

Within salmoniforms, some of our analyses yield high support for a sister group relationship between Coregoninae and Thymallinae. Previous analyses based on mitogenomic sequences did not sample the genus *Prosopium*. Li et al. (2010) found with the inclusion of *Thymallus* and *Coregonus*, moderate support for this relationship with ML (76% boostrap) and high support from Bayesian analyses (1.00 posterior probability). However, in another mitogenomic study with two representatives of *Thymallus*, Thymallinae was found to be more closely related to Salmoninae (Yasuike et al., 2010). Results of a single nuclear locus phylogenetic analysis of the Salmonidae support a Salmoninae + Thymallinae clade (Shedko et al., 2012). Alternatively, multilocus nuclear data and combined mitochondrial and nuclear data support Coregoninae+Salmoninae (Crête-Lafrenière et al., 2012; Near et al., 2012) or Thymallinae+Coregoninae (Burridge et al., 2012). The morphologically-based hypothesis of salmonid relationships (Sanford, 1990; Wilson and Williams, 2010) groups Thymallinae and Salmoninae in a clade that is sister group to the coregonins. If these relationships remain labile under more extensive trait and taxonomic sampling, the lack of agreement may prove to be the result of a rapid salmonid radiation into the three subfamilies.

Divergence Time Estimation

Living and fossil esociforms and salmoniforms are restricted to northern hemisphere landmasses. Given this distribution it is interesting to ask whether or not the timing of origin of the group or the orders coincides with key events in the evolution of the northern hemisphere

geography. The 95% HPD interval for divergence between Esociformes+Salmoniformes and Argentiformes in our study is contained in the early Cretaceous. Our estimate of divergence time between Esociformes and Salmoniformes corresponds to the boundary between the Aptian and Albian of the Cretaceous (Walker and Geismann, 2009). Roughly, the 95% HPD for Esociformes and Salmoniformes divergence spans the younger half of the Early Cretaceous. During that period, the Atlantic Ocean was beginning to form and Eurasia and North America were well separated during the Early Cretaceous (Vullo et al., 2012). It is unlikely that the breakup of Laurasia was a vicariant event marking the split of esociforms and salmoniforms as it happened much earlier than our estimates of this divergence.

Both the ages of Esociformes and Salmoniformes are constrained by fossil calibration points in this study. The age of Salmonidae is constrained by use of *Eosalmo* to date the MRCA of all three salmonid subfamilies. The characters which support the placement of *Eosalmo* as sister to extant salmonins also support a Thymallinae and Salmoninae sister relationship (Wilson and Li, 1999). The contradictory molecular support for ((Coregoninae, Thymallinae), Salmoninae) indicates that an alternative placement of the fossil for calibration may be appropriate or that it should be excluded. The age of the origin of Salmonidae is forced by the *Eosalmo* calibration to be at least 51.8 Ma. Alternatively, if *Eosalmo* is used to constrain the age of a subfamily or two subfamilies, the estimated origin of Salmonidae will be older as in Crête-Lafrenière et al. (2012). By excluding the *Eosalmo* calibration point from the analysis we removed the assumptions required to place the fossil. The age of the Salmonidae was estimated to be 27.0 % younger without a fossil calibration included for this group. Consequently, a more rapid diversification of salmonid lineages is inferred. Regardless of how the *Eosalmo* evidence

is treated, the 95% HPD intervals for the time to MRCA of Esociformes and of Salmoniformes do not overlap and support a smaller time to MRCA for salmoniforms.

The Esociformes and Salmoniformes broadly overlap in distribution and have evolved under similar conditions. A key difference between the two orders is an ancestral polyploidization event in the salmoniform lineage. Salmoniforms also show markedly higher extant species diversity than esociforms. Our data and analyses suggest a markedly higher rate of species accumulation in salmoniforms. Future estimations of age of divergence in the two groups without relying on internal calibration points and incorporating nuclear data will be needed to more precisely compare their diversification rates.

Conclusion

Our results add to the emerging consensus on basal euteleost relationships in which Esociformes and Salmoniformes are sister lineages. Given the stability of this relationship, it may be appropriate at this time to identify an appropriate name for the Esociformes+Salmoniformes clade. A possible solution is to modify the limits of Salmoniformes to encompass both groups, abandon Esociformes and treat the two major lineages in the newly defined salmoniforms as the families Esocidae and Salmonidae. Regardless of nomenclatural choices, the relevant relationships reported here and elsewhere are backed by ample evidence and are consistently supported thus it is advisable to adopt a classification scheme that accurately reflects them. Concerning intraordinal relationships, our analyses support esociform monophyly and the generic inter-relationships proposed by López et al. (2000; 2004). Among salmoniforms, subfamily inter-relationships remain unresolved using mitogenomic data.

Acknowledgements

We would like to thank the following individuals in sample collections: Molly Hallock (Washington Department of Fish and Wildlife), Joseph Buckwalter (Alaska Department of Fish and Game), and Motohiro Kikuchi (Chitose Salmon Park). A gift of tissue was provided by the Academy of Natural Sciences of Philadelphia (ANSP 189305). Robert Marcotte provided helpful comments on a draft manuscript and Sébastien Lavoué provided the geologic time bar used in Figure 2 and assistance with the BEAST analysis. Wataru Iwasaki kindly annotated the sequences for submission. M. Campbell was supported by the joint US NSF EAPSI and Japan JSPS Summer Program (NSF OISE 1015583). Analysis was partially supported by a US NSF grant (DEB 0963767) to J. López. Additional support was provided by grants-in-aid from the Ministry of Education, Culture, Sports, Science, and Technology, Japan to M. Miya (17207007 and 22580229). Computational support was provided by UA Life Science Informatics, Grant Number RR016466 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).

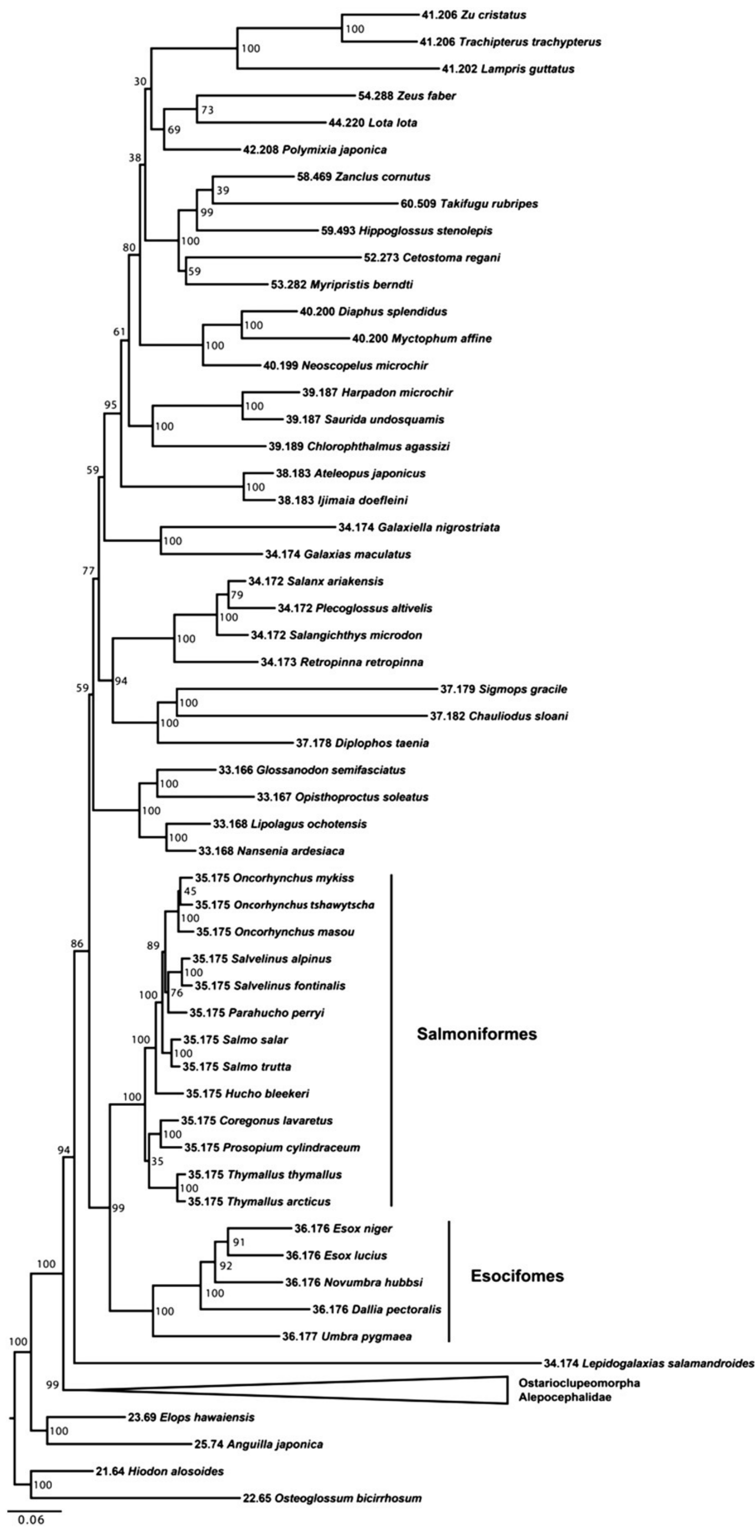


Figure 2.1: Maximum-likelihood (ML) phylogenetic tree of 93 actinopterygiian taxa. Analysis is based on a $1_n2_nR_nT_n$ data partition and coding scheme (details in text). Bootstrap values are shown as node labels.

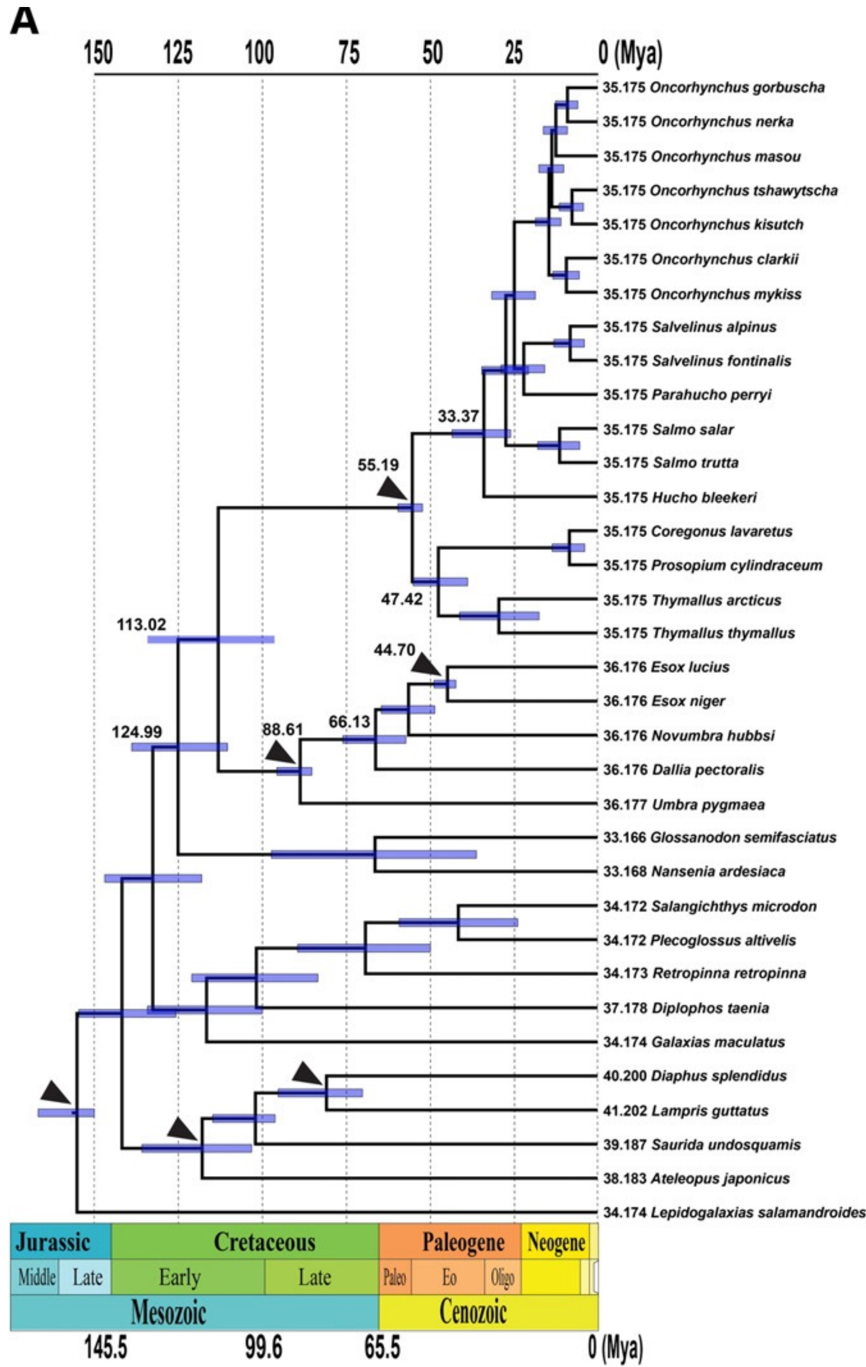


Figure 2.2A Fossil calibrated phylogeny of Salmoniformes + Esociformes generated under a Bayesian relaxed clock model in BEAST with salmonid calibration. 95% HPD intervals are shown as blue bars at nodes. Contains *Eosalmo* as a calibration point for the origin of Salmonidae. Calibration points are indicated by black triangles and twelve outgroup taxa are included.

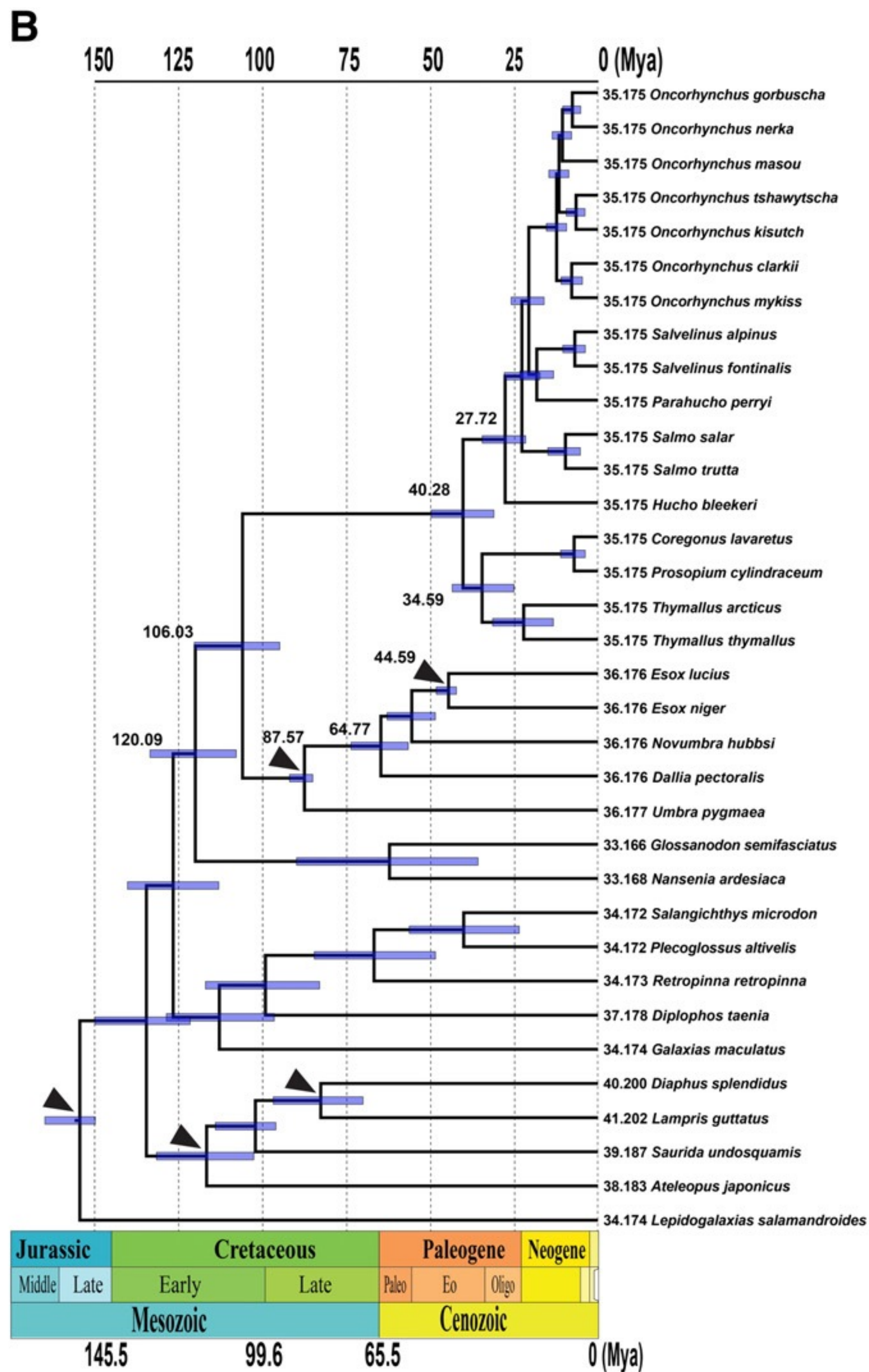


Figure 2.2B: Fossil calibrated phylogeny of Salmoniformes + Esociformes generated under a Bayesian relaxed clock model in BEAST without salmonid calibration. 95% HPD intervals are shown as blue bars at nodes. A tree is shown in which there is no calibration in salmonid lineages. Calibration points are indicated by black triangles and twelve outgroup taxa are included.

Table 2.1A: Taxa included in this study and corresponding GenBank accession numbers. Classification follows Nelson (2006).

		Order or Suborder	Family or Subfamily	Organism	Accession Number
Division	Teleostei				
Subdivision	Osteoglossomorpha				
		Hiodontiformes	Hiodontidae	*Hiodon alosoides*	AP004356
		Osteoglossifomres	Osteoglossidae	*Osteoglossum bicirrhosum*	AB043025
Subdivision	Elopomorpha				
		Elopiformes	Elopidae	*Elops hawaiensis*	AB051070
		Anguilliformes	Anguillidae	*Anguilla japonica*	AB038556
Subdivision	Ostarioclupeomorpha				
		Clupeiformes	Denticipitidae	*Denticeps clupeiodes*	AP007276
			Pristigasteridae	*Pellona flavipennis*	AP009619
			Engraulidae	*Engraulis japonicus*	AB040676
			Chirocentridae	*Chirocentrus dorab*	AP006229
			Clupeidae	*Sardinops melanostictus*	AB032554
		Gonorynchiformes	Chanidae	*Chanos chanos*	AB054133
			Gonorynchidae	*Gonorynchus greyi*	AB054134
				Gonorynchus abbreviatus	AP009402
			Kneriidae	*Cromeria nilotica*	AP011560
				Grasseichthys gabonensis	AP007277
				Kneria sp.	AP007278
				Parakneria cameronensis	AP007279
			Phractolaemidae	*Phractolaemus ansorgii*	AP007280
		Cypriniformes	Cyprinidae	*Cyprinus carpio*	AP009047
				Sarcocheilichthys variegatus	AB054124
			Gyrinocheilidae	*Gyrinocheilus aymonieri*	AB242164
			Catostomidae	*Catostomus commersonii*	AB127394
			Cobitidae	*Lefua echigonia*	AB054126
			Balitoridae	*Schistura balteata*	AB242172
		Characiformes	Distichontidae	*Distichodus sexfasciatus*	AB070242
			Chilodontidae	*Chilodus punctatus*	AP011984
			Alestiidae	*Phenacogrammus interruptus*	AB054129
			Characidae	*Chalceus macrolepidotus*	AB054130
			Lebiasinidae	*Lebiasina astrigata*	AP011995
		Siluriformes	Diplomystoidea	*Diplomystes nahuelbutaensis*	AP012011
			Amphiliidae	*Amphilius sp.*	AP012002
			Callichthyidae	*Corydoras rabauti*	AB054128
			Loricariidae	*Pterygoplichthys disjunctivus*	AP012021
			Bagridae	*Pseudobagrus tokiensis*	AB054127
			Pimelodidae	*Pimelodus pictus*	AP012019
		Gymnotiformes	Gymnotidae	*Electrophorus electricus*	AP011978
			Hypopomidae	*Brachyhypopomus pinnicaudatus*	AP011570
			Sternopygidae	*Eigenmania virescens*	AB054131
			Apteronotidae	*Apteronotus albifrons*	AB054132
Subdivision	Euteleostei				
Superorder	Protacanthopterygii				
		Argentiformes			
		Argentinoidei	Argentinidae	*Glossanodon semifasciatus*	AP004105
			Opisthoproctidae	*Opisthoproctus soleatus*	AP004110
			Microstomatidae	*Nansenia ardesiaca*	AP004106
			Bathylagidae	*Bathylagus ochotensis*	AP004101
		Alepocephaloidei	Platytroctidae	*Platytroctes apus*	AP004107
				Maulisia mauli	AP009404
			Alepocephalidae	*Alepocephalus tenebrosus*	AP004100
				Narcetes stomias	AP009585

Table 2.1B: Taxa included in this study and corresponding GenBank accession numbers. Classification follows Nelson (2006) except Esociformes follow López et al. (2004).

	Osmeriformes			
	Osmeroidei	Osmeridae	*Plecoglossus altivelis*	AB047553
			Salangichthys microdon	AP004109
			Salanx ariakensis	AP006231
		Retropinnidae	*Retropinna retropinna*	AP004108
		Galaxiidae	*Galaxias maculatus*	AP004104
			Galaxiella nigrostriata	AP006853
			Lepidogalaxias salamandroides	HM106490
	Salmoniformes	Salmonidae		
		Coregoninae	*Coregonus lavaretus*	AB034824
			Prosopium cylindraceum	This study.
		Thymallinae	*Thymallus arcticus*	FJ872559
			Thymallus thymallus	FJ853655
		Salmoninae	*Hucho bleekeri*	HM804473
			Oncorhynchus clarkii	AY886762
			Oncorhynchus gorbuscha	EF455489
			Oncorhynchus keta	EF105341
			Oncorhynchus kisutch	EF126369
			Oncorhynchus masou	DQ864465
			Oncorhynchus mykiss	DQ288268
			Oncorhynchus nerka	EF055889
			Oncorhynchus tshawytcha	AF392054
			Parahucho perryi	This study.
			Salmo salar	U12143
			Salmo trutta	AM910409
			Salvelinus alpinus	AF154851
			Salvelinus fontinalis	AF154850
	Esociformes	Umbridae	*Umbra pygmaea*	This study.
		Esocidae	*Dallia pectoralis*	AP004102
			Esox lucius	AP004103
			Esox niger	This study.
			Novumbra hubbsi	This study.
Neoteleostei				
	Stomiiformes	Diplophidae	*Diplophos taenia*	AB034825
		Gonostomidae	*Sigmops gracile*	AB016274
		Stomiidae	*Chauliodus sloani*	AP002915
	Ateleopodiformes	Ateleopodidae	*Ijimaia doefleini*	AP002917
			Ateleopus japonicus	AP002916
	Aulopiformes	Synodontidae	*Harpadon microchir*	AP002919
			Saurida undosquamis	AP002920
		Chlorophthalmidae	*Chlorophthalmus agassizi*	AP002918
	Myctophiformes	Neoscopelidae	*Neoscopelus microchir*	AP002921
		Myctophidae	*Myctophum affine*	AP002922
			Diaphus splendidus	AP002923
	Lampridiformes	Lampridae	*Lampris guttatus*	AP002924
		Trachipteridae	*Trachipterus trachypterus*	AP002925
			Zu cristatus	AP002926
Superorder Polymixiomorpha				
	Polymixiiformes	Polymixiidae	*Polymixia japonica*	AB034286
Superorder Paracanthopterygii				
	Gadiformes	Gadidae	*Lota lota*	AP004412
Superorder Acanthopterygii				
	Beryciformes	Holocentridae	*Myripristis berndti*	AP002940
	Perciformes	Zanclidae	*Zanclus cornutus*	AP009162
	Pleuronectiformes	Pleuronectidae	*Hippoglossus stenolepsis*	AM749126
	Tetraodontiformes	Tetraodontidae	*Takifugu rubripes*	AP006045
	Stephanoberyciformes	Cetomimidae	*Cetostoma regani*	AP004423
	Zeiformes	Zeidae	*Zeus faber*	AP002941

Table 2.2: Fossil calibrations used in divergence time estimation. Taxonomic order to which calibration point is assigned, taxa included in the analysis of which the most recent common ancestor (MRCA) is dated, and priors assigned to the calibration point are shown. Additional information and source details are also included.

		Prior				
Taxonomic Group	**Dating MRCA of Which Taxa**	**Offset**	**Log(Mean)**	**Log(SD)**	**95%**	**Source and Additional Information**
Esociformes	Esocoidei	85.0	1.0	1.00	99.1	From Masstrichian of Cretaceous (Wilson et al. 1992)
	Esox and *Kenoza* subgenera of *Esox*	42.0	1.0	0.85	53.0	The first record of *Kenoza* from the Eocene (Grande 1999).
Salmoniformes	All Salmonine taxa	51.8	1.618	0.80	70.6	*Eosalmo driftwoodensis* as stem salmonine (Wilson 1977; Wilson & Williams 1992). Calibrated as Near et al. (2012).
Aulopiformes	*Saurida, Diaphus,* and *Lampris*	96.0	1.5	1.20	128.3	Santini et al. (2009).
Lampriformes	*Diaphus* and *Lampris*	70.0	1.2	1.32	99.1	Santini et al. (2009)

Table 2.3: Posterior characteristics of selected nodes from a simultaneous Bayesian divergence time and tree search conducted in BEAST. The results from both the inclusion and exclusion of *Eosalmo* as a calibration point are presented. The time to most recent common ancestor of taxa is present as a mean with 95% highest probability density (HPD) upper and lower bounds. The posterior probability (posterior prob.) of the particular node is also included.

Dating MRCA of Which Taxa	***Eosalmo* calibration included**			
	Posterior			
	Mean	**95% Low**	**95% High**	**Posterior Prob.**
Esociformes+Salmoniformes and Argentiformes	125.07	110.68	139.52	0.61
All Esociformes and Salmoniformes	113.35	97.06	135.27	0.99
All Esociformes	88.66	85.09	95.86	0.99
Esocidae	66.12	56.75	75.99	1.00
Novumbra + *Esox*	56.28	48.64	64.81	1.00
All Salmoniformes	55.28	52.15	59.73	1.00
Thymallinae and Coregoninae	47.33	38.08	55.09	1.00
Coregoninae (*Prosopium* and *Coregonus*)	29.18	16.86	41.86	1.00
Salmoninae	34.32	25.76	43.28	1.00
Oncorhynchus	14.52	10.68	18.73	1.00

	***Eosalmo* calibration excluded**			
	Posterior			
	Mean	**95% Low**	**95% High**	**Posterior Prob.**
Esociformes+Salmoniformes and Argentiformes	120.46	107.31	134.22	0.80
All Esociformes and Salmoniformes	107.22	95.22	124.68	0.99
All Esociformes	87.64	85.09	92.36	0.99
Esocidae	64.72	56.19	73.77	1.00
Novumbra + *Esox*	55.42	48.44	63.24	1.00
All Salmoniformes	41.60	31.52	53.14	1.00
Thymallinae and Coregoninae	35.62	25.83	46.80	1.00
Coregoninae (*Prosopium* and *Coregonus*)	22.87	13.79	32.91	1.00
Salmoninae	28.86	21.86	36.78	1.00
Oncorhynchus	12.96	9.58	16.59	1.00

References

Begle, D.P., 1991. Relationships of the osmeroid fishes and the use of reductive characters in phylogenetic analysis. Syst. Zool. 40, 33–53.

Begle, D.P., 1992. Monophyly and relationships of the argentinoid fishes. Copeia 1992, 350–366.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2005. GenBank. Nucleic Acids Res. 33, D34–D38.

Benton, M.J. (Ed.), 1993. The fossil record 2. Chapman and Hall, London.

Benton, M.J., Donoghue, P.C.J., Asher, R.J., 2009. Calibrating and constraining molecular clocks, in: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, Oxford, pp. 35–86.

Bininda-Emonds, O., 2005. GenBankStrip.pl.

Burridge, C.P., McDowall, R.M., Craw, D., Wilson, M.V.H., Waters, J.M., 2012. Marine dispersal as a pre-requisite for Gondwanan vicariance among elements of the galaxiid fish fauna. J. Biogeogr. 39, 306–321.

Cavender, T., 1969. An Oligocene mudminnow (family Umbridae) from Oregon with remarks on relationships with the Esocoidei. Occas. Pap. Mus. Zoo.l Number 660.

Cheng, S., Chang, S.-Y., Gravitt, P., Respess, R., 1994. Long PCR. Nature 369, 684–685.

Crête-Lafrenière, A., Weir, L.K., Bernatchez, L., 2012. Framing the Salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. PLoS ONE 7, e46662.

Diogo, R., Doadrio, I., Vandewalle, P., 2008. Teleostean phylogeny based on osteological and myological characters. Int. J. Morphol. 26, 463–522.

Dirk, N., 2004. Otolithes de poissons aptiens du Maestrazgo (province de Castellon, Espagne orientale). Bull. Inst. R. Sci. Nat. Belg. Sci. Terre. 74, 101–120.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS. Bio. 4, e88.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–73.

Edgar, R.C., 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Edgar, R.C., 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792 –1797.

Fink, W.L., 1984. Basal euteleosts: relationships, in: Moser, H.G., Richardson, W.J., Cohen, D.M., Fahay, M.P., Kendall Jr., A.W., Richards, S.L. (Eds.), Ontogeny and Systematics of Fishes. Am. Soc. Ichthyol. Herpetol. Spec. Publ., Lawrence, pp. 202–206.

Fink, W.L., Weitzman, S.H., 1982. Relationships of the stomiiform fishes (Teleostei), with a description of *Diplophos*. Bull. Mus. Comp. Zool. 150, 31–92.

Grande, L., 1999. The first *Esox* (Esocidae: Teleostei) from the Eocene Green River Formation, and a brief review of esocid fishes. J. Vertebr. Paleontol. 19, 271–292.

Grande, T., Laten, H., López, J.A., 2004. Phylogenetic relationships of extant esocid species (Teleostei: Salmoniformes) based on morphological and molecular characters. Copeia 2004, 743–757.

Greenwood, D.R., Archibald, S.B., Mathewes, R.W., Moss, P.T., 2005. Fossil biotas from the Okanagan Highlands, southern British Columbia and northeastern Washington State: climates and ecosystems across an Eocene landscape. Can. J. Earth Sci. 42, 167–185.

Greenwood, P.H., Rosen, D.E., Weitzman, S.H., Myers, G.S., 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. B. Am. Mus. Nat. Hist. 131, 339–456.

Heath, T.A., Zwickl, D.J., Kim, J., Hillis, D.M., 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst. Biol. 57, 160–166.

Hedtke, S.M., Townsend, T.M., Hillis, D.M., 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55, 522–529.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52, 124–126.

Ishiguro, N.B., Miya, M., Nishida, M., 2003. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". Mol. Phylogenet. Evol. 27, 476–488.

Johnson, G.D., Patterson, C., 1996. Relationships of lower euteloestean fishes, in: Stiassny, M.L.J. (Ed.), Interrelationships of Fishes. Academic Press, pp. 251–328.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059 –3066.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286 –298.

Kriwet, J., 2003. Lancetfish teeth (Neoteleostei, Alepisauroidei) from the Early Cretaceous of Alcaine, NE Spain. Lethaia 36, 323–331.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286 –2288.

Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095 –1109.

Lauder, G.V., Liem, K.F., 1983. The evolution and interrelationships of the actinopterygian fishes. Bull. Mus. Comp. Zool. 150, 95–197.

Lavoué, S., Miya, M., Arnegard, M.E., McIntyre, P.B., Mamonekene, V., Nishida, M., 2011. Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. Pro. R. Soc. B. 278, 1003–1008.

Li, J., Xia, R., McDowall, R.M., López, J.A., Lei, G., Fu, C., 2010. Phylogenetic position of the enigmatic *Lepidogalaxias salamandroides* with comment on the orders of lower euteleostean fishes. Mol. Phylogenet. Evol. 57, 932–936.

López, J.A., Bentzen, P., Pietsch, T.W., 2000. Phylogenetic relationships of Esocoid fishes (Teleostei) based on partial cytochrome b and 16S mitochondrial DNA sequences. Copeia 2000, 420–431.

López, J.A., Chen, W.J., Ortí, G., Wood, R.M., 2004. Esociform phylogeny. Copeia 2004, 449–464.

Löytynoja, A., Milinkovitch, M.C., 2003. A hidden Markov model for progressive multiple alignment. Bioinformatics 19, 1505 –1513.

Maddison, D.R., Maddison, W.P., 2000. MacClade version 4: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland.

Maddison, W.P., Maddison, D.R., 2009. Mesquite: a modular system for evolutionary analysis.

Miya, M., Nishida, M., 1999. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): First example of rransfer RNA gene rearrangements in bony fishes. Mar Biotechnol 1, 416–26.

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., Smith, W.L., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. P. Natl. Acad. Sci. USA. 109, 13698–13703.

Nelson, J.S., 2006. Fishes of the World, 4th ed. John Wiley & Sons, Inc, Hoboken, New Jersey.

Patterson, C., 1994. Bony fishes, in: Prothero, D.R., Schoch, R.M. (Eds.), Major Features of Vertebrate Evolution. Short Courses in Paleontology. Paleontological Society, Knoxville, pp. 57–84.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51, 664–671.

Rosen, D.E., 1974. Phylogeny and zoogeography of salmoniform fishes and relationships of *Lepidogalaxias salamandroides*. B. Am. Mus. Nat. Hist. 153, 265–326.

Rosen, D.E., 1982. Teleostean interrelationships, morphological function and evolutionary inference. Amer. Zool. 22, 261–273.

Sanford, C.P.J., 1990. The phylogenetic relationships of salmonoid fishes. Bull. Br. Mus. Nat. Hist. (Zool.) 56, 145–153.

Santini, F., Harmon, L., Carnevale, G., Alfaro, M., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Bio. 9, 194.

Shedko, S., Miroshnichenko, I., Nemkova, G., 2012. Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: Analysis of nuclear *RAG1* gene. Russ. J. Genet. 48, 575–579.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688 – 2690.

Syŝevskaâ, E.K., Daniltšenko, P.G., 1975. A new genus of the family Palaeoesocidae from the Lower Paleogene of Ukraine, in: Šimanskij, V.N., Solov'ëv, A.N. (Eds.), Razvitie i smena organičeskogo mira na rubeže mezozoâ i kainozoâ. Izdatel'stvo "Nauka," Moskva, pp. 117–125.

Vullo, R., Marugán-Lobón, J., Kellner, A.W.A., Buscalioni, A.D., Gomez, B., de la Fuente, M., Moratalla, J.J., 2012. A new Crested Pterosaur from the Early Cretaceous of Spain: The first european tapejarid (Pterodactyloidea: Azhdarchoidea). PLoS ONE 7, e38900.

Walker, J.D., Geismann, J.W., 2009. 2009 GSA Geologic Time Scale. GSA Today April/May, 60–61.

Williams, R.R.G., 1987. The phylogenetic relationships of the salmoniform fishes based on the suspensorium and its muscles. (Ph. D.). University of Alberta, Edmonton.

Wilson, M.V.H., 1977. Middle Eocene freshwater fishes from British Columbia. Life Sci. Contr., R. Ont. Mus. 113, 1–61.

Wilson, M.V.H., Brinkman, D.B., Neuman, A.G., 1992. Cretaceous Esocoidei (Teleostei): early radiation of the pikes in North American fresh waters. J. Pal. 66, 839–846.

Wilson, M.V.H., Li, G.-Q., 1999. Osteology and systematic position of the Eocene salmonid †*Eosalmo driftwoodensis* Wilson from western North America. Zool. J. Linn. Soc-Lond. 125, 279–311.

Wilson, M.V.H., Veilleux, P., 1982. Comparative osteology and relationships of the Umbridae (Pisces: Salmoniformes). Zool. J. Linn. Soc-Lond. 76, 321–352.

Wilson, M.V.H., Williams, R.R.G., 1992. Phylogenetic, biogeographic, and ecological significance of early fossil records of North American freshwater teleostean fishes, in: Mayden, R.L. (Ed.), Systematics, Historical Ecology, & North American Freshwater Fishes. Stanford University Press, Stanford, California, p. 969.

Wilson, M.V.H., Williams, R.R.G., 2010. Salmoniform fishes: key fossils, supertree, and possible morphological synapomorphies., in: Nelson, J.S., Schultze, H.P., Wilson, M.V.H. (Eds.), Origin and Phylogenetic Interrelationships of Teleosts. Verlag Dr. Friedrich Pfeil, München, pp. 379–409.

Yasuike, M., Jantzen, S., Cooper, G.A., Leder, E., Davidson, W.S., Koop, B.F., 2010. Grayling (Thymallinae) phylogeny within salmonids: complete mitochondrial DNA sequences of *Thymallus arcticus* and *Thymallus thymallus*. J. Fish. Biol. 76, 395–400.

CHAPTER 3: ARE FLATFISHES (PLEURONECTIFORMES) MONOPHYLETIC?[1]

Abstract

All extant species of flatfish (order Pleuronectiformes) are thought to descend from a common ancestor, and therefore to represent a monophyletic group. This hypothesis is based largely on the dramatic bilateral asymmetry and associated ocular migration characteristics of all flatfish. Yet, molecular-based phylogenetic studies have been inconclusive on this premise. Support for flatfish monophyly has varied with differences in taxonomic and gene region sampling schemes. Notably, the genus *Psettodes* has been found to be more related to non-flatfishes than to other flatfishes in many recent studies. The polyphyletic nature of the Pleuronectiformes is often inferred to be the result of weak historical signal and/or artifact of phylogenetic inference due to a bias in the data. In this study, we address the question of pleuronectiform monophyly with a broad set of markers (from six phylogenetically informative nuclear loci) and inference methods designed to limit the influence of phylogenetic artifacts. Concomitant with a character-rich analytical strategy, an extensive taxonomic sampling of flatfish and potential close relatives is used to increase power and resolution. Results of our analyses are most consistent with a non-monophyletic Pleuronectiformes with *Psettodes* always being excluded. A fossil-calibrated Bayesian relaxed clock analysis estimates the age of Pleuronectoidei to be 73 Ma, and the time to most recent common ancestor of Pleuronectoidei, *Psettodes*, and other relative taxa to be 77 Ma. The ages are much older than the records of any

[1] Matthew A. Campbell, Wei-Jen Chen, J. Andrés López. 2013. Are flatfishes (Pleuronectiformes) monophyletic? Molecular Phylogenetics and Evolution 69:664-673.

fossil pleuronectiform currently recognized. We discuss our findings in the context of the available morphological evidence and discuss the compatibility of our molecular hypothesis with morphological data regarding extinct and extant flatfish forms.

Introduction

Current state of flatfish systematics

Flatfish (Percomorpha: Pleuronectiformes) have received attention in evolutionary biology from Darwin's time (Darwin, 1872) because of their pronounced cranial asymmetry, which requires the ontogenetic migration of an eye from one side of the head to the other (Frazzetta, 2012). The lack of extant species with incipient or partial cranial asymmetry opens questions about evolutionary tempo and mode of the morphological change (sudden vs. gradual evolutionary change) and room for speculation on the evolutionary scenarios that would promote the evolution of asymmetry (Janvier, 2008). For example, Lamarck proposed a scenario of adaptive evolution in which flatfish ancestors lived in exceedingly shallow water and lied flatly on the sea bed (Lamarck, 1809). Flatfish and the absence of intermediate forms were discussed as early challenges to theories of evolutionary change through the accumulation of a series of small steps (Darwin, 1872; Mivart, 1871). The recent discovery of fossils showing an intermediate degree of asymmetry casts those early debates in a new light by showing how the current marked asymmetry could have arisen (Friedman, 2008, 2012). However, pleuronectiform monophyly remains a topic of ongoing debate (*e.g.*, Amaoka, 1969; Chabanaud, 1949; Chapleau, 1993; Dettai and Lecointre, 2005).

Support for pleuronectiform monophyly is based largely on the dorsoventrally compressed morphology that the group's common name highlights. Three synapomorphies have been identified in support of flatfish monophyly: (1) cranial asymmetry as a result of the migration of the eyes, (2) the dorsal fin positioned dorsal to the skull, and (3) the presence of the *recessus orbitalis* (Chapleau,1993). The *recessus orbitalis* is a muscular sac in the orbit that can

be filled with fluid enabling the eyes to protrude above the head while a flatfish is lying on the substrate (Cole and Johnstone, 1902; Holt, 1894). Flatfish begin life as bilaterally symmetric larvae, but develop asymmetry through development as one eye migrates dorsally across the head and cranium to the opposite side (Brewster, 1987). Pleuronectiformes is a species-rich group with approximately 700 recognized, extant species, 134 genera, and 14 families. It is considered to be derived from a perciform (perch-like) lineage (Chapleau, 1993; Chen et al., 2003; Munroe, 2005; Nelson, 2006). The core of flatfish species diversity occurs in the tropics but about one fourth of the species are found in temperate waters (Hensley, 1997; Munroe, 2005).

According to the otolith fossil record, early pleuronectiforms could have been present in the Late Paleocene–Early Eocene, 57–53 Ma (Munroe, 2005; Schwarzhans, 1999). The oldest crown flatfish fossil skeleton known is a representative of unknown affinity to extant forms of bothids from the Lutetian, Eocene (around 45 million years ago; Chanet, 1997, 1999; Norman, 1934). Shortly after this period, several different pleuronectiform lineages suddenly appear in the fossil record along with other diverse acanthomorph fishes (Chanet, 1997; Munroe, 2005; Patterson, 1993; Schwarzhans, 1999). Among fossil flatfishes, Soleidae is the best represented family (Chanet, 1999). Extant intermediary forms between symmetrical and asymmetrical fish do not exist, though they are present in the fossil record at approximately 40–50 million years ago (Friedman, 2008, 2012).

Phylogenetic studies appear to be converging on a consensus but not yet fully defined placement of flatfish among one of the major acanthomorph clades: clade L or the Carangimorpha sensu Li et al. (2009). Evidence for clade L was first reported by Chen et al. (2003) from multiple gene sequence data. Currently, this clade comprises disparate perciform

taxa encompassing carangids (jacks), echeneids (remoras), coryphaenids (dolphinfishes), rachycentrids (cobia), sphyraenids (barracudas), menids (moonfish), polynemids (threadfins), xiphiids (swordfish), istiophorids (billfishes), toxotids (archerfishes), centropomids (snooks), latids (Nile perches and allies) (Betancur-R. et al., 2013a, 2013b; Chen et al.,2003, 2007; Li et al., 2009; Little et al., 2010; Near et al., 2012; Smith and Craig, 2007; Smith and Wheeler, 2006; Wainwright et al., 2012). *Lactarius* (false trevally) has been recognized as part of the Carangimorpha in this study. Yet, questions regarding when flatfishes evolved, and how these diverse lineages are related to each other and to other percomorphs in the clade L remain unresolved (Azevedo et al., 2008; Berendzen and Dimmick, 2002; Chapleau, 1993; Chen et al., 2003; Dettai and Lecointre, 2005; Li et al., 2009; Little et al., 2010; Roje, 2010; Shi et al., 2011; Smith and Wheeler, 2006). Moreover, molecular studies have not consistently shown flatfishes to be a monophyletic group with Psettodidae and a few taxa exhibiting base composition bias often excluded (Betancur-R. et al., 2013a; Chen et al., 2003; Dettai and Lecointre, 2005; Li et al., 2009; Near et al., 2012, 2013; Smith and Wheeler, 2006).

Psettodes (spiny turbot) and pleuronectiform polyphyly

Psettodidae contains a single genus (*Psettodes*) with three recognized species (Nelson, 2006). The condition of the three pleuronectiform synapomorphies differs between *Psettodes* and other pleuronectiforms. Generally in pleuronectiforms the eyes are on the same side of the head, but in the case of *Psettodes* one eye is at the dorsal midline (Friedman, 2008). This condition affects the insertion of the dorsal fin in *Psettodes*, which unlike that in other flatfish, is posterior to the eye (Nelson, 2006). The recessus orbitalis is assumed by Chapleau (1993) to be present among all flatfishes including *Psettodes*, but it may in fact not be found in *Psettodes*

(Chabanaud, 1937). Chabanaud (1937) notes that the eyes of *Psettodes* cannot be extended and do not have any skin folds around the eyes unlike pleuronectoids, which can extend the eyes and have skin folds around the eyes. In addition, *Psettodes* has distinct characteristics that are not typical of other flatfish. Populations of species of *Psettodes* may include both left- and right-sided fish, a characteristic termed antisymmetry. In contrast, populations of other pleuronectiform species have a tendency to be uniformly left or right sided (Palmer, 1996). *Psettodes* retains many characters considered to be ancestral in Pleuronectiformes. Chapleau (1993) lists the following: palatine teeth (character 4), toothed plates on basihyal (character 5), a basisphenoid (character 6), spines in median fins (character 7), absent or not well developed sciatic portion of urohyal, (character 8), presence of uroneural 1 (character 9), elongated shape of second infrapharyngobranchial (character 10), a large maxilla (character 11), and a parhypural the articulating with vertebral column (character 35). Other characters that may be considered primitive in *Psettodes* are the presence of a macula neglecta in the inner ear (Platt, 1983) and vertical barring (Hewer, 1931). *Psettodes* bodies are almost rounded and do not have the associated bilateral asymmetry in musculature typical of other flatfishes (Munroe, 2005) and often swim in an upright orientation (Hensley, 1999). The distinct morphology of *Psettodes* earned it an early characterization as ''simply an asymmetric ercoid'' (Regan, 1910). The theory that *Psettodes* arose from a different lineage is not new, and several authors outline the similarities of *Psettodes* to percoids (Amaoka, 1969; Hubbs, 1945; Kyle, 1921; Norman, 1934; Regan, 1910, 1929). The scarcity of shared derived characters among percoid families severely limits the phylogenetic utility of these observations (Chapleau, 1993; Gosline, 1971; Johnson, 1984). *Psettodes* is now considered to be most closely related to other flatfishes and to be the most basal lineage of the Pleuronectiformes (Chapleau, 1993; Friedman, 2012; Munroe, 2005).

In the current system of fish classification, the suborder Psettodoidei (including a sole family Psettodidae) is the sister lineage of all other living flatfish species, which are grouped in the suborder Pleuronectoidei (Nelson, 2006).

From a molecular-based perspective, the inferred phylogenetic placement of *Psettodes* and other flatfish taxa has varied between studies depending on genes surveyed or inference method employed (Dettai and Lecointre, 2005). The salient pattern is that psettodids are not grouped with other pleuronectiform taxa (Dettai and Lecointre, 2005; Li et al., 2009; Smith and Wheeler, 2006). It remains to be determined if methodological artifacts (e.g., base composition or long-branch attraction) are responsible for the non-monophyly of flatfish. For this study we attempt to resolve that question by increasing the number of independent data sources (more genes) and by recognizing and addressing sources of phylogenetic artifacts.

Data sources

To improve phylogenetic resolution we examined six independent sources of characters in the form of single copy protein-coding nuclear genes. Increasing the number of independent data points and sites is a well established strategy for improving the accuracy of phylogenetic inference (e.g., Cao et al., 1994; Chen and Mayden, 2010; Mitchell et al., 2000; Russo et al., 1996; Wolf et al., 2004; Wortley et al., 2005). Studies of acanthomorph phylogeny including a sampling of Pleuronectiformes from divergent lineages and based on evidence from more than one locus have used three data sources (Chen et al., 2003; Smith and Wheeler, 2006) or four data sources (Dettai and Lecointre, 2005; Li et al., 2009). It is important to consider that increasing the number of traits will not circumvent problems due to substitution saturation, difficulties in alignment, and/or lack of information due to strong sequence conservation (Smith and Wheeler,

2006) as noted by Chen et al. (2003, 2008) and Li et al. (2009). The genes sequenced here were selected in part because they can be aligned with little or no ambiguity and have been reported to be phylogenetically informative (Chen, 2001; Chen et al., 2003, 2008; Dettai and Lecointre, 2005; López et al., 2004).

Phylogenetic artifacts

The most commonly used implementations of nucleotide substitution models for phylogenetics assume that nucleotide frequencies remain relatively stable across all the lineages being examined. However, there is ample evidence of base composition shifts at different levels of phylogenetic divergence (Akashi et al., 1998; Eyre-Walker, 1999; Galtier and Gouy, 1995; Mooers and Holmes, 2000). Relying on an incorrect substitution model can mislead phylogenetic inference by affecting branch length estimation (Posada, 2001) or estimating an incorrect topology (Bruno and Halpern, 1999; Penny et al.,1994) and support for resulting topologies can also be biased (Buckley and Cunningham, 2002). Convergent base composition can result in organisms being improperly associated in phylogenies as a result of similarity in overall frequencies of nucleotides (Delsuc et al., 2005; Foster and Hickey, 1999; Phillips et al., 2004; Steel et al., 1993). In cases where a molecular hypothesis opposes a well established morphological hypothesis,it is often thought that base composition bias is at fault (e.g., Betancur-R. et al., 2013b; Li and Ortí, 2007; Sheffield et al., 2009). However, identifying when the degree of deviation from base composition stationarity will mislead phylogenetic inference is problematic (Jermiin et al., 2004). In addition to base composition non-stationarity, long branch attraction can also contribute to artificial support for groupings not corresponding to true clades (Bergsten, 2005; Felsenstein, 1978).

We approach the question of pleuronectiform monophyly with the intent of identifying and eliminating possible biases in the data. We exhaustively evaluate our sequence data by taxon, gene, and codon position for evidence of compositional bias or saturation and remove or recode affected characters. A broad sampling of pleuronectiforms and possible relatives is used to reduce occurrences of long-branch attraction in the dataset and increase accuracy (Hillis, 1998; Hillis et al., 2003). Because simply eliminating data partitions or taxa, and using alternative sequence codings to reduce compositional bias comes at the cost of potential phylogenetic signal, we also employ several phylogenetic methods that have been designed to account for non-stationarity of base composition evolution (e.g., p4; Foster, 2004). We review existing morphological evidence to assess the compatibility between molecular and morphological sources concerning the question of pleuronectiform monophyly.

Materials and methods

Taxon and character sampling

We assembled a taxonomic sample representing all major divisions within the Pleuronectiformes including *Psettodes* (25 taxa). In addition, we sampled heavily (15 taxa) within acanthomorph clade L, the Carangimorpha (Chen et al., 2003; Li et al., 2009) to capture potential sister taxa of pleuronectiforms (see introduction). Finally, a broad sampling of 48 additional percomorph taxa representing main lineages recently identified in molecular analyses (Chen et al., 2003, 2007; Li et al., 2009; Miya et al., 2003; Smith and Craig, 2007; Wainwright et al., 2012) were included to evaluate the support for acanthomorph clade L. Two beryciforms were used as outgroups to root the percomorph tree. Tissue samples were obtained from collections performed by W.-J. Chen or the University of Kansas tissue collections

(Supplemental Table S3.1A, Supplemental Table S3.1B). In addition to newly reported sequences, publicly available sequences from GenBank were included this study (Supplemental Table S3.1A, Supplemental Table S3.1B).

DNA data

Total genomic DNA was isolated from samples using Qiagen DNEasy spin-column or QIAmp kits following manufacturer's directions. Fragments of six nuclear protein-coding genes were amplified for this study. The nuclear protein-coding genes used in the study are Recombination Activating Gene 1 (RAG1), Rhodopsin (RH), Early Growth Response Protein genes 1, 2B, and 3 (E1, E2B, E3), and Mixed-lineage Leukemia (MLL). Primer sequences and sources are given in Supplemental Table S3.2. The temperature cycling profile used for amplification of RAG1 had an initial denaturation step of 95° C for 4 min, followed by 35 cycles of 95° C for 40 s, 53° C for 40 s, and 72° C for 90 s, and a final extension of 72° C for 7 min. For the other five genes a similar profile was used, but the annealing temperature was raised to 55° C and the extension time was reduced to 60 s. Either Takara ExTaq or Promega GoTaq Flexi were used. For amplifications using Promega GoTaq Flexi, PCR reagent concentrations were 1X Promega GoTaq Flexi reaction buffer, 0.25 mMdNTP's, 2.0 mMMgCl2, 0.4 μMforward primer, 0.4 μMreverse primer, 0.025 U/μL GoTaq Flexi Taq polymerase, and 1 μL template DNA (variable concentrations). Reagent concentrations for reactions using Takara ExTaq were 1X Takara ExTaq reaction buffer, 0.8 mM dNTP's, 2.0 mM MgCl2, 0.2 μM forward primer, 0.2 μM reverse primer, 0.5 U/μL Takara ExTaq polymerase. Diluted DNA extractions of varying concentrations were added at a ratio of 2.5 μL for a 25 μL reaction. Unpurified PCR products were sent to multiple commercial institutions for purification and Sanger sequencing. Raw

sequence output was examined, edited and assembled using the features implemented in CodonCode Aligner Version 3.7.1.1 (by CodonCode Corp., Dedham, MA, USA).

Assembled DNA sequences were managed using Se-Al v2.0a11 (available at http://tree.bio.ed.ac.uk/software/seal/) and Mesquite 2.75 (Maddison and Maddison, 2011). Compiled sequences were initially aligned with MUSCLE (Edgar, 2004a, 2004b) using the on-line server at http://www.ebi.ac.uk/Tools/muscle/index.html. Alignments were then adjusted manually to ensure that the placement of inferred insertions/deletions (indels) followed the expected codon structure. Regions containing large indels (e.g. tandem repeats in EGR genes) showing high dissimilarity in sequence length, which may produce invalid assertions of homology were discarded from the phylogenetic analyses. We trimmed the 5'- and 3'-ends of some sequences to reduce the number of sites with missing data.

Stationary phylogenetic analyses

For the initial phylogenetic analyses, we had two expectations for variability in the data since the most constrained codon position is the second and the least is the third position (Alff-Steinberger, 1969; Haig and Hurst, 1991; Kimura, 1980; Woese, 1965). Consequently, we expected stronger and more numerous deviations from base composition homogeneity at the third codon position than at other positions. Secondly, at the time scales we are investigating, third codon positions could be mutationally saturated and recoding to purines and pyrimidines (RY) would be useful for reducing both saturation and base composition bias (e.g., Chen et al., 2008; Delsuc et al., 2003; Phillips and Penny, 2003; Phillips et al., 2004).

To determine if certain taxon/marker combinations showed significant deviation in base composition, we created alignments of variable sites for each codon position and tested each

alignment using the Chi-squared test for base composition homogeneity implemented in TreePuzzle version 5.2 (Schmidt et al., 2002). Systematic biases across markers were evaluated based on repeated failures to pass the test of homogeneity and helped establish whether genome wide biases are present in the taxa in this study. Based on results of tests for stationarity, we generated the following three alternative codings of the data set for phylogenetic inference: (1) all codon positions retained for all genes ($1_N2_N3_N$); (2) all codon positions retained, third codon positions recoded as purines and pyrimidines ($1_N2_N3_{RY}$); and (3) third codon positions discarded (1_N2_N). We also generated alignments following these three data schemes with no missing data to assess the influence of missing data on inferred relationships.

Phylogenetic analysis of the $1_N2_N3_N$, $1_N2_N3_{RY}$, and 1_N2_N datasets for all taxa and those with no missing data was conducted in RAxML 7.2.8 under a partitioned maximum likelihood (ML) approach using the general time reversible model of nucleotide evolution (GTR) (Stamatakis, 2006) with a four category gamma distribution (Γ), invariant sites (I) and automatic stopping of bootstrap replicates. Data were partitioned by gene and codon position. For the alignments containing all taxa, we evaluated the stability of the resulting topology using RogueNaRok (Aberer et al., 2013). Rogue taxa, those that fail to find a consistent placement among pseudoreplicate analyses (Aberer et al., 2013) were removed and the edited alignment re-analyzed.

Alternative phylogenetic analyses

We also conducted analyses implementing models designed to alleviate issues of compositional heterogeneity. We used the three data coding schemes ($1_N2_N3_N$, $1_N2_N3_{RY}$, and 1_N2_N) partitioned by gene in these analyses and used the programs Phylobayes 3.3.b (Lartillot et

al., 2009) and p4 (Foster, 2004). Phylobayes implements a CAT-GTR model (Lartillot and Philippe, 2004) that allows for more variation in nucleotide evolution than the more widely used substitution models. In Phylobayes we ran an analysis for each of the three data schemes with two chains for at least 500 cycles. After 500 cycles, a sampling every 100 cycles was done to check convergence of the two chains. The program was allowed to run until all discrepancies between the chains were less than 0.3 and all effective sample sizes (ESS) were greater than 50 as recommended by the software developers.

We conducted two different analyses in p4 differing on the treatment of rate matrices and base composition vectors. In both cases the estimate of the α shape parameter for the Γ distribution was linked across partitions with unlinked relative rates for each partition. We used the GTR+I+Γ model of nucleotide evolution in p4 for each data partition. When more than one base composition vector or rate matrix was specified, the additional vector or matrix was constrained to represent at least two taxa. The placement of additional base composition vectors and rate matrices was at first placed randomly, then allowed to vary within the MCMC tree search. The first strategy was to retain a single rate matrix and proportion of invariant sites per partition to reduce parameterization. Each partition was then permitted to have multiple base composition vectors. In our second strategy, we allowed multiple rate matrices and base composition vectors in each partition. We began with a basic Markov chain Monte Carlo (MCMC) Bayesian tree search with p4 using four chains, sampled every 1000 steps, and a total run length of 1,000,000 steps. We subsequently modified MCMC parameters to reach adequate sampling and mixing. In all p4 analyses, we discarded 10% of samples as burnin.

Divergence time estimation

We estimated divergence times using the simultaneous Bayesian phylogenetic inference and divergence time approach (Drummond et al., 2006) with a Bayesian relaxed clock model with uncorrelated lognormal rate heterogeneity as implemented in BEAST version 1.7.2 (Drummond et al., 2012). Given the highly congruent phylogenetic trees produced by the analysis described above, we only employed the $1_N2_N3_{RY}$ data coding scheme in the divergence time analysis. We generated a starting tree for this analysis by partitioning the data by gene and constraining ingroups, outgroups, and the time to most recent common ancestor (TMRCA) of the ingroup. We calibrated the root of the tree at 150 million years ago (Ma) using the first appearance of euteleost and ostariophysan fish in the fossil record at a minimum of 149.85 Ma (Benton et al., 2009). The root age was chosen so that subsequent constraints forced on the starting tree would be compatible. An uncorrelated relaxed clock was used to generate the input tree with a Markov chain Monte Carlo (MCMC) chain of 100 million generations sampled every 5000 generations. All partitions were modeled under a GTR+I+Γ model of nucleotide evolution. We applied a 10% burnin and examined the MCMC run output with Tracer v 1.5 to determine whether the analyses resulted in sufficiently sampled parameters (Drummond et al., 2012). The resulting topology was incorporated as the starting tree into the following divergence time analysis.

The alignment was partitioned by gene and each partition was modeled under a GTR+I+Γ model of evolution. We included settings in BEAST to use ambiguities across all partitions and to unlink the uncorrelated relaxed clock for each data partition. Based on the results from the ML tree search in this study, we assigned lognormal fossil constraint distributions at well-supported nodes (Supplemental Table S3.4). We did not use any fossil

pleuronectiform fossils as calibration points to minimize the effect of prior assumptions on pleuronectiform relationships and age of lineages.

Two independent runs of 100 million generations sampled every 10,000 generations were generated. After verifying adequate sampling and convergence with Tracer v 1.5, we applied a 10% burnin and combined the tree files with LogCombiner. The final maximum clade credibility tree with mean heights was generated with TreeAnnotator.

Results

Taxon and character sampling

Sequence data from a total of 90 taxa are examined in this analysis. This taxonomic sample includes 25 pleuronectiforms (Table A). No taxon has more than two missing genes in our data matrix. Sixty-seven of the 90 taxa did not have any missing sequence data.

Alignment

After end-trimming and concatenation, our final alignment spans 5664 nucleotide sites. The aligned sequence matrix of combined genes (90 taxa) includes about 7.6% missing nucleotides and gapped sites; a text file with the concatenated alignment is available from the Dryad repository (doi: 10.5061/dryad.t749b). The $1_N2_N3_N$ alignment includes 3034 variable sites, of which 2525 are parsimony informative. When recoded as $1_N2_N3_{RY}$ the alignment contains 2396 variable sites and 1821 parsimony informative sites. Excluding first codon positions (1_N2_N) produces an alignment of 3776 characters. Of these, 1239 are variable and 838 are parsimony informative.

Base composition changes

Forty-seven taxa contain compositional biases in one or more genes at variable sites (Supplemental Table S3.3A, Supplemental Table S3.3B, Supplemental Table S3.3C and Fig. 3.1). *Psettodes erumei* exhibits compositional bias towards higher GC content only in the MLL gene, but not a broader genome wide base composition bias. In contrast, systematic base composition bias in other pleuronectiform taxa is evident (e.g., Bothidae). In other percomorph taxa,GC and AT bias are only evident in lophiiforms. We detect no evidence of unusual base composition biases in the non-pleuronectiform cargangimorphs included in our sample.

3.4. Stationary phylogenetic analyses

Results of ML phylogenetic analyses using the different combination of data and taxa described in the methods, consistently find a non-monophyletic Pleuronectiformes with *Psettodes* always excluded (Fig. 3.1 and Table 3.1). The monophyly of the suborder Pleuronectoidei (Pleuronectiformes minus *Psettodes*) is supported. Important relationships and bootstrap support are summarized in Table 1. The sister group relationship of Pleuronectoidei and Centropomidae (*sensu* Greenwood (1976), Lates + Centropomus) is consistently inferred across all ML analyses. The placement of *Psettodes* varied with taxon sampling and data scheme. Evaluation of ML results from the alignments containing all taxa with RogueNaRok identified rogue taxa in $1_N2_N3_N$ and 1_N2_N data schemes. Importantly, this analysis does not identify *Psettodes* as a possible rogue taxon. ML searches with rogue taxa removed from $1_N2_N3_N$ and 1_N2_N searches again resolve a non-monophyletic Pleuronectiformes (Table 1). Finally, all of the analyses strongly support the monophyletic ‘‘Carangimorpha’’ (Clade L, ML bootstrap value = 100%; Posterior probability = 1). Carangimorpha in this study includes recognized taxa from

previous molecular studies plus a perciform family, Lactariidae. Lactariidae contains only one species, *Lactarius lactarius*, widely distributed in Indo-West Pacific (Nelson, 2006).

Alternative phylogenetic analyses

Use of the CAT-GTR model in Phylobayes does not result in convergence with the $1_N2_N3_N$ data scheme. In the case of $1_N2_N3_N$ data scheme, long run time permits high ESS for each parameter but variation between chains remains greater than 0.3. With the other two data schemes, convergence was reached and the topologies generated by the $1_N2_N3_{RY}$ and 1_N2_N coding schemes are summarized in Table 3.1.

Analysis of the data with p4 varied in base composition vectors and rate matrices assigned to each data scheme. In analyses of the $1_N2_N3_N$ dataset we assigned six base composition vectors with a run length of 5 million generations; and four base composition vectors and four rate matrices with a run length of 3 million generations in a second analysis. In analyses of the $1_N2_N3_{RY}$ matrix, five base vectors were modeled on the tree in addition to a single rate matrix; and three base vectors and three rate matrices in a second analysis. Both of these $1_N2_N3_{RY}$ analyses ran for 3 million generations. In analyses of the 1_N2_N matrix, we allowed four base vectors and one rate matrix; and three base vectors and three rate matrices in a second analysis. Both of these runs had lengths of 3 million generations. In all six p4 analyes, Pleuronectiformes is polyphyletic. Pleuronectoidei remains monophyletic whereas *Psettodes* is more closely related to non-pleuronectiform taxa (Table 3.1). The placement of *Psettodes* is inconsistent between analyses.

3.6. Divergence time estimation

Simultaneous Bayesian tree inference and divergence time estimation results in a paraphyletic Pleuronectiformes, but monophyletic Pleuronectoidei and Carangimorpha, which includes *Psettodes* (Fig. 3.2 and Table 3.1). We find the estimated divergence time for the split between Pleuronectoidei and Centropomidae to have a mean age of 75.3 Ma (95% highest posterior density (HPD) 67.3–84.5), and the time to MRCA of the Pleuronectoidei to be 73.4 Ma (95% HPD 65.1–82.1). The origin of the carangimorphs dates back to 78.4 Ma (95% HPD 65.2–130.0). The time to the divergence of *Psettodes* from other fishes in our sample is estimated to have a mean of 77.4 Ma (95% HPD 69.7–86.5 Ma).

Discussion

Non-monophyletic Pleuronectiformes and the sister of the Pleuronectoidei

Combined, the results from all our analyses indicate that the six-gene (~5.5 kbp) dataset is incongruent with a monophyletic Pleuronectiformes. The genus *Psettodes* is consistently excluded from the Pleuronectiformes across analyses. We find the Pleuronectoidei to be monophyletic, in agreement with previous molecular and morphological studies (Azevedo et al., 2008; Berendzen and Dimmick, 2002; Chapleau, 1993). We identify the sister-taxa of the Pleuronectoidei to be the Centropomidae (including Latidae; see below). We addressed several potential biases that may have misled phylogenetic inference. Our taxonomic sampling is broad and specifically targets clade L percomorphs as potential sister lineages to pleuronectiform clades. We include multiple independent loci with a substantial degree of variability that could be aligned with high confidence. Further, we evaluated the loci for base composition homogeneity and implemented alternative treatments of third codon positions (RY recoding and

deletion). We evaluated the stability of phylogenetic inference by using only taxa with no missing data and by excluding potentially problematic taxa as identified by the approach implemented in RogueNaRok. We also used several alternative molecular evolution models. No treatment of the data yielded support of a monophyletic Pleuronectiformes. Pleuronectioid sister-taxa, Centropomidae *sensu* Greenwood (1976) includes two currently recognized perciform families Latidae and Centropomidae (Nelson, 2006). The evolutionary affinity for these two families was confirmed by recent molecular studies (Chen et al., 2007; Li et al., 2011; Near et al., 2012) and this study. Two morphological features used to unite these two groups in single assemblage are: (1) expanded neural arch and spine on the 2nd vertebrae often embracing the spine of the first vertebra; (2) and, pored lateral-line scales extending to the posterior edge of the caudal fin. Although determination of these features (e.g., morphology of the second neural spine) remain highly subjective and the posterior extension of the lateral line may be present in other percomorphs (e.g., Sciaenidae; Mooi and Gill, 1995; Otero, 2004), extant and extinct flatfishes share a posterior extensive lateral line with centropomids (Fukuda et al., 2010; Yamanaka et al., 2010).

Psettodes is placed within the Carangimorpha, however there is no consistent support for any particular sister lineage for this genus. It is already recognized that *Psettodes* is a divergent flatfish lineage that has been interpreted as basally divergent among the flatfishes. It is recognized as a separate suborder in morphological studies (Chapleau, 1993; Friedman, 2012). Regardless its morphological distinctiveness, *Psettodes* is thought to possess the synapomorphies proposed for the Pleuronectiformes by Chapleau (1993). However, the presence of the recessus orbitalis has not been systematically evaluated among flatfish groups (Hensley, 1997), and may not be present in *Psettodes*. Chabanaud (1937) while noting that *Psettodes* cannot protrude its

eyes and lacks skin folds around the eyes which would suggest it can, did not examine *Psettodes* for the presence of the recessus orbitalis. Determining the condition of this character in *Psettodes* will help establish the extent to which the morphological and molecular lines of evidence conflict. The traits that *Psettodes* shares with percoids (Amaoka, 1969; Hubbs, 1945; Norman, 1934; Regan, 1910, 1929) do not provide synapomorphies to identify a potential sister group for *Psettodes*. In light of the molecular evidence, does a review of the existing literature reveal potentially informative traits linking *Psettodes* to non-flatfish groups? Work predating the development of cladistics placed *Psettodes* among serranids (Norman, 1934). A cladistic analysis identified four synapomorphies supporting serranid monophyly (Johnson, 1983). *Psettodes* shares two of these four characters with serranids (no third preural cartilage and no procurrent spur). Both traits are reductive and may represent independent losses (Chapleau, 1993; Johnson, 1983). Finally, aspects of the head musculature of *Psettodes* have been used to suggest affinity to carangids (Kyle, 1921), and we find *Psettodes* to be a close relative of carangids with this molecular dataset. A broad analysis of morphological variation among pleuronectiforms and acanthomorphs may add clarity to nature of the apparent conflict between morphology and molecular-based hypotheses of pleuronectiform relationships.

How does accepting a polyphyletic pleuronectiforms affect the interpretation of extinct intermediate flatfish lineages?

†*Amphistium* is found in deposits from the Ypresian and Lutetian (40.4–55.8 Ma) while †*Heteronectes* is documented from the Ypresian (48.6–55.8 Ma; Walker and Geismann, 2009; Friedman, 2012). Both are much younger than the estimated ages of the origin of Pleuronectoidei and Carangimorpha (Fig. 2). We estimate a mean age of 73.4 Ma with a 95% HPD range of

65.1–82.1 Ma for the time to MRCA of extant Pleuronectoidei. The age of crown pleuronectiforms predating known flatfish intermediates is consistent with the fact that †*Amphistium* and †*Heteronectes* occurred in strata that also contain fossils showing complete cranial asymmetry (Chanet, 1997, 1999; Friedman, 2008). Although preservation of these fossils makes it difficult to fully evaluate all relevant characters, †*Amphistium* and †*Heteronectes* are characterized by cranial asymmetry. Cranial asymmetry in †*Amphistium* and †*Heteronectes* is not as complete compared to extant flatfishes (Friedman, 2008). With regards to pleuronectiform synapomorphies, †*Amphistium* has derived pleuronectiform features unrelated to asymmetry that cannot be evaluated in †*Heteronectes*: (1) a dorsal fin that is anteriorly extensive, (2) anteriorly curved neural spines of the abdominal region, and (3) a procumbent first pterygiophore of the dorsal fin (Friedman, 2008).

†*Amphistium* and †*Heteronectes* both have traits in common with *Psettodes* that are considered primitive for pleuronectiforms (Friedman, 2012). †*Amphistium* and *Psettodes* differ with regards to a ventrally directed sciatic process (character 8 of Chapleau (1993), Friedman, 2008, 2012); however share character states considered primitive for flatfish otherwise. †*Heteronectes* can be evaluated for five of seven osteological characters that are considered informative for flatfish relationships, four of which are shared with *Psettodes* in a primitive state (Friedman, 2012). The fifth character, cranial asymmetry, is incomplete and considered by Friedman (2012) to be sufficient to place †*Heteronectes* as a flatfish. Otherwise †*Amphistium* and †*Heteronectes* show general percomorph character states including presence of dorsal and anal fine spines, and in the case of †Heteronectes a procurrent spur, found only in *Psettodes* amongst extant flatfishes. These characteristics have been offered as evidence to place the two fossil taxa as stem lineages of a monophyletic Pleuronectiformes with †*Amphistium* higher up

along the stem (Friedman, 2008, 2012). If the pleuronectiforms as currently defined include representatives of two divergent lineages, then it will be important to re-evaluate the affinities of †*Amphistium* or †*Heteronectes* to adequately characterize the evolution of bilateral asymmetry in fishes. In light of the phylogenetic hypothesis supported by molecular evidence in this study, it will be especially valuable to review morphological variation in *Psettodes* and the two fossil genera to test the stem placement of the fossil taxa. It is possible that either †*Amphistium* or †*Heteronectes* are not stem members of a monophyletic Pleuronectiformes [sensu Chapleau (1993)] or the Pleuronectoidei, and might be a stem lineage of Psettodoidei or related to other lineages of percomorphs (Friedman, 2012).

Our results support parallel evolution of the flatfish body form with pronounced cranial asymmetry in two fish lineages with extant representatives. A growing number of molecular-based phylogenetic studies offer evidence rejecting monophyly of Pleuronectiformes as the result of alternative placements for the genus *Psettodes* (e.g., Betancur-R. et al., 2013a; Near et al., 2012, 2013; but see Betancur-R. et al., 2013b). The evidence is found in different taxonomic and gene fragment samples. The potential biases in base composition across taxa may mislead our conclusion about monophyly or non-monophyly of the Pleuronectiformes, and possibly affect our inference of intra-pleuronectiform phylogenywhen fewer gene markers and/or inappropriate phylogenetic reconstruction methods used. For instance, a monophyletic Cynoglossidae (GC biased) was only found with $1_N2_N3_N$ coding by p4 in our study. However given consistent results across a broad range of treatments of the sequence data, we find it unlikely that our non-monophyletic Pleuronectiformes is the product of artifacts of phylogenetic reconstruction. The results of our study support at least two independent origins of a flatfish body form with pronounced cranial asymmetry. If further phylogenetic analyses corroborate this finding, the

evolution of cranial asymmetry should prove a rich research topic for understanding parallel evolution of complex traits. If parallel evolution of body asymmetry is confirmed by further research, it would suggest that major morphological adaptations can take place in the context of relatively modest degrees of divergence at the coding sequence level and point to important roles for regulatory changes in the evolution of complex morphological adaptations.

Acknowledgments

Sebastien Lavoué provided assistance with the Bayesian divergence time analysis. We want to thank Jhen-Nien Chen, Pei-Chun Lo, and Hsin Lee for their efforts in sequencing. Bruno Chanet and Jordan S. Metzgar provided helpful comments on earlier versions of the manuscript. Our gratitude goes to Bruno Chanet, Samuel Iglésia, Guy Duhamel, Richard L. Mayden, Kwang-Tsao Shao, Mao-Ying Lee, Labbish L. Chao, University of Kansas Natural History Museum and the Scripps Institution of Oceanography Marine Vertebrates Collection for sharing tissue samples (via H.-J. Walker). Computational support was provided by UA Life Science Informatics, Grant Number P20RR016466 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). This work was supported by a joint award from the Taiwanese National Science Council/U.S. National Science Foundation (OISE 1210243 to M.A.C.), the Taiwanese National Science Council (NSC 99-2611-M-002-001-MY2 and NSC 101-2611-M-002 -016 -MY3 to W.J.C.), and the U.S. National Science Foundation (DEB 0963767 to J.A.L.).

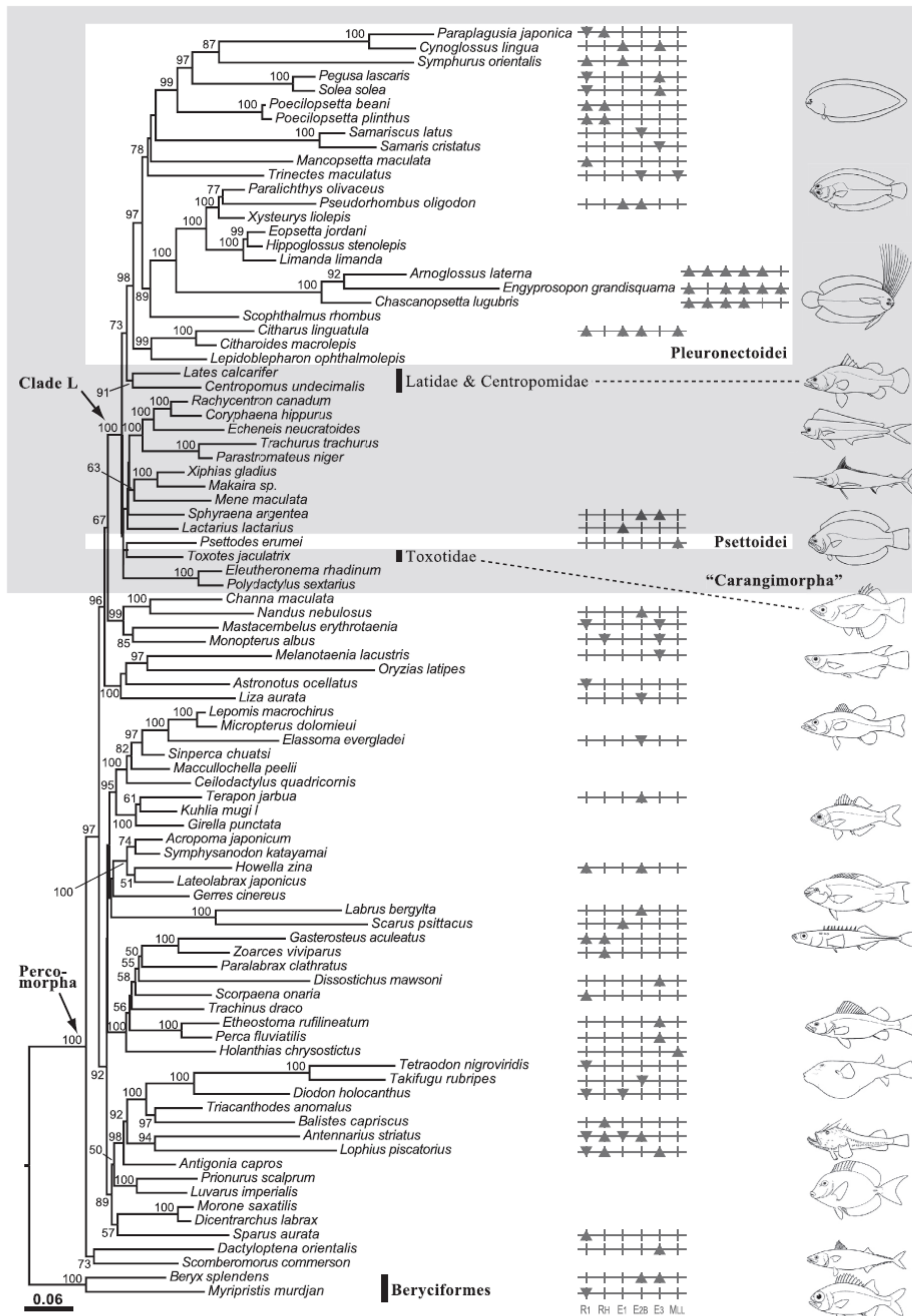


Figure 3.1: A maximum-likelihood (ML) tree generated under a GTR+I+Γ model of sequence evolution in RAxML, depicting phylogenetic positions of the flatfishes (Pleuronectiformes) (taxa within the white rectangle boxes) in relation to other percomorph taxa. All taxa are included in the analysis with data partitioned by gene and codon position and third codons recoded as purines or pyrimidines ($1_N2_N3_{RY}$). Values at nodes represent bootstrap values. Those values below 50% are not shown. Taxa with significant higher GC content and lower GC content with respect to gene partitions, as detected by chi-square tests, are indicated as black up-pointing and down-pointing triangles, respectively, after the taxon names.

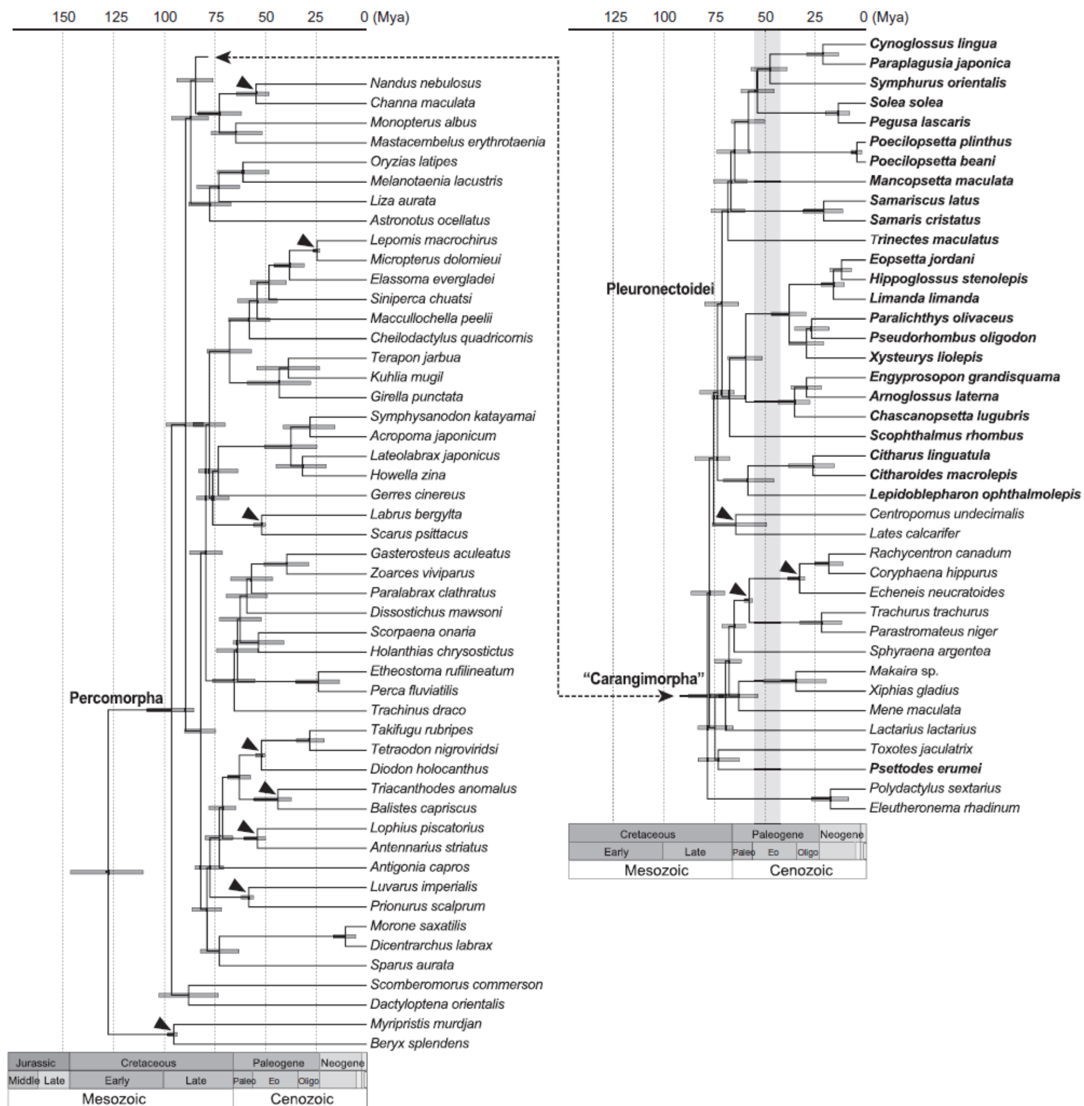


Figure 3.2: Timetree based on a Bayesian relaxed clock calibrated by fossils and distributions described in Supplementary Table S3.4. The timescale is in millions of years ago (Ma). Horizontal bars at nodes represent 95% highest posterior densities and black triangles indicate fossil calibrated nodes. Pleuronectiform taxa are highlighted in bold. Vertical bar in light gray indicates the period when flatfish intermediates were present according to fossil records.

Table 3.1: Summary of key relationships and support values for phylogenetic analyses of different data coding schemes and taxon composition. Centopomidae includes Lates and Centropomus (Greenwood, 1976; Li et al., 2011). For each analysis the basic characteristics and outcomes are reported.

Data Scheme	Numer of Partitions	Pleuronectiformes Monophyletic?	Bootstrap Support or Posterior Probability For Pleuronectoidei	Sister Group of Pleuronectoidei	Bootstrap Support or Posterior Probability for Sister of Pleuronectoidei	Sister of *Psettodes*	Bootstrap Support or Posterior Probability for Sister of *Psettodes*
ML Analysis							
$1_N2_N3_N$	18	No	100	Centropomidae	32	*Toxotes jaculatrix*	27
$1_N2_N3_{RY}$	18	No	98	Centropomidae	73	*Toxotes jaculatrix*	39
1_N2_N	18	No	70	Centropomidae	67	Other Carangimorpha, not Pleuronectifomres + Centropomidae excluding *Polydactylus sextarius* and *Eleutheronema rhadinum*	21
$1_N2_N3_N$ No Missing Data	18	No	100	Centropomidae	34	*Eleutheronema rhadinum*	47
$1_N2_N3_{RY}$ No Missing Data	18	No	99	Centropomidae	81	Pleuronectiformes+Other Carangimorpha excluding *Eleutheronema rhadinum*	34
1_N2_N No Missing Data	18	No	92	Centropomidae	87	Other Carangimorpha excluding *Toxotes jaculatrix* and *Eleutheronema rhadinum*	16
$1_N2_N3_N$ Rogue Taxa Removed	18	No	100	Centropomidae	29	*Polydactylus sextarius* + *Eleutheronema rhadinum*	36
1_N2_N Rogue Taxa Removed	18	No	82	Centropomidae	75	Other Carangimorpha, not Pleuronectifomres + Centropomidae, excluding *Polydactylus sextarius* and *Eleutheronema rhadinum*	42
GTR-CAT Model							
$1_N2_N3_{RY}$	6	No	0.87	Centropomidae	0.5	Part of four branch polytomy at base of Carangimorpha	0.99
1_N2_N	6	No	0.94	Polytomy including Centropomidae	-	Part of five branch polytomy at base of Carangimorpha	-
p4 Multiple Composition Vectors							
$1_N2_N3_N$	6	No	59	Centropomidae	54	*Eleutheronema rhadinum* + *Polydactylus sextarius*	100
$1_N2_N3_{RY}$	6	No	100	Centropomidae	98	Other Carangimorpha excluding *Toxotes jaculatrix, Eleutheronema rhadinum* and *Polydactylus sextarius*	99
1_N2_N	6	No	90	Centropomidae	90	Pleuronectiformes+Other Carangimorpha excluding *Eleutheronema rhadinum* and *Polydactylus sextarius*	66
p4 Multiple Composition Vectors and Rate Matrices							
$1_N2_N3_N$	6	No	61	Centropomidae	56	*Eleutheronema rhadinum* + *Polydactylus sextarius*	100
$1_N2_N3_{RY}$	6	No	100	Centropomidae	97	Other Carangimorpha excluding *Toxotes jaculatrix, Eleutheronema rhadinum* and *Polydactylus sextarius*	99
1_N2_N	6	No	98	Centropomidae	100	Other Carangimorpha excluding *Toxotes jaculatrix, Eleutheronema rhadinum* and *Polydactylus sextarius*	65
BEAST							
$1_N2_N3_{RY}$	6	No	1	Centropomidae	0.98	*Toxotes jaculatrix*	0.52

Table S3.1A: Non-pleuronectiform taxa included in this study, the corresponding accession to the tissue (if any), and corresponding GenBank accession numbers for gene sequences. Newly determined sequences are in bold; * sequences retrieved from complete genomic database, Ensemble (http://www.ensembl.org/).

Taxon Name	Order	Family	R1	RH	E1	E2B	E3	MLL
Melanotaenia lacustris	Atheriniformes	Melanotaeniidae	JN230909	JN231008	JN230961	JN231061	JN231123	
Oryzias latipes	Beloniformes	Adrianichthyidae	Ensemble*	Ensemble*	Ensemble*	Ensemble*	Ensemble*	Ensemble*
Beryx splendens	Beryciformes	Berycidae	EF095636	AY141265	JN230957	JN231057	JN231119	AY362238
Myripristis murdjan	Beryciformes	Holocentridae	KC442204	KC442231	KC442093		KC442166	
Dactyloptena orientalis	Dactylopteriformes	Dactylopteridae	KC442206	KC442232	KC442096	KC442130	KC442169	**KF312007**
Gasterosteus aculeatus	Gasterosteiformes	Gasterosteidae	Ensemble*	Ensemble*	Ensemble*	Ensemble*	Ensemble*	Ensemble*
Antennarius striatus	Lophiiformes	Antennariidae	KC442213	KC442240	KC442109	KC442142	KC442182	AY362215
Lophius piscatorius	Lophiiformes	Lophiidae	JN230911	AY368325	JN230965	JN231065	JN231127	AY362274
Liza aurata	Mugiliformes	Mugilidae	KF017112	KF017144	KF017006	KF017049	KF017077	**KF312008**
Prionurus scalprum	Perciformes	Acanthuridae	KC442211	KC442238	KC442105	KC442139	KC442178	**KF312009**
Acropoma japonicum	Perciformes	Acropomatidae	KF017118	KF017148	KF017013	KF017054	KF017084	**KF312010**
Antigonia capros	Perciformes	Caproidae	AY308785	KC442237	KC442104	KC442138	KC442177	EU638027
Parastromateus niger	Perciformes	Carangidae	EF095654	EF095616	KC442097	KC442131	KC442170	**KF312011**
Trachurus trachurus	Perciformes	Carangidae	**KF311975**	**KF312119**	**KF311941**	**KF312057**	**KF312087**	
Lepomis macrochirus	Perciformes	Centrarchidae	AY430227	AY742577	KC442113	KC442146	KC442186	**KF312012**
Micropterus dolomieu	Perciformes	Centrarchidae	KF017143	AY742590	KF017040	KF017076	KF017111	
Centropomus undecimalis	Perciformes	Centropomidae	KC442207	KC442233	KC442098	KC442132	KC442171	**KF312013**
Channa maculata	Perciformes	Channidae	KF017114	KF017146	KF017008	KF017041	KF017079	
Cheilodactylus quadricornis	Perciformes	Cheilodactylidae	KF017131	KF017159	KF017027	KF017047	KF017097	**KF312014**
Astronotus ocellatus	Perciformes	Cichilidae	EF095671	EF095629	JN230960	JN231060	JN231122	**KF312015**
Coryphaena hippurus	Perciformes	Coryphaenidae	**KF311976**	**KF312120**	**KF311942**	**KF312058**	**KF312088**	**KF312016**
Echeneis neucratoides	Perciformes	Echeneidae	**KF311977**	**KF312121**	**KF311943**	**KF312059**	**KF312089**	AY362245
Elassoma evergladei	Perciformes	Elassomatidae	AY308784	KF017169	KF017037	KF017048	KF017108	
Gerres cinereus	Perciformes	Gerreidae	EF095666	EF095624	JN230966	JN231066	JN231128	**KF312017**
Howella zina	Perciformes	Howellidae	KF017116		KF017010	KF017052	KF017081	**KF312018**
Makaira sp.	Perciformes	Istiophoridae	**KF311978**	**KF312122**	**KF311944**	**KF312060**	**KF312090**	**KF312019**
Kuhlia mugil	Perciformes	Kuhlidae	KF017126	KF017154	KF017022	KF017060	KF017092	
Girella punctata	Perciformes	Kyphosidiae	KC442214	KC442244	KC442114	KC442147	KC442187	
Labrus bergylta	Perciformes	Labridae	EF095669	KC442239	KC442107	KC442141	KC442180	AY362222
Lactarius lactarius	Perciformes	Lactariidae	**KF311979**	**KF312123**	**KF311945**	**KF312061**	**KF312091**	
Lateolabrax japonicus	Perciformes	Lateolabracidae	EF095650	AY141293	KF017011	KF017053	KF017082	AY362253
Lates calcarifer	Perciformes	Latidae	JN230910	AY141294	JN230963	JN231063	JN231125	EU638059
Luvarus imperialis	Perciformes	Luvaridae	KC442212	EU637975	KC442106	KC442140	KC442179	EU638065
Mene maculata	Perciformes	Menidae	EF095659	AY141316	JN230962	JN231062	JN231124	AY362250
Dicentrarchus labrax	Perciformes	Moronidae	EF095651	Y18673	KC442100	KC442134	KC442173	**KF312020**
Morone saxatilis	Perciformes	Moronidae	KC442208	KC442234	KC442099	KC442133	KC442172	**KF312021**
Nandus nebulosus	Perciformes	Nandidae	KF017113	KF017145	KF017007	KF017050	KF017078	**KF312022**
Dissostichus mawsoni	Perciformes	Nototheniidae	KC442215	DQ498794	KC442115	KC442148	KC442188	**KF312023**
Maccullochella peelii	Perciformes	Percichthyidae	KF017134	KF017162	KF017029	KF017066	KF017100	**KF312024**
Etheostoma rufilineatum	Perciformes	Percidae	JN230912	JN231009	JN230967	JN231067	JN231129	**KF312025**
Perca fluviatilis	Perciformes	Percidae	KF017120	AY141295	KF017016	KF017043	KF017087	AY362279
Eleutheronema rhadinum	Perciformes	Polynemidae	**KF311980**	**KF312124**	**KF311946**	**KF312062**	**KF312092**	**KF312026**
Polydactylus sextarius	Perciformes	Polynemidae	**KF311981**	**KF312125**	**KF311947**	**KF312063**	**KF312093**	
Rachycentron canadum	Perciformes	Rachycentridae	**KF311982**	**KF312126**	**KF311948**	**KF312064**	**KF312094**	**KF312027**
Scarus psittacus	Perciformes	Scaridae	EF095675	EF095633	KC442108		KC442181	**KF312028**
Scomberomorus commerson	Perciformes	Scombridae	EF095676	EF095634	KC442094	KC442128	KC442167	**KF312029**
Holanthias chrysostictus	Perciformes	Serranidae	EF095645	AY141290	KF017014	KF017055	KF017085	AY362209
Paralabrax clathratus	Perciformes	Serranidae	KF017122	KF017150	KF017018	KF017058		
Siniperca chuatsi	Perciformes	Sinipercidae	KF017139	KF017167	KF017034	KF017071	KF017105	**KF312030**
Sparus aurata	Perciformes	Sparidae	EF095657	Y18665	KC442101	KC442135	KC442174	**KF312031**
Sphyraena argentea	Perciformes	Sphyraenidae	**KF311983**	**KF312127**	**KF311949**	**KF312065**	**KF312095**	**KF312032**
Symphysanodon katayamai	Perciformes	Symphysanodontida	KF017117	KF017147	KF017012	KF017042	KF017083	**KF312033**
Terapon jarbua	Perciformes	Terapontidae	KF017127	KF017155	KF017023	KF017061	KF017093	
Toxotes jaculatrix	Perciformes	Toxotidae	**KF311984**	**KF312128**	**KF311950**	**KF312066**	**KF312096**	**KF312034**
Trachinus draco	Perciformes	Trachinidae	KF017119	AY141304	KF017015	KF017056	KF017086	AY362277
Xiphias gladius	Perciformes	Xiphiidae	**KF311985**	DQ874811	**KF311951**	**KF312067**	**KF312097**	EU638098
Zoarces viviparus	Perciformes	Zoarcidae	KF017121	KF017149	KF017017	KF017057	KF017088	**KF312035**
Scorpaena onaria	Scorpaeniformes	Scorpaenidae	EF095642	AY141288	JN230968	JN231068	JN231130	AY362236
Mastacembelus erythrotaenia	Synbranchiformes	Mastacembelidae	KF017115	AY141275	KF017009	KF017051	KF017080	AY362249
Monopterus albus	Synbranchiformes	Synbranchidae	KC442205	AY141276	KC442095	KC442129	KC442168	AY362252
Balistes capriscus	Tetraodontiformes	Balistidae	AY700308	KC442242	KC442111	KC442144	KC442184	**KF312056**
Diodon holocanthus	Tetraodontiformes	Diodontidae	AY700325	KC442241	KC442110	KC442143	KC442183	
Takifugu rubripes	Tetraodontiformes	Tetraodontidae	AF108420	AF201471	Ensemble*	Ensemble*	Ensemble*	Ensemble*
Tetraodon nigroviridis	Tetraodontiformes	Tetraodontidae	NC007176	AJ293018	Ensemble*	Ensemble*	Ensemble*	CR649703
Triacanthodes anomalus	Tetraodontiformes	Triacanthodidae	AY308788	KC442243	KC442112	KC442145	KC442185	EU638095

Table S3.1B: Pleuronectiform taxa included in this study, the corresponding accession to the tissue (if any), and corresponding GenBank accession numbers for gene sequences. Newly determined sequences are in bold; * sequences retrieved from complete genomic database, Ensemble (http://www.ensembl.org/). Genes in this study are recombination activating gene 1 (RAG1), rhodopsin (RH), early growth response gene 1 (EGR1), early growth response gene 2B (E2B), early gowth response gene 3 (E3) and Mixed-lineage Leukemia (MLL).

Taxon Name	Order	Family	R1	RH	E1	E2B	E3	MLL
Trinectes maculatus	Pleuronectiformes	Achiridae	AY430224	EF095610	JN230964	JN231064	JN231126	EU638096
Mancopsetta maculata	Pleuronectiformes	Achiropsettidae	**KF311986**	**KF312129**	**KF311952**	**KF312068**	**KF312098**	**KF312036**
Arnoglossus laterna	Pleuronectiformes	Bothidae	**KF311987**	**KF312130**	**KF311953**	**KF312069**	**KF312099**	**KF312037**
Chascanopsetta lugubris	Pleuronectiformes	Bothidae	**KF311988**	**KF312131**	**KF311954**	**KF312070**		
Engyprosopon grandisquama	Pleuronectiformes	Bothidae	**KF311989**		**KF311955**	**KF312071**	**KF312100**	**KF312038**
Brachypleura novaezeelandiae	Pleuronectiformes	Citharidae		**KF312132**	**KF311956**		**KF312101**	**KF312039**
Citharoides macrolepis	Pleuronectiformes	Citharidae	**KF311990**	**KF312133**	**KF311957**		**KF312102**	**KF312040**
Citharus linguatula	Pleuronectiformes	Citharidae	**KF311991**	**KF312134**	**KF311958**	**KF312072**	**KF312103**	AY362232
Lepidoblepharon ophthalmolepis	Pleuronectiformes	Citharidae	**KF311992**	**KF312135**	**KF311959**	**KF312073**	**KF312104**	**KF312041**
Cynoglossus lingua	Pleuronectiformes	Cynoglossidae	**KF311993**		**KF311960**	**KF312074**	**KF312105**	
Paraplagusia japonica	Pleuronectiformes	Cynoglossidae	**KF311994**	**KF312136**	**KF311961**	**KF312075**	**KF312106**	
Symphurus orientalis	Pleuronectiformes	Cynoglossidae	**KF311995**	**KF312137**	**KF311962**			**KF312042**
Paralichthys olivaceus	Pleuronectiformes	Paralichthyidae	KC442210	KC442236	KC442103	KC442137	KC442176	**KF312043**
Pseudorhombus oligodon	Pleuronectiformes	Paralichthyidae	**KF311996**	**KF312138**	**KF311963**	**KF312076**	**KF312107**	**KF312044**
Xystreurys liolepis	Pleuronectiformes	Paralichthyidae	**KF311997**	**KF312139**	**KF311964**	**KF312077**	**KF312108**	**KF312045**
Eopsetta jordani	Pleuronectiformes	Pleuronectidae	**KF311998**	**KF312140**	**KF311965**	**KF312078**	**KF312109**	**KF312046**
Hippoglossus stenolepis	Pleuronectiformes	Pleuronectidae	**KF311999**	**KF312141**	**KF311966**	**KF312079**	**KF312110**	**KF312047**
Limanda limanda	Pleuronectiformes	Pleuronectidae	**KF312000**	**KF312142**	**KF311967**	**KF312080**	**KF312111**	**KF312048**
Poecilopsetta beani	Pleuronectiformes	Pleuronectidae	**KF312001**	**KF312143**	**KF311968**	**KF312081**	**KF312112**	**KF312049**
Poecilopsetta plinthus	Pleuronectiformes	Pleuronectidae	**KF312002**	**KF312144**	**KF311969**	**KF312082**	**KF312113**	**KF312050**
Psettodes erumei	Pleuronectiformes	Psettodidae	KC442209	KC442235	KC442102	KC442136	KC442175	**KF312051**
Samaris cristatus	Pleuronectiformes	Samaridae	**KF312003**	**KF312145**	**KF311970**		**KF312114**	**KF312052**
Samariscus latus	Pleuronectiformes	Samaridae	**KF312004**	**KF312146**	**KF311971**	**KF312083**	**KF312115**	
Scophthalmus rhombus	Pleuronectiformes	Scophthalmidae	**KF312005**	**KF312147**	**KF311972**	**KF312084**	**KF312116**	**KF312053**
Pegusa lascaris	Pleuronectiformes	Soleidae	**KF312006**	**KF312148**	**KF311973**	**KF312085**	**KF312117**	**KF312054**
Solea vulgaris (solea)	Pleuronectiformes	Soleidae	EF095644	Y18672	**KF311974**	**KF312086**	**KF312118**	**KF312055**

Table S3.2: Primers used in this study and their sources.

Gene Name	Primer Name	Primer Sequence 5'-3'	Source
Recombination Activating Gene 1			
	R1 2533F	CTGAGCTGCAGTCAGTACCATAAGATGT	López et al. 2004
	R1 4078R	TGAGCCTCCATGAACTTCTGAAGRTAYTT	López et al. 2004
	R1 4061R	AATACTTGGAGGTGTAGAGCCAGT	Chen et al. 2007
	R1 4090R	CTGAGTCCTTGTGAGCTTCCATRAAYTT	López et al. 2004
Rhodopsin			
	RH 1F	ATGAACGGCACAGARGGAC	Chen et al. 2013
	RH PcoF1	CATCGTCCGGAGTCCTTATG	Chen et al. 2013
	RH 193F	CNTATGAATAYCCTCAGTACTACC	Chen et al. 2003
	RH 1039R	TGCTTGTTCATGCAGATGTAGA	Chen et al. 2003
	RH 1073R-modif	CCRCAGCACAGRGTGGTGATCATG	Chen et al. 2003
Early Growth Response Protein 1			
	E1 225F	CCTGAYATCCCCTTCAACTGTG	Chen et al. 2013
	E1 284F	CCCCCATCTCYTACACAGG	Chen et al. 2013
	E1 290F	TMTCTTACACAGGCCGYTTCAC	Chen et al. 2008
	E1 333F	CAGYAACAGTCTRTGGGCTGAG	Chen et al. 2008
	E1 1104R	CGCAGGTGGATCTTRGTGTG	Chen et al. 2008
	E1 1118R	CTTCTTGTCCTTCTGCCGYAGRT	Chen et al. 2013
	E1 1126R	CTTTYTCTGCTTTCTTGTCCTTCT	Chen et al. 2008
Early Growth Response Protein 2B			
	E2B 252F	CGCAACCAGACTTTCACCTAY	Chen et al. 2013
	E2B 261F	TTCACCTAYATGGGNAAGTTCTCMAT	Chen et al. 2013
	E2B 270F	ATGGGRAAGTTCTCCATCGAC	Chen et al. 2013
	E2B 278F	AGTTTTCCATCGACTCSCAGTA	Chen et al. 2008
	E2B 287F	TTGACTCSCAGTATCCAGGTAAC	Chen et al. 2008
	E2B 1078R	AATTTGCGNCCGCAGSAGTC	Chen et al. 2013
	E2B 1078R-bis	GAACTTACGNCCGCAGAARTC	Chen et al. 2013
	E2B 1108R	TTTTGTGTGTCTCTTTCTYTCGTC	Chen et al. 2008
	E2B 1112R	ATTTTNGTGTGTCGYTTYCTC	Chen et al. 2013
	E2B 1117R	AGGTGGATTTTGGTGTGTCTYTT	Chen et al. 2008
	E2B 1121R	CCTCAGGTGGATTTTAGTGTGTC	Chen et al. 2013
Early Growth Response Protein 3			
	E3 161F	AATATCATGGACYTGGGNATGG	Chen et al. 2008
	E3 254F	GTCACCTAYYTGGGSAAGTTT	Chen et al. 2008
	E3 1068R	GTCCRCAGAACTCGCARGAGA	Chen et al. 2013
Mixed-lineage Leukemia			
	MLL 1459F	TCCCAGACTCARGTTTCCAG	This Study
	MLL U1506	CAGCAGTTCCAGCCYCTSTA	Dettaï & Lecointre 2005
	MLL L2127	CWGNTTTTGGTCTYTTGATNATATT	Dettaï & Lecointre 2005
	MLL 2170R	CTCTGCTGAAKGAGAGTAGTKGG	This Study

Table S3.3A: Taxa that failed at least one X^2 test of base composition at one gene. Mean GC content of each gene is reported. Test values are only repeated for failures. Values are reported for Recombination Activating Gene 1 and Rhodopsin.

Order	Family	Taxon	Recombination Activating Gene 1			Rhodopsin		
			p-value	GC Content	Difference from Mean GC Content	p-value	GC Content	Difference from Mean GC Content
Lophiiformes	Lophiidae	*Lophius piscatorius*	0.54%	0.54	-0.059	0.00%	0.7	0.11
Lophiiformes	Antennariidae	*Antennarius striatus*	0.00%	0.51	-0.089	0.10%	0.68	0.09
Atheriniformes	Melanotaeniidae	*Melanotaenia lacustris*						
Beryciformes	Berycidae	*Beryx splendens*						
Beryciformes	Holocentridae	*Myripristis murdjan*	0.96%	0.56	-0.039			
Incertae sedis	Elassomatidae	*Elassoma evergladei*						
Gasterosteiformes	Gasterosteidae	*Gasterosteus aculeatus*	2.47%	0.64	0.041	0.00%	0.77	0.18
Synbranchiformes	Synbranchidae	*Monopterus albus*				1.99%	0.51	-0.08
Synbranchiformes	Mastacembelidae	*Mastacembelus erythrotaenia*	1.29%	0.54	-0.059			
Scorpaeniformes	Dactylopteridae	*Dactyloptena orientalis*						
Scorpaeniformes	Scorpaenidae	*Scorpaena onaria*	3.68%	0.64	0.041			
Perciformes	Howellidae	*Howella zina*	0.00%	0.71	0.111			
Perciformes	Serranidae	*Holanthias chrysostictus*						
Perciformes	Terapontidae	*Terapon jarbua*						
Perciformes	Percidae	*Etheostoma rufilineatum*						
Perciformes	Percidae	*Perca fluviatilis*						
Perciformes	Lactariidae	*Lactarius lactarius*						
Perciformes	Sparidae	*Sparus aurata*	0.01%	0.67	0.071			
Perciformes	Nandidae	*Nandus nebulosus*						
Perciformes	Cichilidae	*Astronotus ocellatus*	4.96%	0.56	-0.039			
Perciformes	Mugilidae	*Liza aurata*						
Perciformes	Sphyraenidae	*Sphyraena argentea*						
Perciformes	Labridae	*Labrus bergylta*						
Perciformes	Scaridae	*Scarus psittacus*						
Perciformes	Zoarcidae	*Zoarces viviparus*				0.00%	0.72	0.13
Perciformes	Nototheniidae	*Dissostichus mawsoni*						
Pleuronectiformes	Psettodidae	*Psettodes erumei*						
Pleuronectiformes	Citharidae	*Citharus linguatula*	0.67%	0.65	0.051			
Pleuronectiformes	Paralichthyidae	*Pseudorhombus oligodon*						
Pleuronectiformes	Bothidae	*Arnoglossus laterna*	0.00%	0.72	0.121	0.12%	0.68	0.09
Pleuronectiformes	Bothidae	*Chascanopsetta lugubris*	0.00%	0.74	0.141	0.00%	0.74	0.15
Pleuronectiformes	Bothidae	*Engyprosopon grandisquama*	0.00%	0.68	0.081			
Pleuronectiformes	Achiropsettidae	*Mancopsetta maculata*	0.06%	0.66	0.061			
Pleuronectiformes	Pleuronectidae	*Poecilopsetta beani*	0.00%	0.69	0.091	1.99%	0.65	0.06
Pleuronectiformes	Pleuronectidae	*Poecilopsetta plinthus*	0.00%	0.69	0.091	2.10%	0.65	0.06
Pleuronectiformes	Samaridae	*Samaris cristatus*						
Pleuronectiformes	Samaridae	*Samariscus latus*						
Pleuronectiformes	Achiridae	*Trinectes maculatus*						
Pleuronectiformes	Soleidae	*Pegusa lascaris*	1.49%	0.54	-0.059			
Pleuronectiformes	Soleidae	*Solea vulgaris (solea)*	0.74%	0.54	-0.059			
Pleuronectiformes	Cynoglossidae	*Cynoglossus lingua*						
Pleuronectiformes	Cynoglossidae	*Paraplagusia japonica*	2.51%	0.54	-0.059	4.09%	0.64	0.05
Pleuronectiformes	Cynoglossidae	*Symphurus orientalis*	0.00%	0.7	0.101			
Tetraodontiformes	Balistidae	*Balistes capriscus*				4.47%	0.65	0.06
Tetraodontiformes	Tetraodontidae	*Takifugu rubripes*						
Tetraodontiformes	Tetraodontidae	*Tetraodon nigroviridis*	0.09%	0.53	-0.069			
Tetraodontiformes	Diodontidae	*Diodon holocanthus*	0.00%	0.51	-0.089			
Mean GC Content:				**0.60**			**0.59**	
Number of Variable Sites:				**883**			**431**	

Table S3.3B: Taxa that failed at least one X^2 test of base composition at one gene. Mean GC content of each gene is reported. Test values are only repeated for failures. Values are reported for Early Growth Response Protein 1 and Early Growth Response Protein 2B.

Order	Family	Taxon	Early Growth Response Protein 1			Early Growth Response Protein 2B		
			p-value	GC Content	Difference from Mean GC Content	p-value	GC Content	Difference from Mean GC Content
Lophiiformes	Lophiidae	*Lophius piscatorius*						
Lophiiformes	Antennariidae	*Antennarius striatus*	0.29%	0.59	-0.09	0.03%	0.79	0.09
Atheriniformes	Melanotaeniidae	*Melanotaenia lacustris*						
Beryciformes	Berycidae	*Beryx splendens*				1.33%	0.76	0.06
Beryciformes	Holocentridae	*Myripristis murdjan*						
Incertae sedis	Elassomatidae	*Elassoma evergladei*				2.03%	0.63	-0.07
Gasterosteiformes	Gasterosteidae	*Gasterosteus aculeatus*						
Synbranchiformes	Synbranchidae	*Monopterus albus*						
Synbranchiformes	Mastacembelidae	*Mastacembelus erythrotaenia*						
Scorpaeniformes	Dactylopteridae	*Dactyloptena orientalis*						
Scorpaeniformes	Scorpaenidae	*Scorpaena onaria*						
Perciformes	Howellidae	*Howella zina*				0.00%	0.82	0.12
Perciformes	Serranidae	*Holanthias chrysostictus*						
Perciformes	Terapontidae	*Terapon jarbua*				0.34%	0.77	0.07
Perciformes	Percidae	*Etheostoma rufilineatum*						
Perciformes	Percidae	*Perca fluviatilis*						
Perciformes	Lactariidae	*Lactarius lactarius*	3.36%	0.74	0.06			
Perciformes	Sparidae	*Sparus aurata*						
Perciformes	Nandidae	*Nandus nebulosus*				0.00%	0.82	0.12
Perciformes	Cichilidae	*Astronotus ocellatus*						
Perciformes	Mugilidae	*Liza aurata*				3.53%	0.64	-0.06
Perciformes	Sphyraenidae	*Sphyraena argentea*				2.58%	0.76	0.06
Perciformes	Labridae	*Labrus bergylta*				0.08%	0.77	0.07
Perciformes	Scaridae	*Scarus psittacus*	0.52%	0.75	0.07			
Perciformes	Zoarcidae	*Zoarces viviparus*						
Perciformes	Nototheniidae	*Dissostichus mawsoni*						
Pleuronectiformes	Psettodidae	*Psettodes erumei*						
Pleuronectiformes	Citharidae	*Citharus linguatula*	0.00%	0.82	0.14	0.02%	0.79	0.09
Pleuronectiformes	Paralichthyidae	*Pseudorhombus oligodon*	0.37%	0.75	0.07	0.16%	0.78	0.08
Pleuronectiformes	Bothidae	*Arnoglossus laterna*	1.12%	0.75	0.07	0.05%	0.78	0.08
Pleuronectiformes	Bothidae	*Chascanopsetta lugubris*	0.01%	0.77	0.09	0.06%	0.78	0.08
Pleuronectiformes	Bothidae	*Engyprosopon grandisquama*	0.00%	0.88	0.20	0.00%	0.82	0.12
Pleuronectiformes	Achiropsettidae	*Mancopsetta maculata*						
Pleuronectiformes	Pleuronectidae	*Poecilopsetta beani*						
Pleuronectiformes	Pleuronectidae	*Poecilopsetta plinthus*						
Pleuronectiformes	Samaridae	*Samaris cristatus*						
Pleuronectiformes	Samaridae	*Samariscus latus*				4.01%	0.63	-0.07
Pleuronectiformes	Achiridae	*Trinectes maculatus*				2.52%	0.63	-0.07
Pleuronectiformes	Soleidae	*Pegusa lascaris*						
Pleuronectiformes	Soleidae	*Solea vulgaris (solea)*						
Pleuronectiformes	Cynoglossidae	*Cynoglossus lingua*	0.10%	0.76	0.08			
Pleuronectiformes	Cynoglossidae	*Paraplagusia japonica*						
Pleuronectiformes	Cynoglossidae	*Symphurus orientalis*	1.91%	0.73	0.05			
Tetraodontiformes	Balistidae	*Balistes capriscus*						
Tetraodontiformes	Tetraodontidae	*Takifugu rubripes*				0.00%	0.59	-0.11
Tetraodontiformes	Tetraodontidae	*Tetraodon nigroviridis*						
Tetraodontiformes	Diodontidae	*Diodon holocanthus*	0.01%	0.57	-0.11			
Mean GC Content:				**0.68**			**0.70**	
Number of Variable Sites:				**446**			**477**	

Table S3.3C: Taxa that failed at least one X^2 test of base composition at one gene. Mean GC content of each gene is reported. Test values are only repeated for failures. Values are reported for Early Growth Response Protein 3 and Mixed-lineage Leukemia.

Order	Family	Taxon	Early Growth Response Protein 3			Mixed-lineage Leukemia		
			p-value	GC Content	Difference from Mean GC Content	p-value	GC Content	Difference from Mean GC Content
Lophiiformes	Lophiidae	*Lophius piscatorius*	0.08%	0.56	-0.094			
Lophiiformes	Antennariidae	*Antennarius striatus*						
Atheriniformes	Melanotaeniidae	*Melanotaenia lacustris*	0.01%	0.54	-0.114			
Beryciformes	Berycidae	*Beryx splendens*	0.25%	0.74	0.086			
Beryciformes	Holocentridae	*Myripristis murdjan*						
Incertae sedis	Elassomatidae	*Elassoma evergladei*						
Gasterosteiformes	Gasterosteidae	*Gasterosteus aculeatus*						
Synbranchiformes	Synbranchidae	*Monopterus albus*	2.06%	0.58	-0.074			
Synbranchiformes	Mastacembelidae	*Mastacembelus erythrotaenia*	0.10%	0.55	-0.104			
Scorpaeniformes	Dactylopteridae	*Dactyloptena orientalis*	0.53%	0.73	0.076			
Scorpaeniformes	Scorpaenidae	*Scorpaena onaria*						
Perciformes	Howellidae	*Howella zina*						
Perciformes	Serranidae	*Holanthias chrysostictus*				0.00%	0.7	0.14
Perciformes	Terapontidae	*Terapon jarbua*						
Perciformes	Percidae	*Etheostoma rufilineatum*	0.19%	0.74	0.086			
Perciformes	Percidae	*Perca fluviatilis*	0.00%	0.83	0.176			
Perciformes	Lactariidae	*Lactarius lactarius*						
Perciformes	Sparidae	*Sparus aurata*						
Perciformes	Nandidae	*Nandus nebulosus*						
Perciformes	Cichilidae	*Astronotus ocellatus*						
Perciformes	Mugilidae	*Liza aurata*						
Perciformes	Sphyraenidae	*Sphyraena argentea*	1.40%	0.73	0.076			
Perciformes	Labridae	*Labrus bergylta*						
Perciformes	Scaridae	*Scarus psittacus*						
Perciformes	Zoarcidae	*Zoarces viviparus*						
Perciformes	Nototheniidae	*Dissostichus mawsoni*	3.52%	0.72	0.066			
Pleuronectiformes	Psettodidae	*Psettodes erumei*				0.03%	0.66	0.10
Pleuronectiformes	Citharidae	*Citharus linguatula*				0.00%	0.71	0.15
Pleuronectiformes	Paralichthyidae	*Pseudorhombus oligodon*						
Pleuronectiformes	Bothidae	*Arnoglossus laterna*	0.00%	0.85	0.196			
Pleuronectiformes	Bothidae	*Chascanopsetta lugubris*						
Pleuronectiformes	Bothidae	*Engyprosopon grandisquama*	0.00%	0.87	0.216	0.01%	0.66	0.10
Pleuronectiformes	Achiropsettidae	*Mancopsetta maculata*						
Pleuronectiformes	Pleuronectidae	*Poecilopsetta beani*						
Pleuronectiformes	Pleuronectidae	*Poecilopsetta plinthus*						
Pleuronectiformes	Samaridae	*Samaris cristatus*	2.24%	0.62	-0.034			
Pleuronectiformes	Samaridae	*Samariscus latus*						
Pleuronectiformes	Achiridae	*Trinectes maculatus*				2.18%	0.5	-0.07
Pleuronectiformes	Soleidae	*Pegusa lascaris*	0.03%	0.76	0.106			
Pleuronectiformes	Soleidae	*Solea vulgaris (solea)*	0.02%	0.76	0.106			
Pleuronectiformes	Cynoglossidae	*Cynoglossus lingua*	0.23%	0.74	0.086			
Pleuronectiformes	Cynoglossidae	*Paraplagusia japonica*						
Pleuronectiformes	Cynoglossidae	*Symphurus orientalis*						
Tetraodontiformes	Balistidae	*Balistes capriscus*						
Tetraodontiformes	Tetraodontidae	*Takifugu rubripes*						
Tetraodontiformes	Tetraodontidae	*Tetraodon nigroviridis*						
Tetraodontiformes	Diodontidae	*Diodon holocanthus*						
Mean GC Content:				**0.65**		**0.57**		
Number of Variable Sites:				**366**		**443**		

Table S3.4: Prior characteristics of calibration points used in divergence time estimation, the taxa whose most recent common ancestor (MRCA) is dated by the calibration point, and the source of the calibration point.

Calibration	Prior Offset	Mean	SD	Source	Dating MRCA of Which Taxa in This Analysis
Centrarchidae	23	0.776	0.8	Albright 1994	*Lepomis macrochirus* and *Micropterus dolomieui*
Stem Echeneidae	30.1	0.165	0.8	Near et al. 2012	*Coryphaena hippurus, Echeneis neucratoides* and *Rachycentron canadum*
Stem Balistidae	37.2	0.37	0.8	Near et al. 2012	*Balistes capriscus* and *Triacanthodes anomalus*
Channoidea	48	1.71	1.14	Santini et al. 2009	*Channa maculata* and *Nandus nebulosus*
Centropomidae	48.6	1.0	1.0	Otero 2004	*Centropomus undecimalis* and *Lates calcarifer*
Crown Labrids	50	0.9	1.6	Santini et al. 2009	*Labrus bergylta* and *Scarus psittacus*
Stem Diodontidae	50	0.672	0.8	Near et al. 2012	*Diodon holocanthus*, *Takifugu rubripes* and *Tetraodon nigroviridis*
Antennariidae	50	0.776	1.0	Carnevale & Pietsch, 2009	*Antennarius striatus* and *Lophius piscatorius*
Stem Carangidae	55.8	0.776	0.8	Near et al. 2012	*Coryphaena hippurus*, *Echeneis neucratoides, Parastromateus niger*, *Rachycentron canadum and Trachurus trachurus*
Stem Luvaridae	55.8	0.776	0.8	Near et al. 2012	*Luvarus imperialis* and *Prionurus scalprum*
Beryx fossil	93.5	0.5	0.8	Palci et al. 2008	*Beryx splendens* and *Myripristis murdjan*

References

Aberer, A.J., Krompass, D., Stamatakis, A., 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Syst. Biol. 62, 162–166.

Akashi, H., Kliman, R., Eyre-Walker, A., 1998. Mutation pressure, natural selection, and the evolution of base composition in Drosophila. Genetica 102–103, 49–60.

Albright, L.B., 1994. Lower vertebrates from an Arikareean (Earliest Miocene) fauna near Toledo Bend Dam, Newton County, Texas. J. Pal. 68:1131-1145.

Alff-Steinberger, C., 1969. The genetic code and error transmission. Proc. Natl. Acad. Sci. USA 64, 584–591.

Amaoka, K., 1969. Studies on the sinistral flounder found in the waters around Japan. Taxonomy, anatomy, and phylogeny. J. Shimonoseki Univ. Fish. 18, 65–340.

Azevedo, M.F.C., Oliveira, C., Pardo, B.G., Martínez, P., Foresti, F., 2008. Phylogenetic analysis of the order Pleuronectiformes (Teleostei) based on sequences of 12S and 16S mitochondrial genes. Genet. Mol. Biol. 31, 284–292.

Benton, M.J., Donoghue, P.C.J., Asher, R.J., 2009. Calibrating and constraining molecular clocks. In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, Oxford, pp. 35–86.

Berendzen, P.B., Dimmick, W.W., 2002. Phylogenetic relationships of Pleuronectiformes based on molecular evidence. Copeia 2002, 642–652.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.

Betancur-R., R., Broughton, R.E.,Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I.,Arcila, D., Sanciangco, M., Cureton II, J.C., Zhang, F., Buser, T., Campbell, M.A., Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C., Hough, D.J., Lu, G., Grande, T., Arratia, G., Ortí, G., 2013a. The Tree of Life and a new classification of bony fishes. PLoS Curr. (2013 April 18 [last modified 23.04.13].Edition 1. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288).

Betancur-R., R., Li, C., Munroe, T.A., Ballesteros, J.A., Ortí, G., 2013b. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol., in press.

Brewster, B., 1987. Eye migration and cranial development during flatfish metamorphosis: a reappraisal (Teleostei: Pleuronectiformes). J. Fish Biol. 31, 805–833.

Bruno, W.J., Halpern, A.L., 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. 16, 564–566.

Buckley, T.R., Cunningham, C.W., 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. Mol. Biol. Evol. 19, 394–405.

Cao, Y., Adachi, J., Janke, A., Pääbo, S., Hasegawa, M., 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J. Mol. Evol. 39, 519–527.

Carnavale, G., Pietchs, T.W., 2009. An Eocene frogfish from Monte Bolca, Italy: The earliest known skeletal record for the family. Paleontology. 52:745-752.

Chabanaud, P., 1937. Les téléostéens dyssymétriques du Mokkatam Inférieur de Tourah. Mém. Inst. Egypte 31, 1–122.

Chabanaud, P., 1949. Le problème de la phylogénèse des Heterosomata. Bull. Inst. Oceanogr. 950, 1–24.

Chanet, B., 1997. A cladistic reappraisal of the fossil flatfishes record consequences on the phylogeny of the Pleuronectiformes (Osteichthyes: Teleostei). Ann. Sci. Nat. Zool. 18, 105–117.

Chanet, B., 1999. Supposed and true flatfishes [Teleostei: Pleuronectiformes] from the Eocene of Monte Bolca, Italy. Stud. Ric. Giacim. Terz. Bolca. 8, 220–243.

Chapleau, F., 1993. Pleuronectiform relationships: a cladistic reassessment. Bull. Mar. Sci. 52, 516–540.

Chen, W.-J., 2001. La répétitivité des clades omme critère de la fiabilité: application à la phylogénie des Acanthomorpha (Teleostei) et des Notothenioidei (acanthomorphes antarctiques). Ph.D. Thesis. University of Paris VI.

Chen, W.-J., Mayden, R.L., 2010. A phylogenomic perspective on the new era of Ichthyology. Bioscience 60, 421–432.

Chen, W.-J., 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. Phylogenet. Evol. 26, 262–288.

Chen, W.-J., Ruiz-Carus, R., Ortí, G., 2007. Relationships among four genera of mojarras (Teleostei: Perciformes: Gerreidae) from the western Atlantic and their tentative placement among percomorph fishes. J. Fish Biol. 70, 202–218.

Chen, W.-J., Miya, M., Saitoh, K., Mayden, R.L., 2008. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of rayfinned fishes: the order Cypriniformes (Ostariophysi) as a case study. Gene 423, 125–134.

Cole, F.J., Johnstone, J., 1902. Pleuronectes (the plaice). Mem. Liverpool Mar. Biol. Commun. 8, 1–252.

Darwin, C., 1872. The Origin of Species by Means of Natural Selection. John Murray, London.

Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on ‘‘Hexapod origins: Monophyletic or paraphyletic?’’. Science 301, 1482.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.

Dettai, A., Lecointre, G., 2005. Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. C.R. Biol. 328, 674–689.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Edgar, R.C., 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113.

Edgar, R.C., 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Eyre-Walker, A., 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics 152, 675–683.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Foster, P.G., 2004. Modeling compositional heterogeneity. Syst. Biol. 53, 485–495.

Foster, P.G., Hickey, D.A., 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48, 284–290.

Frazzetta, T., 2012. Flatfishes, turtles, and bolyerine snakes: evolution by small steps or large, or both? Evol. Biol. 39, 30–60.

Friedman, M., 2008. The evolutionary origin of flatfish asymmetry. Nature 454, 209–212.

Friedman, M., 2012. Osteology of †*Heteronectes chaneti* (Acanthomorpha, Pleuronectiformes), an Eocene stem flatfish, with a discussion of flatfish sister-group relationships. J. Vertebr. Paleontol. 32, 735–756.

Fukuda, E., Nakae, M., Asaoka, R., Sasaki, K., 2010. Branching patterns of trunk lateral line nerves in Pleuronectiformes: uniformity and diversity. Ichthyol. Res. 57, 148–160.

Galtier, N., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc. Natl. Acad. Sci. USA 92, 11317–11321.

Gosline, W.A., 1971. Functional Morphology and Classification of Teleostean Fishes. University of Hawaii Press, Honolulu.

Greenwood, P.H., 1976. A review of the family Centropomidae (Pisces, Perciformes). Bull. Br. Mus. Nat. Hist. (Zool.) 29, 1–81.

Haig, D., Hurst, L., 1991. A quantitative measure of error minimization in the genetic code. J. Mol. Evol. 33, 412–417.

Hensley, D.A., 1997. An overview of the systematics and biogeography of the flatfishes. J. Sea Res. 37, 187–194.

Hensley, D.A., 1999. Psettoididae. Spiny Turbots. In: Carpenter, K.E., Niem, V. (Eds.), FAO Identification Guide for Fishery Purposes. The Wesern Central Pacific. Food and Agriculture Organization of the United Nations, Rome, pp. 3792–3793.

Hewer, H.R., 1931. Studies in colour-changes in fish.—Part V. The colour-patterns in certain flat-fish and their relation to the environment. Zool. J. Linn. Soc. – Lond. 37, 493–513.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52, 124–126.

Holt, E.W.L., 1894. Studies in teleostean morphology from the Marine Laboratory at Cleethorpes. Proc. Zool. Soc. London 1894, 413–446.

Hubbs, C.L., 1945. Phylogenetic position of the Citharidae, a family of flatfishes.Misc. Pub., Museum Zool. Univ. Mich. 63, 1–38.

Janvier, P., 2008. Palaeontology: squint of the fossil flatfish. Nature 454, 169–170.

Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, A.W.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53, 638–643.

Johnson, G.D., 1983. Niphon spinosus: a primitive epinepheline serranid, with comments on the monophyly and intrarelationships of the Serranidae. Copeia 1983, 777–787.

Johnson, G.D., 1984. Percoidei: development and relationships. In: Moser, H.G., Richards, W.J., Cohen, D.M., Fahay, M.P., Kendall, A.W., Jr., Richardson, S.L. (Eds.), Ontogeny and Systematics of Fishes. American Society of Ichthyology and Herpetology, Lawrence, Kansas, pp. 464–478.

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

Kyle, H.M., 1921. The asymmetry, metamorphosis and origin of flat-fishes. Philos. Trans. Roy. Soc. B 211, 75–129.

Lamarck, J.B., 1809. Philosophie Zoologique. Chez Detu, Paris.

Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.

Li, C., Ortí, G., 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. Mol. Phylogenet. Evol. 44, 386–398.

Li, B., Dettaï, A., Cruaud, C., Couloux, A., Desoutter-Meniger, M., Lecointre, G., 2009. RNF213, a new nuclear marker for acanthomorph phylogeny. Mol. Phylogenet. Evol. 50, 345–363.

Li, C., Ricardo, B.R., Smith, L.W., Ortí, G., 2011. Monophyly and interrelationships of Snook and Barramundi (Centropomidae *sensu* Greenwood) and five new markers for fish phylogenetics. Mol. Phylogenet. Evol. 60, 463–471.

Little, A.G., Lougheed, S.C., Moyes, C.D., 2010. Evolutionary affinity of billfishes (Xiphiidae and Istiophoridae) and flatfishes (Plueronectiformes): independent and trans-subordinal origins of endothermy in teleost fishes. Mol. Phylogenet. Evol. 56, 897–904.

López, J.A., Chen, W.-J., Ortí, G., 2004. Esociform phylogeny. Copeia 2004, 449–464.

Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary Analysis. Version 2.75. <http://mesquiteproject.org>.

Mitchell, A., Mitter, C., Regier, J.C., 2000. More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Syst. Biol. 49, 202–224.

Mivart, G.J., 1871. On the Genesis of Species. MacMillan, London.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 26, 121–138.

Mooers, A.O., Holmes, E.C., 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. 15, 365–369.

Mooi, R.D., Gill, A.C., 1995. Association of epaxial musculature with dorsal-fin pterygiophores in acanthomorph fishes, and its phylogenetic significance. Bull. Br. Mus. Nat. Hist. (Zool.) 61, 121–137.

Munroe, T.A., 2005. Systematic diversity of the Pleuronectiformes. In: Gibson, R.N. (Ed.), Flatfishes: Biology and Exploitation. Blackwell, Oxford, pp. 10–41.

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., Smith, W.L., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. USA 109, 13698–13703.

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbink, F.T., Friedman, M., Wainwright, P.C., 2013. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. USA, in press.

Nelson, J.S., 2006. Fishes of the World. John Wiley & Sons, Inc., Hoboken, New Jersey.

Norman, J.R., 1934. A systematic monograph of the flatfishes (Hetersomata). Psettodidae, Bothidae, Pleuronectidae, vol. 2. Printed by order of the Trustees of the British Museum, London.

Otero, O., 2004. Anatomy, systematics and phylogeny of both recent and fossil latid fishes (Teleostei, Perciformes, Latidae). Zool. J. Linn. Soc. – Lond. 141, 81–133.

Palci, A., Jurkovšek, B., Kolar- Jurkovšek, T., Caldwell, M. W., 2008. New palaeoenvironmental model for the Komen (Slovenia) Cenomanian (Upper Cretaceous) fossil lagerstätte. Cret. Res. 29:316-328.

Palmer, A.R., 1996. From symmetry to asymmetry: phylogenetic patterns of asymmetry variation in animals and their evolutionary significance. Proc. Natl. Acad. Sci. USA 93, 14279–14286.

Patterson, C., 1993. An overview of the early fossil record of acanthomorphs. Bull. Mar. Sci. 52, 29–59.

Penny, D., Steel, M.A., Lockhart, P.J., Hendy, M.D., 1994. The role of models in reconstructing evolutionary trees. In: Scotland, R.W., Siebert, D.J., Williams, D.M. (Eds.), Models in Phylogeny Reconstruction. Oxford University Press, Oxford, pp. 211–230.

Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171–185.

Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.

Platt, C., 1983. Retention of generalized hair cell patterns in the inner ear of the primitive flatfish Psettodes. Anat. Rec. 207, 503–508.

Posada, D., 2001. The effect of branch length variation on the selection of models in molecular evolution. J. Mol. Evol. 52, 434–444.

Regan, C.T., 1910. The origin and evolution of the teleostean fishes of the order Heterosomata. Ann. Mag. Nat. Hist. 8, 484–496.

Regan, C.T., 1929. Fishes. Heterosomata. Encyclopedia Britannica, London, pp. 324–325.

Roje, D.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). Mol. Phylogenet. Evol. 56, 586–600.

Russo, C.A., Takezaki, N., Nei, M., 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. 13, 525–536.

Santini, F., Harmon, L., Carnevale, G., Alfaro, M., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Biol. 9:194.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.

Schwarzhans, W., 1999. Piscium Catalogus: Otolithi Piscium. Verlag Dr. Friedrich Pfeil, Munchen.

Sheffield, N.C., Song, H., Cameron, S.L., Whiting, M.F., 2009. Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. Syst. Biol. 58, 381–394.

Shi, W., Kong, X.Y., Wang, Z.M., Jiang, J.X., 2011. Utility of tRNA genes from the complete mitochondrial genome of Psetta maxima for implying a possible sistergroup relationship to the Pleuronectiformes. Zool. Stud. 50, 665–681.

Smith, W.L., Craig, M.T., 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, 35–55.

Smith, W.L., Wheeler, W.C., 2006. Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. J. Hered. 97, 206–217.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixedmodels. Bioinformatics 22, 2688–2690.

Steel, M.A., Lockhart, P.J., Penny, D., 1993. Confidence in evolutionary trees from biological sequence data. Nature 364, 440–442.

Wainwright, P.C., Smith, W.L., Price, S.A., Tang, K.L., Sparks, J.S., Ferry, L.A., Kuhn, K.L., Eytan, R.I., Near, T.J., 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Syst. Biol. 61, 1001–1027.

Walker, J.D., Geismann, J.W., 2009. 2009 GSA Geologic Time Scale. GSA Today. April/May, 60–61.

Woese, C.R., 1965. On the evolution of the genetic code. Proc. Natl. Acad. Sci. USA 54, 1546–1552.

Wolf, Y.I., Rogozin, I.B., Koonin, E.V., 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. Genome Res. 14, 29–36.

Wortley, A.H., Rudall, P.J., Harris, D.J., Scotland, R.W., 2005. Howmuch data are needed to resolve a difficult phylogeny? Case study in Lamiales. Syst. Biol. 54, 697–709.

Yamanaka, Y., Nakae, M., Fukuda, E., Sasaki, K., 2010. Monophyletic origin of the dorsally arched lateral line in Teleostei: evidence from nerve innervations patterns. Ichthyol. Res. 57, 49–61.

CHAPTER 4: MITOCHONDRIAL GENOMIC INVESTIGATION OF FLATFISH MONOPHYLY[1]

Abstract

We present the first study to examine phylogenetic patterns across a broad sample of flatfish mitochondrial genomes. The flatfishes (Pleuronectiformes) have attracted attention in evolutionary biology since the early history of the field. Understanding the evolutionary history and patterns of diversification of the group will shed light on the evolution of novel body plans. Because recent molecular studies have yielded conflicting results, it is important to examine phylogenetic signal in different genomes and genome regions. We aligned and analyzed mitochondrial genome sequences from thirty-nine pleuronectiforms including nine newly reported here, and sixty-six non-pleuronectiforms (twenty additional clade L taxa [Carangimorpha or Carangimorpharia] and forty-six secondary outgroup taxa). The analyses yield strong support for clade L and weak support for the monophyly of Pleuronectiformes. The Pleuronectoidei receives moderate support, and as with other molecular studies the putatively basal lineage of Pleuronectiformes, the Psettodoidei is frequently not most closely related to other pleuronectiforms. Within the Pleuronectodei, the basal branching sequence in the group is poorly resolved, however several flatfish subclades receive stable and uncontradicted support. The affinities of *Lepidoblepharon* and *Citharoides* among pleuronectoids are particularly poorly resolved with these data.

[1]Campbell MA, López JA, Satoh TP, Chen W-J, and Miya M. Mitochondrial genomic investigation of flatfish monophyly. Submitted to Gene. In revision.

Introduction

Flatfishes (Pleuronectiformes) are a distinctive group of vertebrates characterized by bilateral asymmetry (Chapleau, 1993; Frazzetta, 2012). The remarkable body plan of flatfishes fed debate questioning the adequacy of natural selection as a theory of anatomical diversification and much speculation on the speed of such a change, in part due to the lack of extant intermediates (Janvier, 2008; Mivart, 1871). Only recently have intermediate flatfish forms been recognized in the fossil record (Friedman, 2012, 2008).

Complicating the topic of flatfish origins, support for monopyly for Pleuronectiformes is not universal. Evidence for flatfish paraphyly was offered in several studies (Amaoka, 1969; Chabanaud, 1949; Norman, 1934) predating a cladistic synthesis that concluded in support of the monophyly of the group (Chapleau, 1993). In this light, results of molecular-based studies that offer evidence for flatfish paraphyly are intriguing (Betancur-R. et al., 2013a, 2013b; Campbell et al., 2013a; Chen et al., 2003; Dettai and Lecointre, 2005; Li et al., 2009; Near et al., 2013, 2012; Smith and Wheeler, 2006). When the evidence does support monophyly of the flatfishes, the result is often sensitive to the particular combination of analyses and datasets examined (Betancur-R. et al., 2013b; Campbell et al., 2014). The debate surrounding what DNA sequences say about monophyly of flatfishes continues (Betancur-R. and Ortí, 2014; Campbell et al., 2014). While GC-bias can be shown to play a role in disrupting pleuronectiform monophyly when particular taxa are examined, that effect cannot explain the consistent placement of the genus *Psettodes* (spiny turbots) outside a restricted pleuronectiform clade (Campbell et al., 2013a). The placement of *Psettodes* apart from other pleuronectiforms may be the product of incomplete lineage sorting and/or the inability to correctly infer gene trees in nuclear datasets focusing on pleuronectiform monophyly (Campbell et al., 2014).

The three known species of *Psettodes* form the pleuronectiform suborder Psettodoidei. All other species of flatfishes (>700) are assigned to the suborder Pleuronectoidei in approximately 14 families and 34 genera (Munroe, 2005; Nelson, 2006). Three putative pleuronectiform synapomorphies (Chapleau, 1993) are not shared by *Psettodes* (Chabanaud, 1937; Nelson, 2006). The only morphological character uniting Pleuronectiformes appears to be correlates of bilateral asymmetry, which takes a distinct form in *Psettodes* (Friedman, 2008). To date, phylogenetic studies show that the monophyly of pleuronectoids is well supported (Campbell et al., 2013a) and that the phylogenetic affinities of all flatfishes (Psettodoidei and Pleuronectoidei) are with the Carangimorpha or clade L *sensu* Chen et al. (2003). Molecular evidence highlighted a close relationship between carangids and pleuronectids first with whole mitochondrial genome (mitogenome) data (Miya et al., 2003). This placement is well established and consistently supported (Betancur-R. et al., 2013a; Chen et al., 2007, 2003; Little et al., 2010; Miya et al., 2003; Near et al., 2012; Smith and Craig, 2007; Smith and Wheeler, 2006; Wainwright et al., 2012). Clade L contains an array of perciform taxa with diverse morphologies such as Toxotidae (archerfishes), Carangidae (jacks), Centropomidae+Latidae (snooks, Nile perches and allies), Xiphiidae (swordfish), Istophoridae (billfishes), Polynemidae (threadfins), Echeneidae (remoras), Coryphaenidae (dolphinfishes), Rachycentridae (cobia), Sphyraenidae (barracudas), Menidae (moonfish), and *Lactarius* (false trevally).

Flatfish, then, are in a curious position. Clade L is consistently found with high indices of support in molecular studies, although it contains a diverse array of morphological forms. In contrast, a monophyletic Pleuronectiformes receives only weak and inconsistent support in some concatenated phylogenetic analyses (Betancur-R. et al., 2013b). Futhermore, only only one gene trees to species tree analysis of many (Betancur-R et al., 2013b, Betancur-R. and Ortí, 2014) has

demonstrated pleuronectiform monophyly despite the striking bilateral asymmetry characteristic of the group. In addition, evaluation of different species trees from gene tree frameworks, datasets without missing data, accommodating for divergent base composition, and different configurations of concatenated analyses of nuclear gene data yield paraphyletic arrangements of the two main pleuronectiform lineages (Betancur-R. and Ortí, 2014; Betancur-R. et al., 2013a, 2013b; Campbell et al., 2014, 2013a).

Here we report results of an extensive examination of phylogenetic signal in mitochondrial genomes to infer pleuronectiform inter- and intra-relationships. Mitogenomes have a long history of use in fish molecular phylogenetics and have proven effective in resolving many areas of the fish tree of life (e.g. Campbell et al., 2013b; Doosey et al., 2009; Inoue et al., 2003, 2001; Miya and Nishida, 2000; Saitoh et al., 2003) while offering a number of practical advantages for phylogenetic inference (e.g. extremely conserved organization and uniparental/haploid inheritance). Because mitochondrial sequences show faster rates of substitution and smaller effective population size when compared to nuclear genomes, they have the potential to retain phylogenetic signal for diversification events that nuclear sequences may not (Charlesworth, 2009; Felsenstein, 2004). Our central goal is to establish to what extent patterns of mitogenomic variability among living flatfishes and their close relatives are congruent or in contradiction with expectations derived from flatfish monophyly.

Materials and Methods

Mitogenomes from twenty non-pleuronectiform clade L taxa representing maximal diversity of sampled lineages (Miya et al., 2013) were obtained from GenBank (Table 4.1A and

Table 4.1B). An additional forty-six candidate outgroups following Campbell et al. (2013a) were obtained from available mitogenome sequences (Table 1C). Among pleuronectiforms, we included all mitogenomic sequences available in GenBank removing a duplicate mitogenome sequences. We then targeted maximal divergences in unrepresented lineages to increase the accuracy of phylogenetic inference (Adkins and Honeycutt, 1994; Hillis, 1998; Hillis et al., 2003; Pollock et al., 2002). Mitogenome sequencing was conducted through long PCR then Sanger sequencing of short amplicons (Miya and Nishida, 1999). Multiple sequence alignments (MSA) were made for the protein-coding genes excluding ND6 due to compositional heterogeneity. First, amino acid sequences were aligned with MUSCLE version 3.8.31 (Edgar, 2004a, 2004b) and the corresponding DNA sequences aligned following the amino acid alignment. Ribosomal RNA (rRNA) sequences were aligned to an existing alignment (Miya et al., 2013) and a new and transfer RNA (tRNA) alignment was made with MUSCLE version 3.8.31 and regions of uncertain positional homology in alignments were excluded from subsequent analyses. We then conducted a maximum likelihood (ML) phylogenetic analyses using RAxML version 8.0.0 under GTR+Γ model of nucleotide evolution (Stamatakis and Ott, 2008) using twenty-three different configurations. These alternative configurations differ in sequence region inclusion/exclusion, coding of purines and pyrimidines at third codon positions ($1_N2_N3_{RY}$) to improve phylogenetic performance in the case of saturation and compositional bias (Phillips et al., 2004; Phillips and Penny, 2003), exclusion of third codon positions (1_N2_N) and partitioning scheme. The full dataset was partitioned by codon positions for each gene with third codons included, recoded, or removed, rRNA (R), and tRNA (T) partitions ($1_N2_N3_N$RT, $1_N2_N3_{RY}$RT, and 1_N2_NRT). In addition, we used partition schemes identified with PartitionFinder (Lanfear et al., 2012) on eight alternative data schemes and conducted ML phylogenetic analyses

on the un-partitioned dataset. Support from each component of the dataset was investigated separately such as protein coding genes by codon positions only, rRNA only, rRNA+tRNA, etc.

Results

A total of nine new mitogenomes from flatfishes were determined for this study and accessioned in the DDBJ/GenBank/EMBD under accessions AP014586-AP014594. Details of gene composition and organization, and molecular evolution of these newly available mitogenomes will be presented elsewhere.

Our alignment consists of 105 total taxa. Each codon position contained 3,636 sites. Our total alignment of unrecoded data ($1_N2_N3_N$RT) contains 13,742 sites with 9,091 distinct alignment patterns. The proportion of missing data was 0.21%. Partitioned ML analyses of the complete dataset partitioned by: codon positions for protein coding genes (with and without recoding of third codons), ribosomal RNAs, and transfer RNAs ($1_N2_N3_N$RT and $1_N2_N3_{RY}$RT) yield a monophyletic Pleuronectiformes (Figure 4.1; Table 4.1) with low support (bootstrap values of 8 and 22, respectively), monophyletic Pleuronectoidei with low support (20 and 46, respectively) and a monophyletic clade L with high support (100 in both cases). Exclusion of third codon positions (1_N2_NRT) did not result in a monophyletic Pleuronectiformes or Pleuronectoidei, but had high support for clade L (100).

Pleuronectiform monophyly is evident in only eight of the twenty-three analysis configurations (Table 2A and Table 2B) with all those cases showing invariably low support for monophyly of the group (bootstrap support < 23, average of 12.00). Support for Pleuronectoidei is common, found in eighteen of twenty-three analyses, but weak (bootstrap support < 46).

Support for clade L is found in twenty-two of the twenty-three analyses, and bootstrap support for clade L is frequently greater than 97. A monophyletic clade L was not found only with a single partition anlalysis of tRNA.

Considering only the full dataset partitioned by codon position (the $1_N2_N3_N$RT, $1_N2_N3_{RY}$RT, and 1_N2_NRT datasets) and relationships within Pleuronectiformes, we find evidence of Paralichthyidae comprising two distinct lineages. Otherwise family level divisions within Pleuronectodei were monophyletic. Strong support from the full datasets ($1_N2_N3_N$RT, $1_N2_N3_{RY}$RT, and 1_N2_NRT) indicates the genus *Paraplagusia* is nested with *Cynoglossus*. Pleuronectoidei in our analyses is comprised of several stable groupings which are uncertain in affinity at higher levels. Pleuronectidae is highly supported and most closely related to Paralichthyidae (*Paralichthys* + *Pseudorhombus*). Bothidae is highly supported as well as its relationship to Paralichthyidae (*Cyclopsetta*). We find Scopthalmidae, Achiropsettidae, and Rhombosoleidae to form a grouping as well. Cynoglossidae and Soleidae have high support to be most closely related to each other. In results that include pleuronectiform monophyly, the *Psettodes*-pleuronectoid divergence is the most basal among flatfish inferred diversification events.

Discussion

Our analyses yielded weak and inconsistent evidence for pleuronectiform monophyly. Alternative alignments of tRNA and rRNA sites had noticeable influence on inferred pleuronectiform relationships, which we do not include in this study. Interestingly, even pleuronectoid monophyly was not consistently or highly supported by bootstrap values in our analyses. In contrast, studies of pleuronectiform monophyly using multi-locus nuclear data there

is strong support for the Pleuronectodei (Betancur-R. et al., 2013b; Campbell et al., 2013a). The discrepancy may be evidence of the different ability of nuclear and mitochondrial DNA sequences to preserve information from internode segments of different relative duration.

Partitioning appeared to have a strong effect on potential outcomes. If we assume the monophyly of flatfishes as a starting point then a pattern of under-, appropriate, and over-parameterization emerges in results from alternative analysis configurations (Table 4.2A and Table 4.2B). However, the true relationships are rarely known in phylogenetic studies and we cannot use these results as a true evaluation of PartitionFinder's performance. Analyses of two of the six datasets consisting of only protein coding genes ($1_N2_N3_N$, $1_N2_N3_{RY}$, and 1_N2_N) or protein coding genes and RNA ($1_N2_N3_N$RT, $1_N2_N3_{RY}$RT, and 1_N2_NRT) produce evidence of pleuronectiform and pleuronectoid monophyly when unpartitioned. These are the 1_N2_N and 1_N2_NRT configurations with pleuronectiform bootstrap values of 16 and 7 respectively. Increasing parameterization by considering that each codon position, rRNA, and tRNA sites should be modeled with separate parameters results in more frequent recovery of the monophyletic Pleuronectiformes and Pleuronectoidei (found in results from five of these six datasets). However, the bootstrap support for pleuronectiform monophyly from 1_N2_N declined from 16 to 13 with two partitions, and 1_N2_NRT under four partitions does not support pleuronectiform monophyly. The results suggest that optimal partitioning for 1_N2_N and 1_N2_NRT datasets is a single partition. Increased parameterization was produced by PartitionFinder in datasets including third codon positions from protein coding genes. PartitionFinder always increased the total number of partitions over the subjective partitioning schemes, with poor success at recovering pleuronectiform monophyly (one instance, eighteen partitions, bootstrap support of 8). For example, coding scheme $1_N2_N3_{RY}$RT, which produces the best support values

for pleuronectiform monophyly does not produce a monophyletic pleuronectiformes when unpartitioned, does under five partitions and does not under seven and fifteen partitions. Continuing with assumed pleuronectiform monophyly as outcome indicative of performance, PartitionFinder appears to over-parameterize this dataset through the introduction of many partitions and does not improve the results of phylogenetic inference.

There is no strong evidence for or against pleuronectiform monophyly with existing nuclear sequence data (Campbell et al., 2014), and our results here arrive at the same conclusion. Only few nuclear gene sequences yield a monophyletic Pleuronectiformes when evaluated separately (Betancur-R. et al., 2013b; Campbell et al., 2014). As indicated by Campbell et al. (2014), an inability to correctly infer gene trees and/or a high degree of incomplete lineage sorting present in the clade L fishes is likely affecting these phylogenetic inferences. A benefit of mitogenomes is that each data partition should support the same underlying tree (i.e., there is a single gene) boosting the number of characters that can be soundly included in a concatenated analysis. Mitogenomes are generally non-recombining and uniparentally inherited. Furthermore, the effective population size of mitochondrial genomes is much smaller (1/4) than that of nuclear gene data, and mitogenomic data should not be affected by incomplete lineage sorting to the degree that nuclear genomic data are. The results we present do indicate that there is very little signal in mitochondrial genome data supporting pleuronectiform monophyly or the affinity of Psettodoidei to some other clade L lineage. A tree of clade L taxa with short internode distance as a result of the rapid radiation of the group would generate a low amount of phylogenetic signal with a high degree of homoplasy (or noise), and consequently inconsistent and weakly supported results.

Conclusions

Mitogenomic evidence does not provide strong evidence for flatfish monophyly, nor does it support an alternative placement for *Psettodes*. The highest support for Pleuronectiformes and Pleuronectoidei is 22 and 46 (bootstrap support) generated in the same analysis, neither of which can be considered strong statistical support. It is intriguing that a group of fishes with such striking morphologies arguing in favor of its monophyly (i.e., bilateral asymmetry) should exhibit such low support for monophyly from molecular data. Additional study of molecular evolution of clade L fishes and alternative sources of evidence should be pursued to help resolve the question of flatfish origins. In particular, methodologies that are designed to accommodate for incomplete lineage sorting can use Pleuronectiformes as a model system to explore the effects of highly discordant phylogenetic signal among loci as these methods have not been effective so far (Betancur-R. et al., 2013b). As molecular datasets continue to increase in size, it is important to avoid relying solely on analyses of concatenated alignments, which are known to obscure the underlying variation in phylogenetic signal.

Acknowledgements

M. A. C. wishes to thank his graduate committee members N. Takebayashi and J. Rhodes for their helpful discussions in relevance to this manuscript. For providing tissue and access to vouchers we thank the Center for Molecular Biodiversity Research (NSMT). This study was supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, Numbers 19207007 and 22370035, Japan Society for the Promotion of

Science KAKENHI Grant Number 24770082 and the Ministry of Science and Technology, Taiwan, Grant Number 101-2611-M-002 -016 -MY3.

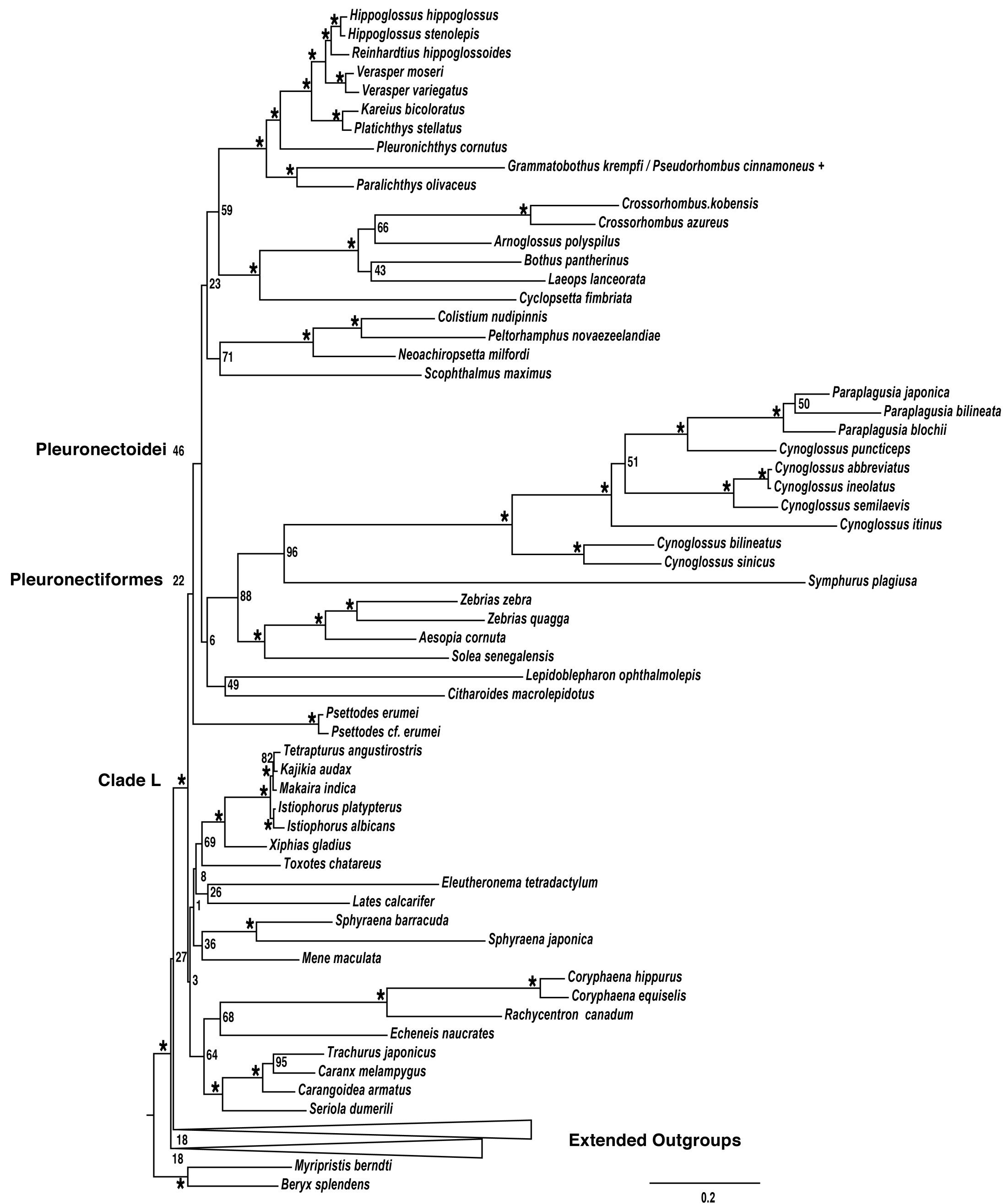


Figure 4.1: A maximum-likelihood (ML) tree generated in RAxML version 8.0.0 under a GTR+Γ model of nucleotide evolution. Mitogenomes were partitioned by codon position with third codons recoded, rRNA, and tRNA ($1_N2_N3_{RY}$RT). Values at nodes indicate bootstrap support values, and asterisk (*) indicates a value of 100. +Sequences for *Grammatobothus krempfi* and *Pseudorhombus cinnamoneus* retrieved from GenBank were identical and only one copy was included in this study.

Table 4.1A: Pleuronectiform mitogenomes included in this study. Family, species name, and sequence number are included. And asterisk (*) denotes mitogenomes generated for this study. Sequences for *Grammatobothus krempfi* and *Pseudorhombus cinnamoneus* retrieved from GenBank were identical and only one copy was included in this study.

Family	Species	Accession or Reference Number
Psettodidae	*Psettodes erumei*	FJ606835
Psettodidae	*Psettodes sp. (cf.erumei)*	AP014594*
Achiropsettidae	*Neoachiropsetta milfordi*	AP014593*
Bothidae	*Arnoglossus polyspilus*	AP014586*
Bothidae	*Bothus pantherinus*	AP014587*
Bothidae	*Crossorhombus azureus*	JQ639068
Bothidae	*Crossorhombus kobensis*	AP014589*
Bothidae	*Grammatobothus krempfi*	NC_022447.1
Bothidae	*Laeops lanceorata*	AP014591*
Citharidae	*Lepidoblepharon ophthalmolepis*	AP014592*
Citharidae	*Citharoides macrolepidotus*	AP014588*
Cynoglossidae	*Cynoglossus abbreviatus*	GQ380410
Cynoglossidae	*Cynoglossus bilineatus*	JQ349000
Cynoglossidae	*Cynoglossus itinus*	JQ639062
Cynoglossidae	*Cynoglossus lineolatus*	JQ349004
Cynoglossidae	*Cynoglossus puncticeps*	JQ349003
Cynoglossidae	*Cynoglossus semilaevis*	EU366230
Cynoglossidae	*Cynoglossus sinicus*	JQ348998
Cynoglossidae	*Paraplagusia bilineata*	JQ349001
Cynoglossidae	*Paraplagusia blochii*	JQ349002
Cynoglossidae	*Paraplagusia japonica*	JQ639066
Cynoglossidae	*Symphurus plagiusa*	JQ639061
Paralichthyidae	*Cyclopsetta fimbriata*	AP014590*
Paralichthyidae	*Paralichthys olivaceus*	AB028664
Paralichthyidae	*Pseudorhombus cinnamoneus*	JQ639069
Pleuronectidae	*Hippoglossus hippoglossus*	AM749122
Pleuronectidae	*Hippoglossus stenolepis*	AM749126
Pleuronectidae	*Kareius bicoloratus*	AP002951
Pleuronectidae	*Platichthys stellatus*	EF424428
Pleuronectidae	*Pleuronichthys cornutus*	JQ639071
Pleuronectidae	*Reinhardtius hippoglossoides*	AM749130
Pleuronectidae	*Verasper moseri*	EF025506
Pleuronectidae	*Verasper variegatus*	DQ403797
Rhombosoleidae	*Colistium nudipinnis*	JQ639063
Rhombosoleidae	*Peltorhamphus novaezeelandiae*	JQ639065
Scophthalmidae	*Scophthalmus maximus*	EU419747
Soleidae	*Aesopia cornuta*	KF000065
Soleidae	*Solea senegalensis*	AB270760
Soleidae	*Zebrias quagga*	JQ348999
Soleidae	*Zebrias zebra*	JQ700100

Table 4.1B: Non-pleuronectiform clade L mitogenomes included in this study. Family, species name, and sequence number are included.

Family	Species	Accession or Reference Number
Centropomidae	*Lates calcarifer*	DQ010541
Toxotidae	*Toxotes chatareus*	AP006806
Coryphaenidae	*Coryphaena hippurus*	AB355908
Coryphaenidae	*Coryphaena equiselis*	AB355907
Rachycentridae	*Rachycentron canadum*	FJ154956
Echeneidae	*Echeneis neucratoides*	AB355905
Carangidae	*Carangoides armatus*	AP004444
Carangidae	*Caranx melampygus*	AP004445
Carangidae	*Trachurus japonicus*	AP003091
Carangidae	*Seriola dumerili*	AB517558
Meneidae	*Mene maculata*	AB355909
Polynemidae	*Eleuthronema tetradactylum*	KC878730
Sphyraenidae	*Sphyraena barracuda*	AP006828
Sphyraenidae	*Sphyraena japonica*	AP012501
Xiphiidae	*Xiphias gladius*	AB470301
Istiophoridae	*Istiophorus albicans*	AP006035
Istiophoridae	*Istiophorus platypterus*	AB470306
Istiophoridae	*Makaira indica*	AB470305
Istiophoridae	*Tetrapturus angustirostris*	AB470303
Istiophoridae	*Kajikia audax*	AB470302

Table 4.1C: Non-clade L mitogenomes included in this study. Family, species name, and sequence number are included.

Family	Species	Accession or Reference Number
Acanthuridae	*Zebrasoma flavescens*	AP006032
Acropomatidae	*Doederleinia berycoides*	AP009181
Adrianichthyidae	*Oryzias latipes*	AP004421
Ammodytidae	*Ammodytes hexapterus*	KC422441
Balistidae	*Balistes vetula*	AP009204
Berycidae	*Beryx splendens*	AP002939
Caproidae	*Antigonia capros*	AP002943
Centrachidae	*Lepomis macrochirus*	JN389795
Centrachidae	*Micropterus dolomieu*	AB378749
Channidae	*Channa maculata*	JX978724
Cichlidae	*Astronotus ocellatus*	AP009127
Dactylopteridae	*Dactyloptena peterseni*	AP002947
Diodontidae	*Diodon holocanthus*	AP009177
Elassomatidae	*Elassoma evergladei*	AP002950
Gasterosteidae	*Gasterosteus aculeatus*	AP002944
Holocentridae	*Myripristis berndti*	AP002940
Kuhliidae	*Kuhlia mugil*	AP011065
Kyphosidae	*Girella punctata*	AP011060
Labridae	*Pseudolabrus sieboldi*	AP006019
Lateolabracidae	*Lateolabrax japonicus*	JQ860109
Lophiidae	*Lophius americanus*	AP004414
Luvaridae	*Luvarus imperialis*	AP009161
Mastecembelidae	*Mastacemblus favus*	AP002946
Melanotaeniidae	*Melanotaenia lacustris*	AP004419
Moronidae	*Morone saxatilis*	HM447585
Mugilidae	*Liza affinis*	JF911709
Nototheniidae	*Dissostichus eleginoides*	AB723627
Percichthyidae	*Nannoperca australis*	JF519732
Percidae	*Etheostoma radiosum*	AY341348
Percidae	*Perca flavescens*	JX629442
Scaridae	*Scarus fosteni*	FJ619271
Scombridae	*Scomberomorus semifasciatus*	JX559745
Sebastidae	*Sebastes marmoratus*	NC_013812
Serranidae	*Cephalophis argus*	KC593377
Serranidae	*Hypoplectrus gemma*	FJ848375
Sinipercidae	*Siniperca chuatsi*	JF972568
Sinipercidae	*Siniperca knerii*	JN378751
Sinipercidae	*Siniperca obscura*	KC567664
Sinipercidae	*Siniperca sherzeri "China: Poyang Lake"*	JQ010985
Sparidae	*Pagrus major*	AP002949.
Synbranchidae	*Monopterus albus*	AP002945
Terapontidae	*Rhynchopelates oxyrhynchus*	AP011064
Tetraodontidae	*Takifugu rubripres*	AJ421455
Tetraodontidae	*Tetraodon nigroviridis*	DQ019313
Triacanthodidae	*Triacanthodes anomalus*	AP009172
Zoarcidae	*Lycodes toyamensis*	AP004448

Table 4.2A: Summary of fifteen of the twenty-three analyses conducted for this study. All analyses were conducted on the same alignment under a GTR+Γ model of evolution in RAxML version 8.0.0. Data included, purine/pyrimidine recoding of third codon positions, and partitioning scheme varied between analyses. Partition schemes were single or subjective- based on biological knowledge. If the Pleuronectiformes, Pleuronectoidei, or clade L/Carangimorpha were found to be monophyletic, the associated bootstrap support is reported.version 8.0.0. Data included, purine/pyrimidine recoding of third codon positions, and partitioning scheme varied between analyses. Partition schemes were single or subjective- based on biological knowledge. If the Pleuronectiformes, Pleuronectoidei, or clade L/Carangimorpha were found to be monophyletic, the associated bootstrap support is reported.

				Pleuronectiformes		Pleuronectoidei		Clade L/Carangimorpha	
Included Data	Coding	Included Partitions	Partition Scheme	Monophyletic	Support	Monophyletic	Support	Monophyletic	Support
Protein Coding Genes	$1_N2_N3_N$	First, second, third codon positions	Single	No	-	Yes	10	Yes	98
Protein Coding Genes	$1_N2_N3_{RY}$	First, second, third codon positions	Single	No	-	Yes	13	Yes	99
Protein Coding Genes	1_N2_N	First and second codon positions	Single	Yes	16	Yes	32	Yes	99
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_N$	First, second, third codon positions, rRNA, tRNA	Single	No	-	Yes	12	Yes	99
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_{RY}$	First, second, third codon positions, rRNA, tRNA	Single	No	-	Yes	27	Yes	100
Protein Coding Genes, 12S, 16S, tRNA	1_N2_N	First and second codon positions, rRNA, tRNA	Single	Yes	7	Yes	25	Yes	100
12S, 16S	-	-	Single	No	-	No	-	Yes	66
tRNA	-	-	Single	No	-	No	-	No	-
			Partition Scheme						
Protein Coding Genes	$1_N2_N3_N$	First, second, third codon positions	$1_N2_N3_N$	Yes	4	Yes	17	Yes	99
Protein Coding Genes	$1_N2_N3_{RY}$	First, second, third codon positions	$1_N2_N3_{RY}$	Yes	18	Yes	45	Yes	100
Protein Coding Genes	1_N2_N	First and second codon positions	1_N2_N	Yes	13	Yes	28	Yes	98
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_N$	First, second, third codon positions, rRNA, tRNA	$1_N2_N3_N$ rRNA tRNA	Yes	8	Yes	20	Yes	100
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_{RY}$	First, second, third codon positions, rRNA, tRNA	$1_N2_N3_{RY}$ rRNA tRNA	Yes	22	Yes	46	Yes	100
Protein Coding Genes, 12S, 16S, tRNA	1_N2_N	First and second codon positions, rRNA, tRNA	1_N2_N rRNA tRNA	No	-	No	-	Yes	100
12S, 16S, tRNA	-	-	rRNAtRNA	No	-	No	-	Yes	17

Table 4.2B: Summary of eight of the twenty-three analyses conducted for this study. All analyses were conducted on the same alignment under a GTR+Γ model of evolution in RAxML version 8.0.0. Data included, purine/pyrimidine recoding of third codon positions, and partitioning scheme varied between analyses. Partition schemes were generated in Partition Finder. Partitions are described with codon positions appended if applicable (e.g. atp61, atp62, and atp63) If the Pleuronectiformes, Pleuronectoidei, or clade L/Carangimorpha were found to be monophyletic, the associated bootstrap support is reported.

				Pleuronectiformes		Pleuronectoidei		Clade L/Carangimorpha	
Included Data	**Coding**	**Included Partitions**	**Partition Finder Best Partition Scheme**	**Monophyletic**	**Support**	**Monophyletic**	**Support**	**Monophyletic**	**Support**
Protein Coding Genes	$1_N2_N3_N$	Protein Coding Genes	(atp6, atp8) (co1) (co2) (co3) (cytb) (nd1, nd4l, nd5) (nd2, nd3, nd4)	No	-	Yes	13	Yes	97
Protein Coding Genes	$1_N2_N3_{RY}$	Protein Coding Genes	(atp6, nd4) (atp8) (co1, co3) (co2, nd3, nd4l) (cytb, nd1) (nd2, nd5)	No	-	Yes	11	Yes	100
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_N$	Protein Coding Genes, 12S, 16S, tRNA	(atp6) (atp8, nd2, nd3, nd4) (co1) (co2) (co3) (cytb) (nd1, nd4l, nd5) (12S, 16S) (tRNA)	No	-	Yes	20	Yes	99
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_{RY}$	Protein Coding Genes, 12S, 16S, tRNA	(atp6, nd2, nd4, nd5) (atp8) (co1, co3) (co2, nd3, nd4l) (cytb, nd1) (12S, 16S) (tRNA)	No	-	Yes	21	Yes	100
Protein Coding Genes	$1_N2_N3_N$	Codon Positions for Protein Coding Genes	(atp61, nd11, nd41) (atp62, nd4l2, nd52) (atp63, atp83, co23, co33, nd4l3) (atp81, nd21, nd32, nd51) (atp82) (co11) (co12, co32) (co13) (co21, nd4l1) (co22) (co31, cytb1) (cytb2, nd12) (cytb3, nd53) (nd13, nd23, nd31, nd43) (nd22, nd33, nd42)	No	-	Yes	23	Yes	99
Protein Coding Genes	$1_N2_N3_{RY}$	Codon Positions for Protein Coding Genes	(atp61, nd11, nd41) (atp62, nd4l2, nd52) (atp63, atp83, cytb3, nd13, nd31, nd43, nd53) (atp81, nd21, nd32, nd51) (atp82) (co11) (co12, co32) (co13, co23, co33, nd23, nd4l3) (co21, nd4l1) (co22) (co31, cytb1) (cytb2, nd12) (nd22, nd33, nd42)	No	-	No	-	Yes	99
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_N$	Codon Positions for Protein Coding Genes, 12S, 16S, tRNA	(atp61, nd11, nd41) (atp62, nd4l2, nd52) (atp63, cytb3) (atp81, nd21, nd32, nd51) (atp82) (atp83, nd4l3) (co11) (co12, co32) (co13) (co21, co31, cytb1, nd4l1) (co22) (co23, co33) (cytb2, nd12) (nd13, nd31, nd43, nd53) (nd22, nd33, nd42) (nd23) (12S, 16S) (tRNA)	Yes	8	Yes	9	Yes	99
Protein Coding Genes, 12S, 16S, tRNA	$1_N2_N3_{RY}$	Codon Positions for Protein Coding Genes, 12S, 16S, tRNA	(atp61, atp81, nd11, nd41) (atp62, nd4l2, nd52) (atp63, atp83, co23, co33, cytb3, nd13, nd23, nd31, nd43, nd4l3, nd53) (atp82) (co11) (co12, co32) (co13) (co21, nd4l1) (co22) (co31, cytb1) (cytb2, nd12) (nd21, nd32, nd51) (nd22, nd33, nd42) (12S, 16S) (tRNA)	No	-	Yes	21	Yes	100

References

Adkins, R.M., Honeycutt, R.L., 1994. Evolution of the primate cytochrome c oxidase subunit II gene. J. Mol. Evol. 38, 215–231.

Amaoka, K., 1969. Studies on the sinistral flounder found in the waters around Japan. Taxonomy, anatomy, and phylogeny. J. Shimonoseki Univ. Fish. 18, 65–340.

Betancur-R., R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Cureton II, J.C., Zhang, F., Buser, T., Campbell, M.A., Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C., Hough, D.J., Lu, G., Grande, T., Arratia, G., Ortí, G., 2013a. The tree of life and a new classification of bony fishes. PLoS Curr. doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288

Betancur-R., R., Li, C., Munroe, T.A., Ballesteros, J.A., Ortí, G., 2013b. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol. 62, 763–785. doi:10.1093/sysbio/syt039

Betancur-R., R., Ortí, G., 2014. Molecular evidence for the monophyly of flatfishes (Carangimorpharia: Pleuronectiformes). Mol. Phylogenet. Evol. 73, 18–22. doi:10.1016/j.ympev.2014.01.006

Campbell, M.A., Chen, W.-J., López, J.A., 2013a. Are flatfishes (Pleuronectiformes) monophyletic? Mol. Phylogenet. Evol. 69, 664–673. doi:10.1016/j.ympev.2013.07.011

Campbell, M.A., Chen, W.-J., López, J.A., 2014. Molecular data do not provide unambiguous support for the monophyly of flatfishes (Pleuronectiformes): A reply to Betancur-R and Ortí. Mol. Phylogenet. Evol. 75, 149–153. doi:10.1016/j.ympev.2014.02.011

Campbell, M.A., López, J.A., Sado, T., Miya, M., 2013b. Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. Gene 530, 57–65. doi:10.1016/j.gene.2013.07.068

Chabanaud, P., 1937. Les téléostéens dyssymétriques du Mokkatam Inférieur de Tourah. Mém. Inst. Egypte 31, 1–122.

Chabanaud, P., 1949. Le problème de la phylogénèse des Heterosomata. Bull. Inst. Ocean. 950, 1–24.

Chapleau, F., 1993. Pleuronectiform relationships: A cladistic reassessment. B. Mar. Sci. 52, 516–540.

Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10, 195–205. doi:10.1038/nrg2526

Chen, W.-J., 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. Phylogenet. Evol. 26, 262–288. doi:10.1016/S1055-7903(02)00371-8

Chen, W.-J., Ruiz-Carus, R., Ortí, G., 2007. Relationships among four genera of mojarras (Teleostei: Perciformes: Gerreidae) from the western Atlantic and their tentative placement among percomorph fishes. J. Fish Biol. 70, 202–218. doi:10.1111/j.1095-8649.2007.01395.x

Dettai, A., Lecointre, G., 2005. Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. C. R. Biol. 328, 674–689. doi:10.1016/j.crvi.2005.04.002

Doosey, M.H., Simons, A.M., Tang, K.L., Wood, R.M., Agnew, M.K., Yang, L., Hirt, M.V., Clements, M.D., Saitoh, K., Sado, T., Miya, M., Nishida, M., Mayden, R.L., Chen, W.-J., Bart, H.L., 2009. Reconstructing the phylogenetic relationships of the earth's most diverse clade of freshwater fishes--order Cypriniformes (Actinopterygii: Ostariophysi): A case study using multiple nuclear loci and the mitochondrial genome. Mol. Phylogenet. Evol. 51, 500–514. doi:doi: DOI: 10.1016/j.ympev.2008.12.015

Edgar, R.C., 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Edgar, R.C., 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792 –1797. doi:10.1093/nar/gkh340

Felsenstein, J., 2004. Inferrring Phylogenies. Sinauer Associates, Inc., Sunderland.

Frazzetta, T., 2012. Flatfishes, turtles, and bolyerine snakes: Evolution by small steps or large, or both? Evol. Biol. 39, 30–60. doi:10.1007/s11692-011-9142-7

Friedman, M., 2008. The evolutionary origin of flatfish asymmetry. Nature 454, 209–212. doi:10.1038/nature07108

Friedman, M., 2012. Osteology of †*Heteronectes chaneti* (Acanthomorpha, Pleuronectiformes), an Eocene stem flatfish, with a discussion of flatfish sister-group relationships. J. Vertebr. Paleontol. 32, 735–756. doi:10.1080/02724634.2012.661352

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52, 124–126.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2001. A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences. Mol. Phylogenet. Evol. 20, 275–285. doi:doi: DOI: 10.1006/mpev.2001.0970

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol. Phylogenet. Evol. 26, 110–120. doi:doi: DOI: 10.1016/S1055-7903(02)00331-7

Janvier, P., 2008. Palaeontology: Squint of the fossil flatfish. Nature 454, 169–170. doi:10.1038/454169a

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. doi:10.1093/molbev/mss020

Li, B., Dettaï, A., Cruaud, C., Couloux, A., Desoutter-Meniger, M., Lecointre, G., 2009. RNF213, a new nuclear marker for acanthomorph phylogeny. Mol. Phylogenet. Evol. 50, 345–363. doi:10.1016/j.ympev.2008.11.013

Little, A.G., Lougheed, S.C., Moyes, C.D., 2010. Evolutionary affinity of billfishes (Xiphiidae and Istiophoridae) and flatfishes (Plueronectiformes): Independent and trans-subordinal origins of endothermy in teleost fishes. Mol. Phylogenet. Evol. 56, 897–904. doi:10.1016/j.ympev.2010.04.022

Mivart, G.J., 1871. On the Genesis of Species. MacMillan., London.

Miya, M., Friedman, M., Satoh, T.P., Takeshima, H., Sado, T., Iwasaki, W., Yamanoue, Y., Nakatani, M., Mabuchi, K., Inoue, J.G., Poulsen, J.Y., Fukunaga, T., Sato, Y., Nishida, M., 2013. Evolutionary origin of the Scombridae (Tunas and Mackerels): Members of a Paleogene adaptive radiation with 14 other pelagic fish families. PLoS ONE 8, e73535. doi:10.1371/journal.pone.0073535

Miya, M., Nishida, M., 1999. Organization of the mitochondrial genome of a deep-sea fish, Gonostoma gracile (Teleostei: Stomiiformes): First example of transfer RNA gene rearrangements in bony fishes. Mar. Biotechnol. 1, 416–26.

Miya, M., Nishida, M., 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol. Phylogenet. Evol. 17, 437–455. doi:doi: DOI: 10.1006/mpev.2000.0839

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 26, 121 – 138.

Munroe, T.A., 2005. Systematic diversity of the Pleuronectiformes, in: Gibson, R.N. (Ed.), Flatfishes: Biology and Eploitation. Blackwell, Oxford, UK, pp. 10–41.

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbrink, F.T., Friedman, M., Wainwright, P.C., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci. doi:10.1073/pnas.1304661110

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., Smith, W.L., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. 109, 13698–13703. doi:10.1073/pnas.1206625109

Nelson, J.S., 2006. Fishes of the World, 4th ed. John Wiley & Sons, Inc, Hoboken, New Jersey.

Norman, J.R., 1934. A systematic monograph of the flatfishes (Hetersomata), Vol. 2. Psettodidae, Bothidae, Pleuronectidae.

Phillips M.J., Delsuc, F., Penny, D. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455-1458.

Phillips M.J., Penny, D. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogent. Evol. 28, 171-185.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51, 664–671.

Saitoh, K., Miya, M., Inoue, J.G., Ishiguro, N.B., Nishida, M., 2003. Mitochondrial genomics of ostariophysan fishes: Perspectives on phylogeny and biogeography. J. Mol. Evol. 56, 464–472. doi:10.1007/s00239-002-2417-y

Smith, W.L., Craig, M.T., 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, 35–55.

Smith, W.L., Wheeler, W.C., 2006. Venom evolution widespread in fishes: A phylogenetic road map for the bioprospecting of piscine venoms. J. Hered. 97, 206 –217. doi:10.1093/jhered/esj034

Stamatakis, A., Ott, M., 2008. Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. Philos. Trans. R. Soc. B Biol. Sci. 363, 3977–3984. doi:10.1098/rstb.2008.0163

Wainwright, P.C., Smith, W.L., Price, S.A., Tang, K.L., Sparks, J.S., Ferry, L.A., Kuhn, K.L., Eytan, R.I., Near, T.J., 2012. The Evolution of pharyngognathy: A phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Syst. Biol. 61, 1001–1027. doi:10.1093/sysbio/sys060

Chapter 5: Conclusion

In this dissertation, empirical datasets were evaluated with respect to three major challenges in phylogenetic inference: Chapter 2 - polyploidy, Chapter 3 - base composition bias, and Chapter 4 - incomplete lineage sorting.

In Chapter 2, the importance of homology assessment in phylogenies of polyploid groups is highlighted. The major conclusions of Chapter 2 is that (1) the subfamily relationships of Salmonidae were not found to be as previously accepted. Secondarily, (2) separate families within the pike order (Esociformes) were not supported. In particular, the conclusion (1) of a sister relationship between graylings (Thymallinae) and whitefishes (Coregoninae) may be a direct result of appropriate homology. The impact of homology in fish phylogenetics is sufficiently addressed and examples may be taken from the plant literature, where gene duplication events are a recognized evolutionary force (Duarte et al., 2010) and polyploidization is frequent (Wood et al., 2009). Likewise, fishes contain a wide range of ancient and recent whole genome duplications (Sato and Nishida, 2010). Studies of salmonid phylogeny more recently have not considered the paralogy of the loci included, such as Shedko et al. (2012) and Crête-Lafrenière et al. (2012), although the importance of paralogy when constructing phylogenies in this group was demonstrated some time ago (Oakley and Phillips, 1999). The conclusion of a sister Thymallinae and Coregoninae interestingly contradicts the leading morphological hypothesis for salmon interrelationships at the subfamily level of a sister relationship between Thymallinae and salmons, trouts, and charrs (Salmoninae) (Sanford, 1990; Wilson and Williams, 2010). Molecular studies have not been consistent in conclusions of salmonid interrelationships, however some papers support the conclusion reached in Chapter 2 (Burridge et al., 2012; Li et al., 2010; Macqueen and Johnston, 2014). The treatment of

MacQueen and Johnston (2014) of identifying paralogous and homologous gene sequences from sequenced RNA lends strong corroborative support to my results. The paralogous nature of nuclear data sets undoubtedly is a contributor the inconsistent results of molecular studies of salmonid relationships. The use of supposed single copy nuclear genes in large scale phylogeny is appropriate to place the Salmonidae among higher taxonomic levels. However, results from these studies such as Betancur-R. et al. (2013) and Near et al. (2013) with respect to the arrangements of Salmonidae are undoubtedly affected by paralogy and should not be considered valid hypotheses of salmonid relationships. Nuclear loci I sequenced (unpublished) shared with both Betancur et al. (2013) and Near et al. (2013) were obviously not single copy in nature. Both of these papers indicate alternative relationships to the results of Chapter 2. Specific work targeting salmonidae (Crête-Lafrenière et al., 2012), incorporating both mitochondrial data, few nuclear loci, and much missing data reaching an alternative conclusion to Chapter 2 is also affected by paralogy. While considering the placement of the Salmonidae, the use of paralogous loci is suitable. Any inferences of relationships within Salmonidae are incorrect as they are affected by paralogy; however peer review required me to include these hypotheses as valid in Chapter 2. Mitochondrial data is subject to potential sampling error as it is a single locus, but congruence between my mitochondrial data and MacQueen and Johnston (2014) which was published after Chapter 2, is a strong indicator that Thymallinae and Coregoninae are most closely related and the mitochondrial hypothesis presented here is correct. Due to the occurrence of ancient whole genome duplication events such as in the ancestor of all euteleosts (Santini et al., 2009) and lineages that are more recent polyploids such as the suckers (Catostomidae) (Chen and Mayden, 2012), phylogenetics of fishes will be improved with increased emphasis on orthology assessment.

Conclusion (2) of Chapter 2 of not finding support for the Umbridae and Esocidae within Esociformes is part of building evidence to this effect. All analyses undertaken in Chapter 2 support a topology of Esociformes of (*Umbra*, (*Dallia*, (*Novumbra, Esox*))). Umbridae of Wilson and Veilleux (1982) contains the genera *Umbra*, *Dallia*, and *Novumbra*. My results are consistent with the morphological hypothesis of Wilson and Williams (2010). The combined evidence clearly points towards a reclassification of Esociformes over the currently accepted taxonomy. However two possible options for Esosciformes+Salmoniformes can be employed. Firstly, a single family should be considered for Esociformes. Then, both Esociformes and Salmoniformes would contain a single family each (Esocidae and Salmonidae). It would be sensible to further reduce the amount of taxonomic categories to a single order for both families. The Salmoniformes *sensu* Greenwood et al. (1966) contained many forms (Salmonoidei, Plecoglossidae, Osmeridae, Argentinoidei, Galaxioidei, Esocoidei, Stomiatoidei, etc.), and a second option to combine Esocidae+Salmonidae into a single order would not be unheard of and would simplify our current taxonomy.

Chapter 3 found that with consideration for base composition bias, we found the flatfishes (Pleuronectiformes) to not be monophyletic. However, strong support for Pleuronectoidei and clade L was apparent. Clade L (Carangimorpha or Carangimorpharia) contains taxa of diverse morphologies such as barracuda, dolphinfish, archerfish, marlins, all extant flatfishes (*Psettodes* and Pleuronectoidei) and many others. Despite the morphological diversity of clade L it is consistently found across molecular analyses. Base composition bias was identified widely across flatfishes. In particular, the lineage of Bothidae was highly biased in base composition. The Psettodoidei does not appear to be compositionally biased, nor was the placement of Psettodoidei found to be unstable by the RogueNaRok (Aberer et al., 2013)

algorithm. I achieved largely congruent results from phylogenetic analyses allowing for greater flexibility in nucleotide evolution such as GTR-CAT (Lartillot et al., 2009) and the individual models implemented in p4 (Foster, 2004) as well as standard phylogenetic tree search methods. However, while the placement of *Psettodes* remained unaffected by the alternative models, intraordinal results for pleuronectiform taxa were affected in some circumstances. Those analyses incorporating third codon positions and not recoding them in this chapter differed in results between models implemented in standard phylogenetic frameworks and those implemented in p4. Notably, the families Cynoglossidae and Soleidae are very similar in DNA composition. In neutral models (stationarity and homogeneity assumed) the monophyly of these two families was not found, and the two families are mixed together in the phylogenetic results. Morphological evidence clearly distinguishes these two families, and recoding third codon positions or omitting them produces monophyletic Cynoglossidae and Soleidae in neutral phylogenetic analyses. By relaxing the assumptions of stationarity and homogeneity in p4, we achieved results by including third codon positions that were not found otherwise.

Third codon positions are the least constrained codon positions, and most varied in composition. Recoding is a common strategy to make use of these data. Recoding often is undertaken by making only two character states instead of four based on biochemical groupings. Purines (A and G) are recoded as R, and pyrimidines (C and T) as Y. Recoding in this fashion affects several aspects of inference, such as saturation (Adkins and Honeycutt, 1994) and compositional heterogeneity (Woese et al., 1991). But, recoding reduces the total amount of information available and biases outside third codon positions are present (Chen et al., 2014). Although recoding is used widely in phylogenetics with 4-state Markov models (*e.g*. Campbell et al., 2013; Crête-Lafrenière et al., 2012; Li and Ortí, 2007), it is not correct to use 4-state Markov

models to model recoded data although it is an accepted practice (Phillips and Penny, 2003; Vera-Ruiz et al., 2014). Success with including third codon positions in non-neutral phylogenetic analyses speaks volumes about the utility of non-neutral models in phylogenetics. Trends in phylogenetic analyses are away from implementing non-neutral models since datasets are growing so large. Very large datasets are difficult to resolve, and methods such as RAxML (Stamatakis 2006; Stamatakis and Ott 2008) which are fast, are inflexible in model choice. Many options are available to account for the potential of base composition bias in phylogenetics (*e.g.* Boussau and Gouy, 2006; Foster, 2004; Galtier and Gouy, 1998; Jayaswal et al., 2005), but these programs are infrequently used (Table 5.1).

It is evident that a large difference in popularity exists between computer programs, and it is true that a more complex model may not be needed in all cases. Certain barriers exist to the widespread use of alternative models. For example, the program p4 is much slower than standard model programs, partly as a result of increased model complexity but also for two other key reasons. Instead of being compiled, p4 is an interpreted program only allows non-neutral models to be explored in a Bayesian framework. It would be a great benefit if programs such as p4 were produced that implement tree searching and non-neutral models, but also required less computational time.

Support for Pleuronectoidei monophyly was strong, not only in terms of replication across analyses, but also in statistical support. Likewise for clade L. The placement for Psettodoidei, was however, inconsistent. The stastical support for the placement of Psettodoidei among analyses, was low. No clear conclusion can be made then on the monophyly or not of Pleuronectiformes in Chapter 3. But, it is clear base composition bias is an unlikely influence on the placement of the psettoid lineage, but incomplete lineage sorting and/or an inability to

correctly infer phylogeny is present (Campbell et al., 2014). A careful review of literature while composing Chapter 3 highlighted the fact that the cladistics synthesis of Chapleau (1993) was flawed. Two of the three putative synapomorphies for Pleuronectiformes are absent in Psettodoidei, in part since the *recessus orbitalis* has not been observed in *Psettodes*. I attempted to address the condition of the *recessus orbitalis* in *Psettodes* through a collaborator mailing a specimen of *Psettodes* to a specialist. Unfortunately, the condition the specimen did not permit the presence of the *recessus orbitalis* to be observed or not. Therefore, it remains for a morphological specialist to evaluate *Psettodes* for aspects of morphology which may tie it to other fishes or to the Pleuronectoidei. The flatfishes represent a case where morphologists and molecular phylogeneticists can work to advance knowledge together. The conclusions of Chapter 2 for a sister Esociformes and Salmoniformes relationship and a single family of esociforms (Esocidae) was contradictory to accepted morphological hypotheses when first advanced. Consistent molecular results led morphologists to re-evaluate evidence as evidenced by the morphological hypothesis of Wilson and Williams (2010) which is highly congruent with the hypothesis of Esociformes + Salmoniformes relationships advanced here.

Chapter 4 again focuses on the question of flatfish monophyly. As opposed to nuclear gene datasets, mitogenomes have three helpful properties in this chapter. Firstly, all parts of mitochondrial genomes share the same history, that is they are a single locus. Therefore, concatenated analyses are appropriate for different mitochondrial genome data partitions. The size of mitochondrial genomes, ~16.5 thousand base pairs, provides many characters for phylogenetic analysis. And, hopefully better parameter estimates. Secondly, the smaller effective population size of mitochondrial genomes causes lineage sorting to occur at a faster rate

compared to nuclear data sources. In additional to the first two benefits listed, mitogenomes are an independent source of data from nuclear genomes.

The major result of the twenty-three analyses conducted in Chapter 4 is that support for flatfish monophyly is weak statistically and inconsistently found in analyses. Support for the monophyly of Pleuronectoidei is not strongly supported statistically and the placement of Psettodoidei is inconsistent among analyses. Combined with the results from Chapter 3 and my other work (Campbell et al., 2014), the lack of resolution in clade L is result of a rapid radiation and consequent short-internode distance. Molecular data as it is now, does not conclusively support pleuronectiform monophyly or not.

Chapter 4 illustrated how data is modeled and concatenated has important effects while lack of recombination in mitochondria has been used to justify concatenation of mitochondrial data. In Chapter 4, by our choice of inference program, we were limited to two choices in model General Time Reversible (GTR) + rate variation (Γ) or GTR + Γ + a proportion of invariant sites (I). Following the suggestions of the program manual for RAxML, we used GTR + Γ and there is some potential for error due to model misspecification (Sullivan and Joyce, 2005). It was clear that partitioning had a large effect on results. Datasets in Chapter 4 inconsistently support flatfish monophyly across partitioning schemes. Partitioning is a strategy to appropriately capture the variation among aligned sites in DNA sequences. In concept, sites that have evolved under similar processes should be pooled into partitions and separate model parameters estimated (Nylander et al., 2004). Identifying partitions is problematic, and in Chapter 4 three approaches were made: single (no partitioning), subjective (based upon my biological intuition) and objective (Lanfear et al., 2012). Variability between data partitions and partition schemes in

mitochondria can be attributed to incorrect model choice, homoplasy, or some other difficulty in phylogenetic inference (artefactual) since all parts of the mitochondrial genome present the same history. In nuclear genomes where discordance between partitions is expected it may additionaly be the result of biological reasons (Galtier and Daubin, 2008). It is evident that there is much contradiction and ambiguity in molecular phylogenetics which is covered up in many phylogenetic analyses and not included in published papers. Variability across data partitions is lost in concatenation, and evidence indicates that concatenation leads to inflated support values even with conflict and systematic error present (Chen et al., 2003; Felsenstein, 1978; Hillis and Bull, 1993; Huelsenbeck, 1997; Salichos and Rokas, 2013). Across the analyses in Chapter 4, which theoretically should have the same result, the same result was not observed. I believe that we should highlight contradiction and ambiguity in phylogenetics and attempt to resolve and understand them instead of presenting only the "best" results, which match preconceptions - such as flatfish monophyly. Without a strong preconception of monophyly for Pleuronectiformes based on cranial asymmetry there would not be support for pleuronectiform monophyly based on the outcome of molecular studies, whose outcome appears to be highly influenced by incomplete lineage sorting, model choice, and partitioning strategies.

Overall this dissertation has shown how a careful methodological approach can result in conclusions that are contrary to widely accepted doctrine. Promising future work for phylogenetics is uncovered in this dissertation in the genomics age. Proper treatment of large datasets to find orthologous and paralogous sequences for analysis will be an advantage in polyploid lineages. Large datasets and increased computational ability should allow non-standard models of nucleotide evolution to be used more, not less. Datasets in the genomic age should not continued to be concantenated into ever larger matrices which obscures phylogenetic

heterogeneity. Addressing the distribution of phylogenetic signal across genomes will be much more informative and insightful.

Table 5.1: Selected references for phylogenetic inference methods and number of citations for each reference from Web of Science. Retrieved on 08/27/2013. *Indicates citation count retrieved from Google Scholar.

Standard Phylogenetic Approaches	Web of Science Citation Count
Ronquist F, Heulsenbeck, JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.	10,343
Stamataki A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.	3,052
Swofford DL. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates.	13,520*
Zwickl, DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation. University of Texas at Austin.	1,104
Non-Standard Phylogenetic Approaches	
Boussau B, Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. Systematic Biology 55:756-768.	33
Foster PG. 2004. Modeling compositional heterogeneity. Systematic Biology 53:485-495	113
Galtier N, Gouy M. 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Molecular Biology and Evolution 15:871-879	176
Jayaswal V, Jermiin LS, Poladian L, Robinson J. 2011. Two stationary nonhomogeneous models of nucleotide sequence evolution. Systematic Biology 60:74-86	5
Jayaswal V., Jermiin LS., Robinson J. 2005. Estimation of phylogeny using a General Markov model. Evol. Bioinforma. Online. 1:62–80.	24

References

Aberer, A.J., Krompass, D., Stamatakis, A., 2013. Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. Syst. Biol. 62, 162–166. doi:10.1093/sysbio/sys078

Adkins, R.M., Honeycutt, R.L., 1994. Evolution of the primate cytochrome c oxidase subunit II gene. J. Mol. Evol. 38, 215–231.

Betancur-R., R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Cureton II, J.C., Zhang, F., Buser, T., Campbell, M.A., Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C., Hough, D.J., Lu, G., Grande, T., Arratia, G., Ortí, G., 2013. The tree of life and a new classification of bony fishes. PLoS Currents. doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288

Boussau, B., Gouy, M., 2006. Efficient Likelihood Computations with Nonreversible Models of Evolution. Syst. Biol. 55, 756–768.

Burridge, C.P., McDowall, R.M., Craw, D., Wilson, M.V.H., Waters, J.M., 2012. Marine dispersal as a pre-requisite for Gondwanan vicariance among elements of the galaxiid fish fauna. J. Biogeogr. 39, 306–321. doi:10.1111/j.1365-2699.2011.02600.x

Campbell, M.A., Chen, W.-J., López, J.A., 2013. Are flatfishes (Pleuronectiformes) monophyletic? Mol. Phylogenet. Evol. 69, 664–673. doi:10.1016/j.ympev.2013.07.011

Campbell, M.A., Chen, W.-J., López, J.A., 2014. Molecular data do not provide unambiguous support for the monophyly of flatfishes (Pleuronectiformes): A reply to Betancur-R and Ortí. Mol. Phylogenet. Evol. 75, 149–153. doi:10.1016/j.ympev.2014.02.011

Chapleau, François. 1993. "Pleuronectiform Relationships: A Cladistic Reassessment." Bull. Mar. Sci. 52 (1): 516–40.

Chen, J.-N., López, J.A., Lavoué, S., Miya, M., Chen, W.-J., 2014. Phylogeny of the Elopomorpha (Teleostei): Evidence from six nuclear and mitochondrial markers. Mol. Phylogenet. Evol. 70, 152–161. doi:10.1016/j.ympev.2013.09.002

Chen, W.-J., Mayden, R.L., 2012. Phylogeny of suckers (Telostei: Cypriniformes: Catostomidae): further evidence of relationships provided by he single-copy nuclear gene IRBP2. Zootaxa 3586, 195–210.

Chen, W.-J., 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. Phylogenet. Evol. 26, 262–288. doi:10.1016/S1055-7903(02)00371-8

Crête-Lafrenière, A., Weir, L.K., Bernatchez, L., 2012. Framing the Salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. PLoS ONE 7, e46662. doi:10.1371/journal.pone.0046662

Duarte, J., Wall, P.K., Edger, P., Landherr, L., Ma, H., Pires, J.C., Leebens-Mack, J., dePamphilis, C., 2010. Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis and Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evol. Bio. 10, 61.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410. doi:10.2307/2412923

Foster, P.G., 2004. Modeling compositional heterogeneity. Syst. Biol. 53, 485–495.

Galtier, N., Daubin, V., 2008. Dealing with incongruence in phylogenomic analyses. Philos. T. Roy. Soc. B. 363, 4023–4029.

Galtier, N., Gouy, M., 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15, 871 –879.

Greenwood, P.H., Rosen, D.E., Weitzman, S.H., Myers, G.S., 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. B. Am. Mus. Nat. Hist. 131, 339–456.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192. doi:10.1093/sysbio/42.2.182

Huelsenbeck, J.P., 1997. Is the Felsenstein Zone a fly trap? Syst. Biol. 46, 69–74. doi:10.1093/sysbio/46.1.69

Jayaswal, V., Jermiin, L.S., Robinson, J., 2005. Estimation of phylogeny using a General Markov model. Evol. Bioinform. Online 1, 62–80.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. doi:10.1093/molbev/mss020

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286 –2288. doi:10.1093/bioinformatics/btp368

Li, C., Ortí, G., 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. Mol. Phylogenet. Evol. 44, 386–398. doi:doi: DOI: 10.1016/j.ympev.2006.10.030

Li, J., Xia, R., McDowall, R.M., López, J.A., Lei, G., Fu, C., 2010. Phylogenetic position of the enigmatic *Lepidogalaxias salamandroides* with comment on the orders of lower euteleostean fishes. Mol. Phylogenet. Evol. 57, 932–936. doi:10.1016/j.ympev.2010.07.016

Macqueen, D.J., Johnston, I.A., 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc. Royal Soc. B. 281. doi:10.1098/rspb.2013.2881

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbrink, F.T., Friedman, M., Wainwright, P.C., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci. doi:10.1073/pnas.1304661110

Nylander, J.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst Biol 53, 47 – 67.

Oakley, T.H., Phillips, R.B., 1999. Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. Molecular Phylogenetics and Evolution 11, 381–393. doi:doi: DOI: 10.1006/mpev.1998.0599

Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171–185. doi:10.1016/S1055-7903(03)00057-5

Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331.

Sanford, C. P. J. 1990. The phylogenetic relationships of salmonoid fishes." Bull. Br. Mus. (Nat. Hist.) Zool. 56: 145–53.

Santini, F., Harmon, L., Carnevale, G., Alfaro, M., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Bio. 9, 194.

Sato, Y., Nishida, M., 2010. Teleost fish with specific genome duplication as unique models of vertebrate evolution. Environ. Biol. Fish. 88, 169–188. doi:10.1007/s10641-010-9628-7

Shedko, S., Miroshnichenko, I., Nemkova, G., 2012. Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: Analysis of nuclear *RAG1* gene. Russ. J. Genet. 48, 575–579. doi:10.1134/S1022795412050201

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688 – 2690.

Stamatakis, A., Ott, M., 2008. Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. Philosophical Trans. Royal Soc. B. 363, 3977–3984. doi:10.1098/rstb.2008.0163s

Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36, 445–466. doi:10.1146/annurev.ecolsys.36.102003.152633

Vera-Ruiz, V., Lau, K., Robinson, J., Jermiin, L., 2014. Statistical tests to identify appropriate types of nucleotide sequence recoding in molecular phylogenetics. BMC Bioinformatics 15, S8.

Wilson, M.V.H., Veilleux, P., 1982. Comparative osteology and relationships of the Umbridae (Pisces: Salmoniformes). Zool. J. Linn. Soc-Lond. 76, 321–352.

Wilson, M.V.H., Williams, R.R.G., 2010. Salmoniform fishes: key fossils, supertree, and possible morphological synapomorphies., in: Nelson, J.S., Schultze, H.P., Wilson, M.V.H. (Eds.), Origin and Phylogenetic Interrelationships of Teleosts. Verlag Dr. Friedrich Pfeil, München, pp. 379–409.

Woese, C.R., Achenbach, L., Rouviere, P., Mandelco, L., 1991. Archaeal phylogeny: Reexamination of the phylogenetic position of *Archaeoglohus fulgidus* in light of certain composition-induced artifacts. Syst. Appl. Microbiol. 14, 364–371. doi:10.1016/S0723-2020(11)80311-5

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L.H., 2009. The frequency of polyploid speciation in vascular plants. P. Natl. Acad. Sci. USA. doi:10.1073/pnas.0811575106