Fall 2022

Phylogenetic Relationships Among Fishes in the Order Zeiformes Based on Molecular Data from Three Mitochondrial Loci

Lindsay Scarpitta

LOYOLA UNIVERSITY CHICAGO

PHYLOGENETIC RELATIONSHIPS AMONG FISHES IN THE ORDER ZEIFORMES BASED ON MOLECULAR DATA FROM THREE MITOCHONDRIAL LOCI

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN BIOLOGY

BY

LINDSAY A. SCARPITTA

CHICAGO, IL

MAY 2022

ACKNOWLEDGEMENTS

I would like to thank the many people who have helped me in the preparation of this degree. I especially would like to thank my adviser Dr. Terry Grande for welcoming me into her lab and for her invaluable support at every step of the project. The breadth of Dr. Grande's knowledge and expertise in her field has been inspiring since the very beginning of my time at Loyola. Her enthusiasm for fishes is infectious, and her encouragement and confidence in the strength of my work have buoyed me through the ups and downs of research and the curveballs life will throw at you. I am deeply grateful to have had the opportunity to build the relationship I have with Dr. Grande, without which my efforts to unravel the evolutionary history of these fascinating fishes would not be what it is today.

I am incredibly grateful to my co-advisor Dr. Mark Wilson for his far-reaching expertise and unflinching generosity in sharing that knowledge with me. His guiding hand was instrumental in everything from technical troubleshooting to asking just the right questions that would lead to my fitting tough evolutionary puzzle pieces together, and I cannot thank him enough for taking the time and care to steer me back on-course when I found myself struggling with the bigger picture.

Many thanks go to the members of my committee, Dr. Howard Laten and Dr. Marty Berg, for their valuable time, instruction, and attention to detail over the course of this project. I am eminently grateful to their patience and encouragement through the entire process, and I am privileged to have reaped the benefits of their involvement in my growth as a researcher.

I would also like to thank former Grande Lab post-doc Cal Borden for all his help in the initial stages of the project, through contacting research collections for tissue acquisition and for his aid in getting the molecular bench work off the ground.

I am grateful to the members of the Grande Lab for their camaraderie and friendship, as well as all of their generous help in troubleshooting software and instruments, studying for shared classes, and teaching me programs and techniques that contributed to the strength of my project. Many thanks to Meg Malone for welcoming me into the lab, to Cheryl Theile, John Belcik, Josh Hittie, and Jeff Peters for being wonderful labmates and classmates during my tenure working on-campus, and for their always-thoughtful questions and suggestions during lab meetings. Thanks also go to Michael Hanson for his beautiful line drawings of the fishes under study for this thesis.

Many thanks are also due to Audrey Berry, Virginia Lorenzo, and all the Loyola University Department of Biology staff for providing me with the funding to complete my graduate study, for their daily close attention and advocacy for Biology graduate students, and for guiding me through the administrative requirements of this program. Thanks go as well to Dr. Sushma Reddy for the use of her molecular biology laboratory space to carry out the bench work for this thesis.

Throughout my course of study, the support of my friends in the Biology Department, family, and loved ones has been a much-needed port in the storm when I felt daunted by the scope of what I hoped to accomplish in this program. I could not have done it without them. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Zeiformes (dories) are mid-water or deep (to 1000 m) marine acanthomorph fishes with a global, circumtropical, and circumtemperate distribution. Some species have a near-worldwide distribution, while others appear to be regional endemics, e.g., near New Zealand. Six families, 16 genera, and 33 species are currently recognized as valid. Relationships among them, however, remain unsettled, especially in light of recent proposals concerning the phylogenetic placement of zeiforms within the Paracanthopterygii rather than allied with beryciforms or percomorphs. The present study uses mtDNA characters to investigate zeiform interrelationships given their revised phylogenetic placement and attendant changes to their close outgroups, carried out as part of a larger study by Grande et al. (2018) also including nDNA + morphological characters in their assessment of zeiform phylogeny. Results indicate that revised outgroups affected the phylogenetic conclusions, particularly among genus and species level relationships, and that mtDNA analyses recover a different arrangement of family and genus relationships than proposed by prior morphology-only hypotheses. All analyses recovered monophyletic Zeidae, Cyttidae, and Oreosomatidae, and Zeniontidae, and non-monophyletic Parazenidae. Overall, results reflect the particular usefulness of mtDNA characters for examination of recent evolutionary events that shaped genus and species level relationships within Zeiformes, and the necessity of considering multiple lines of evidence to reveal the wider picture of zeiform evolution.

INTRODUCTION

Overview of Zeiform Fishes

The Zeiformes (including dories, lookdown dories, tinselfishes, and oreos) are an order of mid-to-deep water marine acanthomorph fishes known for their extremely protrusible jaws used to capture prey such as small schooling fishes and crustaceans (Heemstra, 1980; Scott and Scott, 1988; Heemstra, 1995). They are largely benthopalagic fishes living at depths between 50–1000 m. The order has a global distribution spanning tropical to temperate waters, with some species exhibiting near-worldwide distributions (e.g., *Zeus faber*), while others (e.g., *Neocyttus psilorhynchus*) exhibit limited home ranges (such as the waters off Australia and New Zealand; Fig. 1). Fishes within the genus *Zeus* are physically the largest zeiforms reaching a total length of approximately 90 cm, while species within *Zenion* represent the smallest, reaching a maximum total length of approximately 15 cm. The zeiform fossil record dates to the Late Cretaceous (late Campanian/early Maastrichtian, 72 mya; Tyler et al., 2000; Baciu et al., 2005; Tyler and Santini, 2005; Davesne et al., 2017). Thirty-three extant species are currently recognized as valid and are distributed among sixteen genera (Tyler et al., 2003; Tyler and Santini, 2005; Nelson et al., 2016; Fig. 2).

Traditionally the fishes of order Zeiformes are arranged within the families Cyttidae, Oreosomatidae, Parazenidae, Zeniontidae, Grammicolepididae, and Zeidae (Fig. 2). Cyttidae (Lookdown or Big-eye dories) are large-bodied fishes ranging primarily through the southeast

Atlantic and Indo-Pacific oceans. Oreosomatidae (Oreos) primarily range throughout the Southern Hemisphere, undergo exceptional metamorphosis from their juvenile to adult forms, and are known to have a deep-sea, elongated lifespan up to 210 years, making these fishes some of the longest living vertebrates. Fishes of what was historically considered family Parazenidae (Slender or Smooth dories) are found across the Atlantic and Pacific oceans, and exhibit a more elongate body form than most other zeiform genera. Zeniontidae (Armor-eye and Capro dories) are primarily found in the waters off the southern coast of Africa and in the western Pacific, and typically occupy a depth range from 300-600m. Grammicolepidae (Dwarf dories) range throughout the Atlantic and Pacific Oceans, and are typically collected as by-catch from deep-water trawling. Zeidae (Buckler or John dories, St. Peter's fish) are found in the Atlantic, Indian and Pacific oceans, and are often caught via deep-sea trawling for preparation and sale as restaurant food fishes.

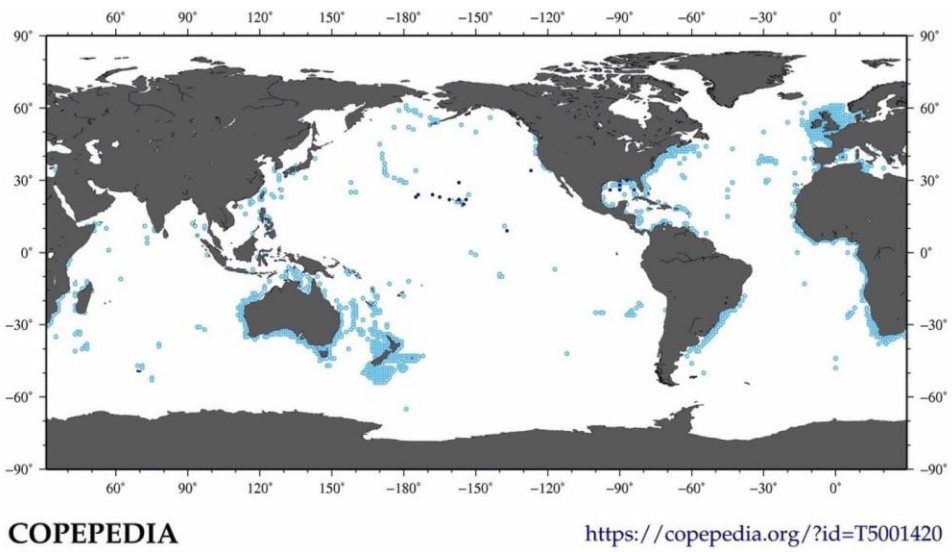

Figure 1. Geographic distribution of zeiform fishes. Zeiform fishes exhibit global distribution across the Atlantic, Pacific, and Indian oceans, representing regional endemics as well as more generalist species whose natural range spans temperate and tropical waters. High densities of zeiforms have been recorded along the coasts of Australia and New Zealand, the eastern coast of North America, and the northern coasts of Europe.

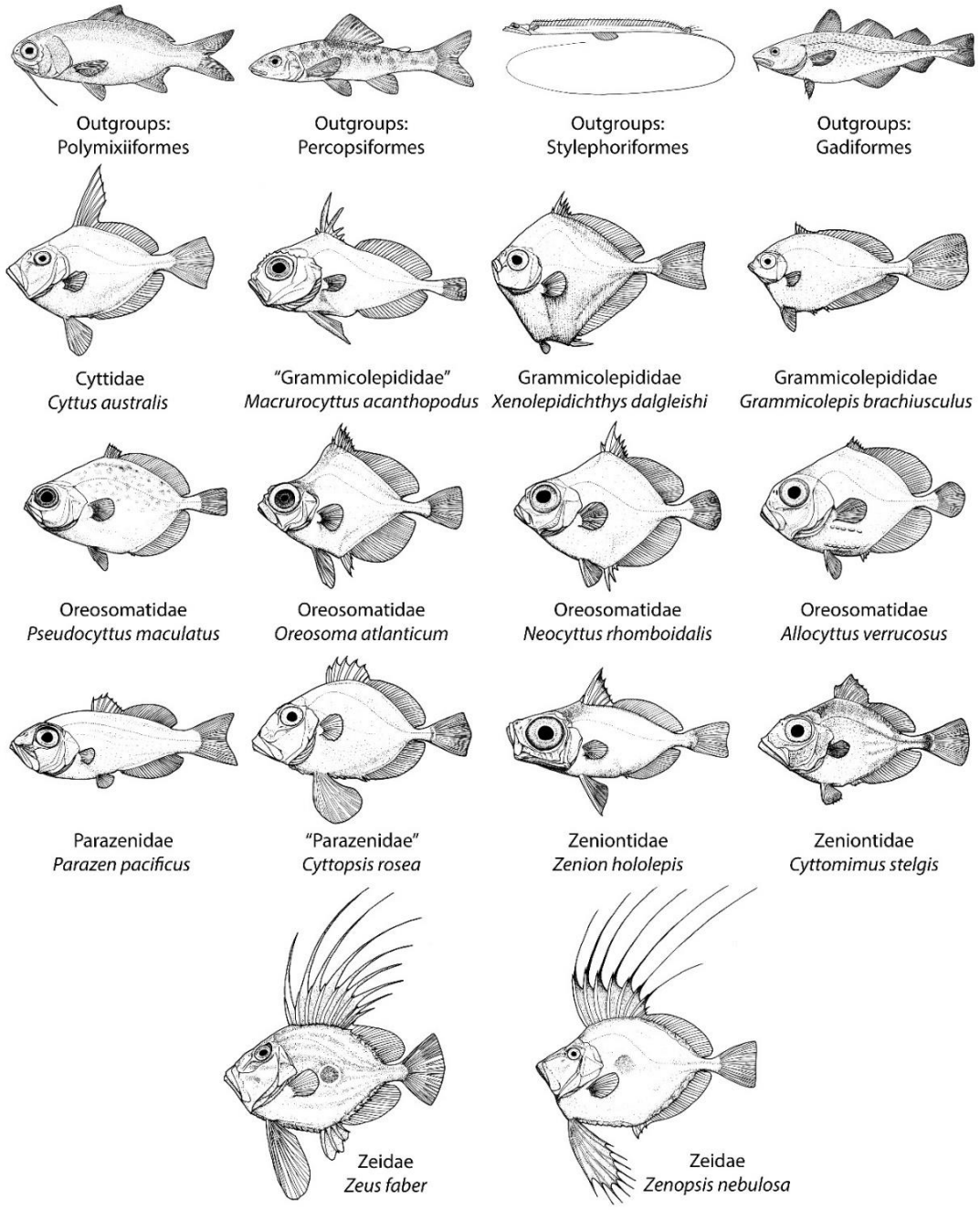

Figure 2. Zeiform fishes and selected outgroups. Taxa representative of the six families within Zeiformes as recognized by Tyler et al. 2003, along with selected outgroup taxa as per the Zeiformes + (Stylephorus + Gadiformes) clade recognized by Grande et al. (2013) and Grande et al. (2018). Illustrations drawn by Michael Hanson.

Zeiform Classification and Controversy

The interesting phylogenetic history of zeiform fishes dates back to Linnaeus' diagnosis of *Zeus* (*sensu lato*) in the *Systema Naturae* (Linnaeus 1758). More recently, zeiforms have been placed within the superorder Paracanthopterygii. Paracanthopterygii, erected by Greenwood et al. (1966), have been an enigmatic group of fishes with respect to membership, and characterized as a more primitive group of bony fishes with comparable morphological diversity to that of Acanthopterygii (i.e., the remainder of fishes also belonging to the Acanthomorpha) (Rosen, 1982; Grande et al., 2013). Since the conception of the group Paracanthopterygii, its taxon composition has been in flux. Most recently, however, there appears to be consistency among both morphological and molecular based studies that Zeiformes belongs within the Paracanthopterygii and that they are very closely related to gadiforms (i.e., cods and relatives).

In contrast, however, Rosen (1984) proposed seven morphological synapomorphies uniting Zeiformes with the order Tetraodontiformes (containing boxfishes, pufferfishes, filefishes and triggerfishes) and the order Caproidae (boarfishes). These characters included otolith shape, symmetry and composition of dorsal and anal fin radials, shape of the parasphenoid, interopercle, and premaxillary bones, and principal caudal fin ray count. In the same study, Rosen (1984) cited another four synapomorphies uniting Tetraodontiformes + Zeiformes, (i.e., length and organization of the caudal skeleton, hyoid arch structure, and arrangements of opercular bone elements).

Johnson and Patterson (1993), in their study of Acanthomorpha (spiny-rayed fishes; a highly diverse group comprising the "bush" at the top of the fishes' tree of life, containing approximately one-third of all vertebrate species on the planet). Those authors provided the first

thorough morphological evidence specifically in support of zeiform monophyly. Their big-picture morphological survey included representative species of all five zeiform families recognized at the time of publication, and cited seven synapomorphies diagnosing the order, including the placement of Baudelot's ligament, the structure of dorsal fin osteological elements, the reduction of the metapterygoid bone, the flexion of vertebral centra and rib placement, the configuration of gill arch elements, the configuration of cranial median cartilage, and the configuration of certain caudal-fin osteological elements. In addition, Johnson and Patterson (1993) hypothesized that Zeiformes might form a sister-relationship with the order Beryciformes (i.e., soldierfishes, squirrelfishes, and lanterneyes), citing the mode of articulation between the parietal and extrascapular cranial bones as a potential synapomorphy. However, Johnson and Patterson (1993) went on to reject such a clade as a possibility on the basis of parsimony, and instead upheld Zeiformes as a sister group to their temporarily named Euacanthopterygii (Beryciformes plus Percomorpha), a large clade containing thousands of species in over 250 families, though they admitted that they did not expect the clade so formed to endure further scientific scrutiny (Johnson and Patterson, 1993).

Neither Rosen's (1984) nor Johnson and Patterson's (1993) studies made any strides towards characterizing the relationships between or within the zeiform families. Later morphological studies, most notably that of Tyler et al. (2003), closely examined Rosen's (1984) proposed synapomorphies and determined that, while some characters did indeed seem to be indicative of a potential relationship between zeiforms and tetraodontiforms, only one was accepted as a consistent synapomorphy within their data set. The remainder of Rosen's proposed synapomorphies were determined to be either too inconsistent among the taxa under study for

adequate consideration, not unique to the taxa in question, or exhibiting evidence of convergent evolution rather than being diagnostic of a tetraodontiform-zeiform clade. Though largely outside the scope of their study, Tyler et al. (2003) noted that three of their parsimony analyses exhibited support of the zeiform + tetraodontiform topology hypothesized by Rosen (1984), while the fourth (the strict consensus phylogeny generated from their 26 most-parsimonious trees; Tyler et al., 2003:fig. 13), which was concluded to be the best-justified and most rational tree configuration, left Rosen's proposed relationships unresolved. Furthermore, Tyler et al. (2003) stated that Johnson and Patterson's (1993) proposed relationship between Zeiformes and Beryciformes was neither supported nor rejected by any of the resulting parsimony analyses. In addition, Tyler et al. (2003) were the first to comprehensively examine the monophyly of the order Zeiformes and the intra-relationships of its subgroups. Their study provided an extensive list of 103 diagnostic morphological characters delineating the order as a whole and providing synapomorphies for each of its component genera, thus establishing evidence for the monophyly of each of the clades within the order. The maximum-parsimony and maximum-likelihood phylogenies and the final consensus tree constructed by Tyler et al. (2003), later amended with the inclusion of fossil taxa in Tyler and Santini (2005), remained the definitive standard on zeiform systematics for many years.

The abbreviated classification of Zeiformes *sensu* Tyler et al. (2003) is as follows:

Order Zeiformes

Suborder Cyttoidei

Family Cyttidae (*Cyttus*)

Suborder Zeioidei

Family Oreosomatidae (*Pseudocyttus*, *Oreosoma*, *Neocyttus*, *Allocytus*)

Family Parazenidae (*Parazen*, *Cyttopis*, *Stethopristes*)

Family Zeniontidae (*Zenion*, *Capromimus*, *Cyttomimus*)

Family Grammicolepididae (*Macrurocyttus*, *Xenolepidichthys*, *Grammicolepis*)

Family Zeidae (*Zeus*, the type genus, and *Zenopsis*)

Molecular studies of Wiley et al. (2000) and Chen et al. (2003), in particular, also failed to recover evidence of Rosen's (1984) proposed sister relationship with Tetradontiformes, placing Zeiformes as a more basal order within Acanthomorpha. Wiley et al. (2000), in their twofold molecular and morphological test of Johnson and Patterson's (1993) hypotheses regarding the whole of Acanthomorpha, were the first to recover Zeiformes as the sister group to Gadiformes (i.e., cods and their allies, which have long been held as members of Paracanthopterygii) on the basis of the mitochondrial 12S ribosomal RNA gene and nuclear 28S large ribosomal subunit gene rDNA sequences. The proposed zeiform + gadiform clade was highly supported by the consensus tree generated from their molecular and morphological data, as opposed to the more traditional placement of zeiforms with either Tetraodontiformes or Beryciformes. Chen et al. (2003), although also recovering zeiforms and gadiforms as sister clades in their molecular study of Acanthomorpha, cited evidence of potential zeiform paraphyly. Miya et al. (2003), in attempting to construct a full-scale phylogeny of higher teleostean relationships, also recovered the zeiform + gadiform sister relationship based on their survey of whole mitogenomic data. In their subsequent molecular study of the phylogenetic position of *Stylephorus chordatus* (a rare deep-sea fish commonly called the Tube-eye), Miya et al. (2007)

were also the first to suggest that *Stylephorus* was more closely related to paracanthopterygians (most specifically to gadiforms and zeiforms) than previously hypothesized, citing a possible (*Stylephorus* + gadiforms) clade, which in turn formed the sister clade to zeiforms. More recently, Grande et al. (2013), as part of the NSF funded Euteleost Tree of Life initiative, corroborated the hypothesized Zeiformes + (*Stylephorus* + Gadiformes) relationship in a wide, dual-focused, molecular and morphological survey of basal acanthomorphs, citing multiple synapomorphies once thought to be diagnostic of zeiforms alone as diagnostic of the Zeiformes + (*Stylephorus* + Gadiformes) clade. In these and other such studies (e.g., Wiley et al., 2000; Chen et al., 2003; Miya et al., 2003; Li et al., 2007; Near et al., 2012; Grande et al., 2013), however, the zeiform taxa sampled were judiciously sparse, because until now, sorting out the intra-relationships within Zeiformes was beyond the broad focus of these far-reaching systematic investigations. Until publication of Grande et al. (2018) and completion of the present study, there have been no other molecular studies examining the intra-relationships of zeiforms to the same level of detail as Tyler (2003) and Tyler and Santini's (2005) morphological assessments, or examining the order Zeiformes in light of its current phylogenetic position: as a member of Paracanthopterygii, closely related to cods.

Purpose

The purpose of my study was to examine the phylogenetic intra-relationships of the order Zeiformes, given its newly accepted phylogenetic position [Zeiformes + (*Stylephorus* + Gadiformes)], using mitochondrial DNA analysis. (2) compare the results of my study with those proposed by Tyler et al. (2003) and Tyler and Santini (2005), which were based exclusively on morphological characters. (3) determine the extent to which mitochondrial data contribute to

phylogenetic reconstruction. This study differs from previous works because of its large taxon sampling and increased genetic markers, covering three mitochondrial genes: the non-coding 12s and 16s ribosomal RNA genes, and the protein-coding Coenzyme I (COI) gene.

The results of my mitochondrial research contributed to a larger collaborative work, in which I am a co-author (Grande et al., 2018). This larger work incorporated my mitochondrial sequence data in a study with five nuclear genes and a comprehensive reanalysis of the morphological data of Tyler et al. (2003) and Tyler and Santini (2005).

The balance of this thesis is divided into the following chapters:

Materials and Methods:

I. Genes

II. Tissue Acquisition and Extraction

III. Molecular Sample Preparation

IV. Sequence Alignment and Data Analysis

V. Phylogenetic Reconstruction

Results:

I. Phylogeny Based on Individual Gene Datasets

II. Phylogeny Based on Combined Gene Dataset

Discussion:

Use of Mitochondrial Sequence Data to Assess Zeiform Relationships

Comparison with the Results of Tyler et al. (2003) and Tyler and Santini (2005)

Comparison with the Results of Grande et al. (2018)

Contribution of mtDNA to Phylogeny within Zeiformes

MATERIALS AND METHODS

Section 1: Genes

While this research was conducted as part of a larger zeiform phylogenetic study spanning mitochondrial and nuclear genetic markers plus morphological characters, the methods outlined here will only cover the mitochondrial portion of the study.

The mitochondrial genome of fishes consists of a short circular chromosome of roughly 17,000 base pairs, encompassing 2 ribosomal RNA genes, 22 transfer RNA genes, and 13 protein-coding genes corresponding to enzymatic subunits for the electron transport chain and ATP production. These genomic elements, as well as the lack of introns and short non-coding intergenic sequences, are conserved across bilateric animals; this means that all animals exhibiting a bilaterally symmetrical body plan contain copies of these genes in their mitochondrial chromosome at the same chromosomal locations (Ladouakis and Zouros, 2017). The structural conservation of the mitochondrial genome is the first of several useful features that make mitochondrial genetic markers particularly well suited to the study of closely related taxa such as zeiform sub-groups. The nucleotide composition of mitochondrial genes can vary significantly among species, but the fact that all bilateric animals contain copies of the same mitochondrial genes allows for evolutionary comparisons across a wide range of taxa. Furthermore, as vertebrate reproduction enables only the mitochondria present in the maternal egg cell to be transmitted from parent to offspring, mitochondrial DNA represents a direct

evolutionary history along the maternal lineage. As an organism grows from the point of fertilization to maturity, the mitochondria present in each cell replicate along with the process of mitotic division, meaning that the multiple copies of the mitochondrial chromosome replicate with each cellular replication, and result in a higher cumulative amount of mtDNA per cell than the singular copies of nDNA chromosomes present in each cell's nucleus. The high copy number of mitochondrial chromosomes per cell make mtDNA easily extracted in high amounts for molecular study (Gissi et al., 2008; James et al., 2015). mtDNA also lacks the repair mechanisms encoded in nDNA, meaning that mtDNA aggregates mutations (insertions, deletions, and substitutions) more rapidly than nDNA. More mutations mean more variable sites within the mitochondrial genome, and more molecular characters from which to draw evolutionary comparisons among taxa under study (Yakes and Van Houten, 1997). Another feature of mtDNA replication that sets it apart from nDNA and makes it well suited for use in the study of closely related taxa is the nearly nonexistent potential for recombination; lack of segment swapping between chromosomes generates no confounding interference that could obscure the evolutionary signal present in the mitochondrial chromosome (Hagström et al., 2014).

While the structure of the mitochondrial genome is highly conserved among animals, the sequences of the mitochondrial genes, especially non-coding genes, display a high degree of variability among animals (Ladoukakis and Zouros, 2017). The sequences of these mitochondrial elements can vary a good deal particularly at the family and genus level, especially with respect to non-protein-coding genes, and can provide much insight into these relationships between closely related groups. mtDNA exhibits evolutionary rate heterogeneity across different genes along the chromosome, as non-coding genes and protein-coding genes experience different

functional constraints that determine what mutations will allow the products of those genes to continue to function unimpeded, and thus what mutations will persist in the genome along subsequent generations. Non-coding genes, due to the lack of resultant protein products encoded by those genes, will typically experience more mutations and evolve faster than protein-coding genes. Within protein-coding genes, rates of evolution can also vary across codon positions; sites corresponding to the first and second codon positions will typically experience slower rates of evolution, as the first and second codon positions typically determine which amino acids will be incorporated into the protein product of that gene. Nucleotide sites corresponding to the third codon position typically evolve much faster, as mutations in this position will often manifest silently without affecting the amino acid sequence of the resultant protein product (Jia and Higgs, 2008).

Due to the rapid rate of evolution experienced by the mitochondrial genome, mitochondrial genes may exhibit genetic saturation at deeper nodes on the evolutionary tree, meaning that multiple substitutions at a single site may seem to indicate that the rate of evolution at that site is lower than has actually occurred (Philippe et al., 2011). This makes mitochondrial DNA particularly suited to the study of closely related taxa, as the rapid rate of mitochondrial genetic evolution can best elucidate recent evolutionary changes without the threat of saturation that confounds mitochondrial evolutionary signal in older, more divergent lineages.

Three mitochondrial genes were chosen for sequencing: the complete small (12S) and large (16S) rRNA genes (Kocher et al., 1989; Titus, 1992; Palumbi, 1996; Feller and Hedges, 1998), and the protein-coding coenzyme I (COI) gene (Ward et al., 2005; Ivanova et al., 2007). The 12s, 16s, and COI genes have already been widely studied among fishes and their sequences

are available across a vast array of taxa, including a wide variety of species within Gadiformes, the recognized sister clade to Zeiformes (Wiley et al., 2000; Chen et al., 2003; Miya et al., 2003; Li et al., 2007; Near et al., 2012; Grande et al., 2013). The 12S and 16S rRNA genes were chosen because they are two non-protein-coding mitochondrial genes that evolve at a faster rate than the protein-coding mitochondrial gene COI, the third gene included in this study. Though the 12s and 16s rRNA genes are highly conserved themselves, especially in vertebrates, these genes nonetheless exhibit a greater deal of variation across closely related species than protein-coding genes, making them good candidates for investigation into zeiform family- and genus-level relationships.

The protein-coding nature of the COI gene, while also highly conserved among vertebrates due to the essential nature of the cytochrome c oxidase I (*cox1*) protein in the electron transport phase of cellular respiration, necessitates that nucleotide substitutions in the COI sequence must result in a functional *cox1* protein. This largely limits nucleotide substitution sites to the third codon position, typically resulting in synonymous codon triplets and ensuring the same amino acid is integrated into the protein at the site of substitution, maintaining its function in electron transport. While most mutations in the COI gene sequence do not result in changes to the protein sequence, related species with similar metabolic requirements will experience similar selective pressures and exhibit similar codon sequences in the COI gene. As a result, the COI gene displays greater intraspecific sequence diversity than interspecific. These factors make the COI gene useful for resolving lower-level relationships among taxa.

Section 2: Tissue Acquisition and Extraction

Tissue samples from every zeiform family representing 13 out of 16 genera (excepting *Capromimus*, *Macrurocyttus*, and *Stethopristes*) and 24 of 33 species, with multiple specimens per taxon where available, plus selected outgroup taxa of the orders Myctophiformes, Lampriformes, Beryciformes, Percopsiformes, Polymixiiformes, Gadiformes, and Stylephoriformes were chosen for the study (Appendix 1). Polymixiiforms (represented here by *Polymixia lowei*), percopsiforms (represented by *Percopsis transmontana*, *Aphredoderus sayanus*, and *Amblyopsis spelea*), gadiforms (represented by *Gadus morhua* and *Muraenolepis microps*) and stylephoriforms (i.e., *Stylephorus chordatus*), are all members of Paracanthopterygii. The more distant outgroups include Myctophiformes (lanternfishes, represented here by *Benthosema glaciale*), which are the sister group to Lampriformes + [Paracanthopterygii plus Acanthopterygii] (Nelson et al., 2016). Lampriformes are represented here by *Lampris guttatus*, and Acanthopterygii by *Beryx splendens*. Tissues were obtained from U.S. and international museums and research collections (Appendix 1) and stored under 95% ethanol at 4^{o}C upon receipt, prior to DNA extraction. Extraction was carried out using the DNEasy Blood and Tissue Kit (Qiagen) according to manufacturer's protocol with adjustments to the final elution volume based on the quality of tissue sample used for the extraction. DNA extracted from high-quality starting tissue was eluted into 80 ul EB final volume, whereas DNA extracted from low-quality tissues (desiccated samples where the containment Eppendorf tube was not sealed properly or opened during transit) was eluted into 30-40 ul EB final volume. The desiccated tissue samples in question were the only representative tissues obtainable for the

ingroup taxon *Cyttus novazealandiae*, and were re-hydrated and included for DNA extraction to determine whether enough intact DNA could be recovered for analysis.

Section 3: Molecular Sample Preparation

Following DNA extraction, 12S and 16S rRNA and COI genetic markers were amplified from each specimen via PCR reaction from primer sequences previously published by the Euteleost Tree of Life project (Appendix 2). Amplification was initially carried out using ThermoFisher GoTAQ DNA polymerase according to marker-specific conditions [Appendix 3]. Fragment sizes were assessed via 1% agarose gel electrophoresis and PCR carried out again with increasing annealing temperature stringency as necessary. Incremental adjustments to primer annealing temperature prevented the primers from annealing imperfectly at regions other than the intended loci, and prevented amplification of extraneous additional regions of DNA that would compromise the quality of sequencing data obtained. Once optimal annealing temperatures were established for all three markers, final PCR amplification was carried out for all ingroup and selected outgroup specimens using ThermoFisher High Fidelity Platinum TAQ DNA polymerase, and PCR product fragment sizes were verified by 1% agarose gel electrophoresis at a run length of 25-30 minutes to ensure clear band separation. Sample quality and concentrations were then assessed by NanoDrop for all samples for each genetic marker, at which point sequencing plates were prepared and sent to the University of Washington DNA Sequencing and Gene Analysis Center for capillary Sanger sequencing.

Section 4: Sequence Alignment and Data Analysis

Upon completion of sequencing, raw sequence data for all taxa were retrieved and FASTA sequence files uploaded into the Geneious v7.1.9 (Kearse et al., 2012; www.geneious.com) sequence analysis software for alignment and editing. Additional sequence data, corresponding to relevant ingroup and outgroup taxa for which tissue samples had not been obtainable, were included via download of previously published sequences made available via the NCBI GenBank database [Appendix 1]. Raw novel sequences and published sequences obtained from GenBank for all specimens were uploaded separately for each genetic marker under study and forward reading frame sequences were aligned in Geneious via progressive pairwise global alignment, according to the algorithm developed by Feng and Doolittle (1987). All ingroup and outgroup sequences within each alignment for each genetic marker were then grouped by taxon and the alignments further edited by eye. Base-pair credibility was judged on the basis of chromatograph quality per nucleotide position, and edits were made via assessment of base quality as reported by the raw sequence chromatographs compared to the sequence quality at the base position in question across closely related taxa in the dataset (e.g., position 157 as reported across all specimens of the genus *Zeus*). Additional editing was performed for the protein-coding COI locus where the sequences were edited by eye on the basis of chromatograph quality at each nucleotide position, and then translated to amino acids. The translation was then further assessed by eye based on base call quality per position as indicated in the sequence chromatographs. This final editing step ensured that no stop codons had been mistakenly edited into the COI sequences included in the dataset in the initial edits made on the basis of raw chromatograph quality alone.

Section 5: Phylogenetic Reconstruction

Final sequence alignments for each genetic marker were run through JModelTest (Posada, 2008) in order to determine the model(s) of evolution best fitting the nucleotide substitution patterns evident in each dataset. Non-protein-coding 12S and 16S datasets were tested without partitioning, while the protein-coding COI dataset was partitioned by codon position and each codon position tested individually for model fit. Best-fit models of evolution for each dataset partition are listed in Table 1.

Marker	**Marker partition**	**Model**
12S	n/a	GTR + G
16S	n/a	GTR + G
COI	First codon position	HKY
COI	Second codon position	HKY
COI	Third codon position	GTR + G

Table 1. Best-fit evolutionary models as determined by JModelTest. Evolutionary models were applied to Maximum Likelihood and Bayesian inference tree-building criteria, according to each dataset; non-coding 12S and 16S rRNA datasets exhibit a roughly equivalent rate across all nucleotide positions, while the protein-coding COI dataset experiences differential rates of evolution across codon positions that preserve protein function. The COI dataset was partitioned by codon position and assessed for the best-fit evolutionary model for each partition.

Once the appropriate evolutionary models had been determined for each locus and codon partition (where applicable), sequence alignments for each of the three genetic markers were run separately, and as one dataset containing all three markers, according to each of three different tree-building algorithms: Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI). The 12S and 16S datasets were run unpartitioned and the COI dataset was run both unpartitioned and separated into three codon partitions according to the model of evolution determined by JModeltest for each locus dataset. MP analyses were constructed using PAUP

version 4.0; Swofford, 2002), ML analyses used GARLI version 2.1 (Bazinet et al., 2014), and BI analyses used MrBayes version 3.2 software (Ronquist and Huelsenbeck, 2003; Ronquist et al., 2012). Initial trees were built according to each method for each marker at 250 thousand replicates (GARLI, MrBayes), or until the program had determined the shortest tree found per replicate (PAUP). [Appendix 4].

At this stage, all sequences that had been obtained for the available taxa in the study were included in nexus-run code files (.NEX) and the nexus files uploaded to each tree builder software's command prompt interface in order to generate initial-alignment trees with a low number of program iterations. These preliminary trees were assessed to determine the reliability of the sequence alignment on the basis of ingroup vs. outgroup sorting. The tree topologies generated by this initial low number of program iterations were expected to sort study taxa into at least two large but distinct groupings on the basis of outgroup vs ingroup sequences, with more closely related taxa sorting nearer each other within those two loose categories. Sequence alignments for each genetic marker were then re-assessed for specimens that had sorted far out-of-place in the initial trees, or for specimens that generated wildly different branch lengths on the tree than specimens of related taxa. First, any accidentally misaligned or frame-shifted sequences were corrected. Once misaligned sequences were realigned back to the correct frame, alignments were then re-assessed more closely for sequence quality. If it was determined that low-quality sequence was the most likely cause of anomalous sorting or anomalous branch length in the preliminary tree structure, the raw DNA extract corresponding to the specimen in question was re-PCR amplified and re-sequenced where possible prior to re-inclusion in the overall dataset. If the low-quality sequence had been generated from a low-quality tissue sample (as was the case

with taxon *Cyttus novazealandiae*), re-sequencing was attempted and the higher-quality sequence was selected for inclusion in the dataset. If the sequence quality did not improve with re-sequencing and additional tissue could not be obtained for re-extraction, the specimen/taxon was removed from the dataset.

Upon performing re-sequencing where necessary and adding new sequences for new specimen tissue samples where obtainable, additional minor alignment edits were made to close base gaps overlooked during initial alignment edits and to confirm sequence quality for all specimen sequences in each dataset. The final edited alignments were then re-run for each tree-building program at 10 million iterations, at which point genus- and species-level relationships could become well established in the trees. The resulting trees were then evaluated again for any outlier branches of excessive length relative to closely related branches, or any taxon sorting blatantly out-of-place on the tree (e.g., distant outgroups sorting within an ingroup clade or vice versa) that would indicate another reassessment of that specimen's sequence within the overall alignment.

At this stage, the trees generated at 10 million program iterations and a 25% burn-in before which the topologies of each tree had stabilized (ML, BI), or until the shortest possible tree had been determined (MP), were then assessed for remaining polytomies that had not been resolved with increased run length. Where polytomies were expected to be a result of identical aligned sequences representing multiple specimens of a single taxon, these identical sequences were detected in Geneious and removed from the alignment.

Individual marker alignments, with identical sequences removed, were the finalized datasets then used to generate a series of combined datasets for gene-by-gene jackknifing, where

contribution of individual markers to overall tree topology could be assessed. The individual 12S, 16S, and COI datasets were combined into three configurations of two markers, partitioned according to non-coding markers and codon position where applicable, for a total of three additional datasets, each containing two alignments, to be evaluated together according to each build criterion. A seventh partitioned dataset was built containing the alignments of all three markers, which would be used to generate a consensus tree for each build criterion (MP, ML, and BI). The resulting seven trees generated per build criterion were evaluated amongst themselves for ingroup clade congruence, at which point each genetic marker could be assessed on the basis of its contribution to the overall tree topology and within each zeiform family. Any clades persisting from a single-marker tree to the two-and three-mitochondrial gene consensus trees were determined to be a result of a single marker's greater signal influence on the overall topology

RESULTS

First, trees generated from each of the non-coding 12s and 16s gene datasets will be discussed individually, followed by the tree generated from the individual coding COI gene dataset, and finally the tree generated from the combined dataset of all three genes under study.

Section 1: Phylogeny based on Individual Gene Datasets - Non-Coding Genes:

Each of the three different analytical conditions yielded dissimilar branching patterns between the 12s and 16s gene sequence datasets, though all were rooted with the same selection of outgroup taxa. Maximum Likelihood (ML), Bayesian (BI), and Maximum Parsimony (MP) analyses all yielded tree topologies with variable degrees of resolution within and between loci, as well as differential ability to resolve relationships at the family and genus level. Within each locus dataset, there was no consensus clade resolution at the family level among methods. Overall, BI analyses recover a significant degree of polytomy in both inter- and intra-family relationships, and low overall clade resolution in comparison to ML and MP methods, though the 12s dataset does yield higher resolution at the genus level than the 16s dataset under all three conditions.

The 12s dataset yielded a higher degree of clade resolution when analysed under ML and MP conditions than it did under BI conditions. Under BI conditions, the 12s tree recovered relationships at the order and genus levels with greater resolution than at the family or species levels, though some relationships were recovered with relatively low support (posterior probabilities; pp: 0.56–1.0; Fig. 3). All families except Parazenidae were recovered as

monophyletic (pp: 0.67–1.0) with memberships consistent with those of Tyler et al. (2003) and Tyler and Santini (2005). While BI methods for the 12s dataset resolved Zeiformes as monophyletic, it was unable to resolve which family is the most basal among zeiform fishes.

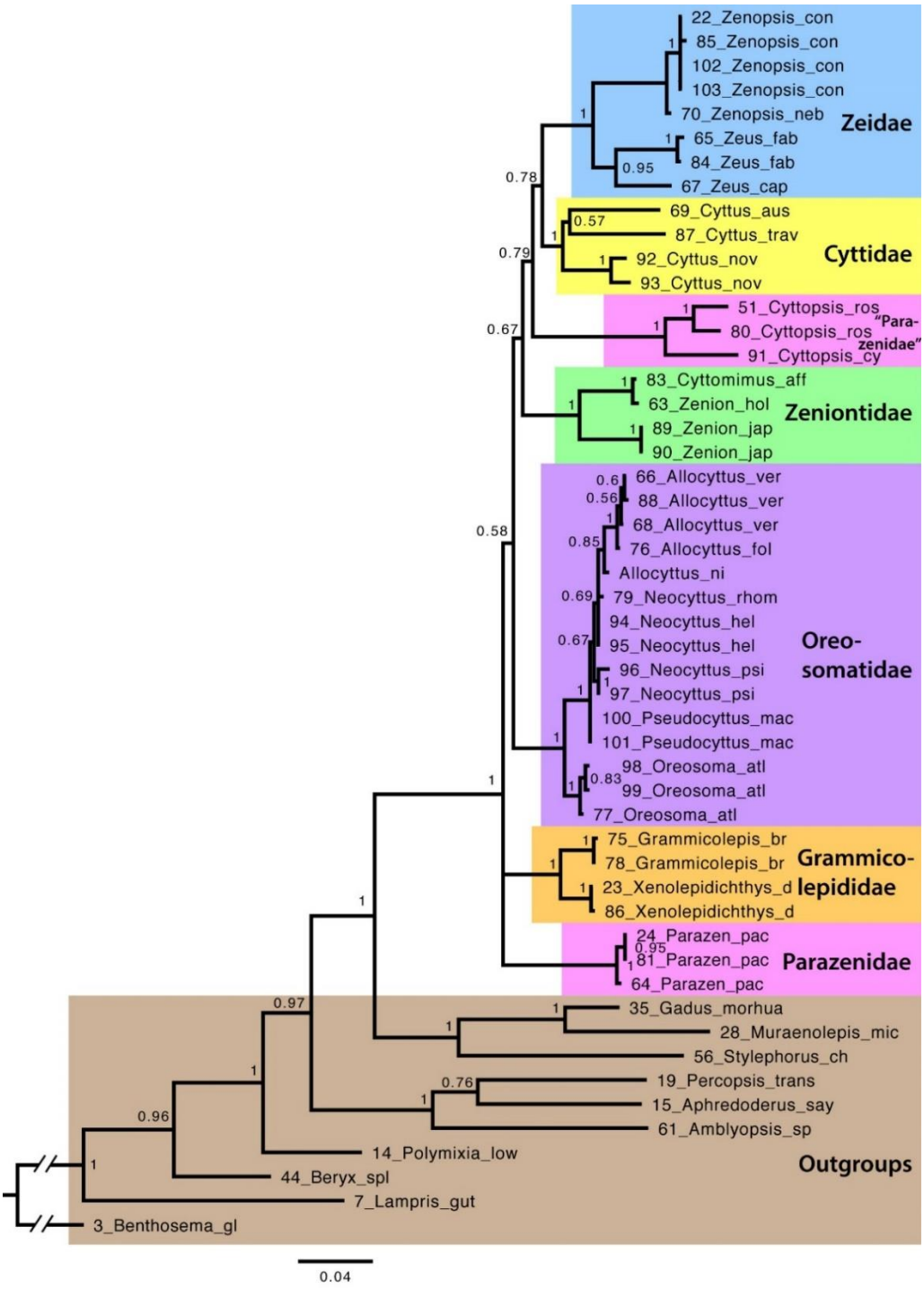

Figure 3. Bayesian Inference (BI) 12s consensus tree as generated by MrBayes v3.2 (Ronquist et al., 2012) using sequence data for the non-coding mitochondrial 12s gene, under the substitution model given in Table 1. Support values at nodes are posterior probabilities. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1. Colors indicate zeiform families as recognized by Tyler et al. (2003).

Monophyletic Grammicolepididae and genus *Parazen* form a polytomy with all other zeiforms (pp: 0.58) and together sort as more basal within the order. A monophyletic clade consisting of Oreosomatidae + {Zeniontidae + [*Cyttopsis* + (Cyttidae + Zeidae)]} was recovered at a posterior probability of 0.67. Cyttidae were recovered as the sister clade to Zeidae (pp: 0.78), representing the most derived clade within Zeiformes. Genera *Cyttopsis* and *Parazen* were recovered each as a discrete clade, splitting the family "Parazenidae" as represented by Tyler et al. (2003). Within each family, genus-level relationships were recovered with moderate to high support (pp: 0.56-1.0), where they could be resolved; exceptions include genera *Neocyttus* and *Pseudocyttus* within Oreosomatidae, where very short branch lengths indicate few distinguishing molecular characters for these relationships.

ML conditions for the 12s dataset recovered a monophyletic Zeiformes (Fig. 4) and showed a higher degree of resolution at the family level. Families Grammicolepididae, Oreosomatidae, and Zeniontidae were recovered as monophyletic. Family Cyttidae did not resolve as monophyletic, with *Cyttus novazealandiae* resolved as sister to *Zeus faber* among Zeidae, and the analysis was unable to unite *Cyttus australis* and *Cyttus traversi* in a monophyletic clade. Conversely, non-monophyletic Cyttidae would break family Zeidae into two clades. The first clade resolved genus *Cyttopsis* + (*Cyttus novazealandiae* + *Zeus faber*) allied with non-monophyletic *Cyttus australis* and *Cyttus traversi*. The remaining zeids were resolved as *Zeus capensis* sister to genus *Zenopsis*. Grammicolepididae were recovered as most basal, i.e., sister to all other Zeiform families, followed sequentially by genus *Parazen*, family Oreosomatidae, and family Zeniontidae, with non-monophyletic Cyttidae and Zeidae as the most derived groups under ML conditions.

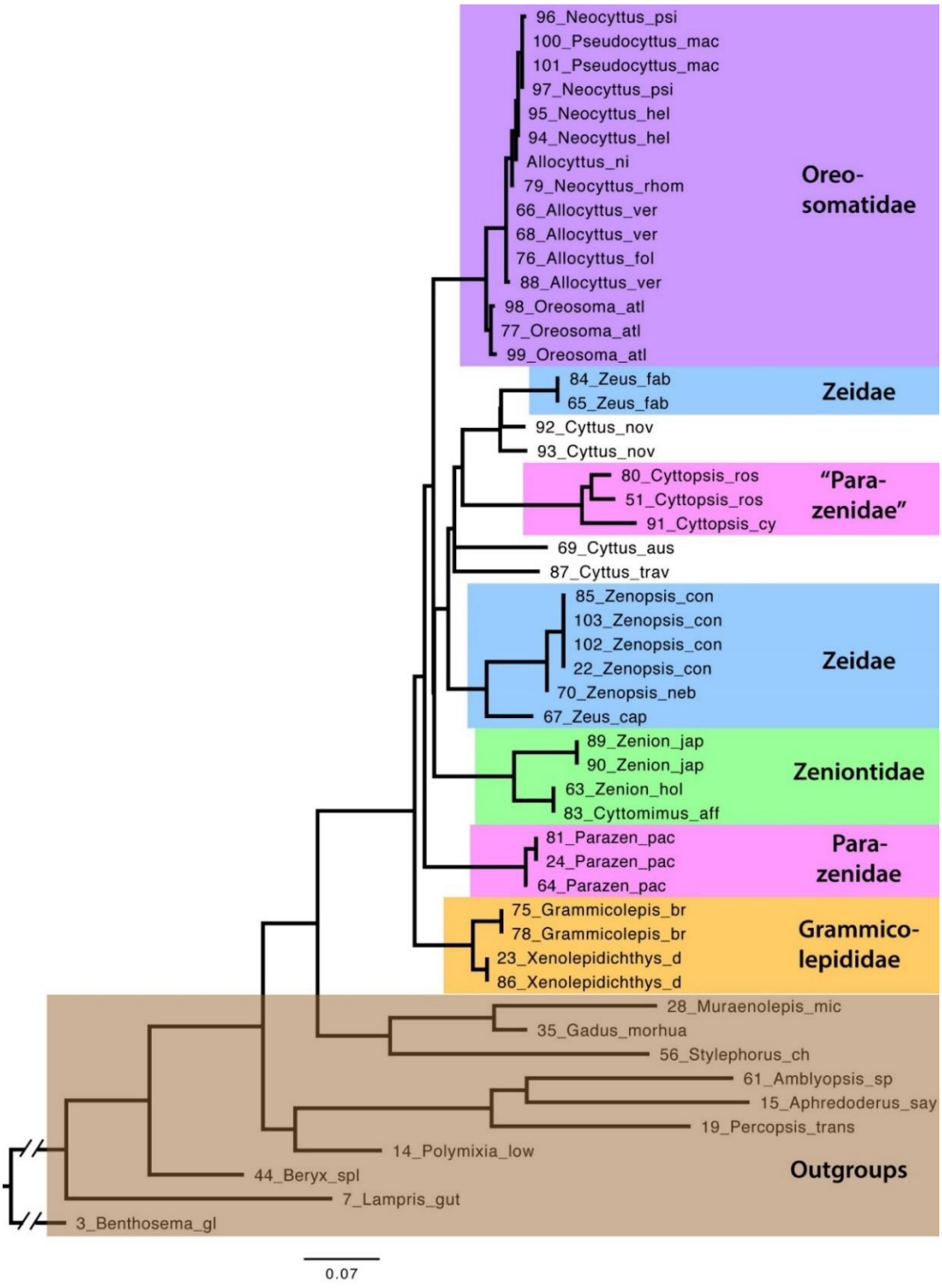

Figure 4. Maximum Likelihood (ML) 12s consensus tree as generated by Garli v2.0 (Zwickl, 2006), using sequence data for the non-coding mitochondrial 12s gene, under the substitution model given in Table 1. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1. Colors indicate zeiform families as recognized by Tyler et al. (2003).

"Parazenidae" were again recovered as non-monophyletic under ML. Oreosomatidae showed the least resolution at the genus level, with *Allocyttus* and *Neocyttus* not resolved beyond what is effectively a polytomy.

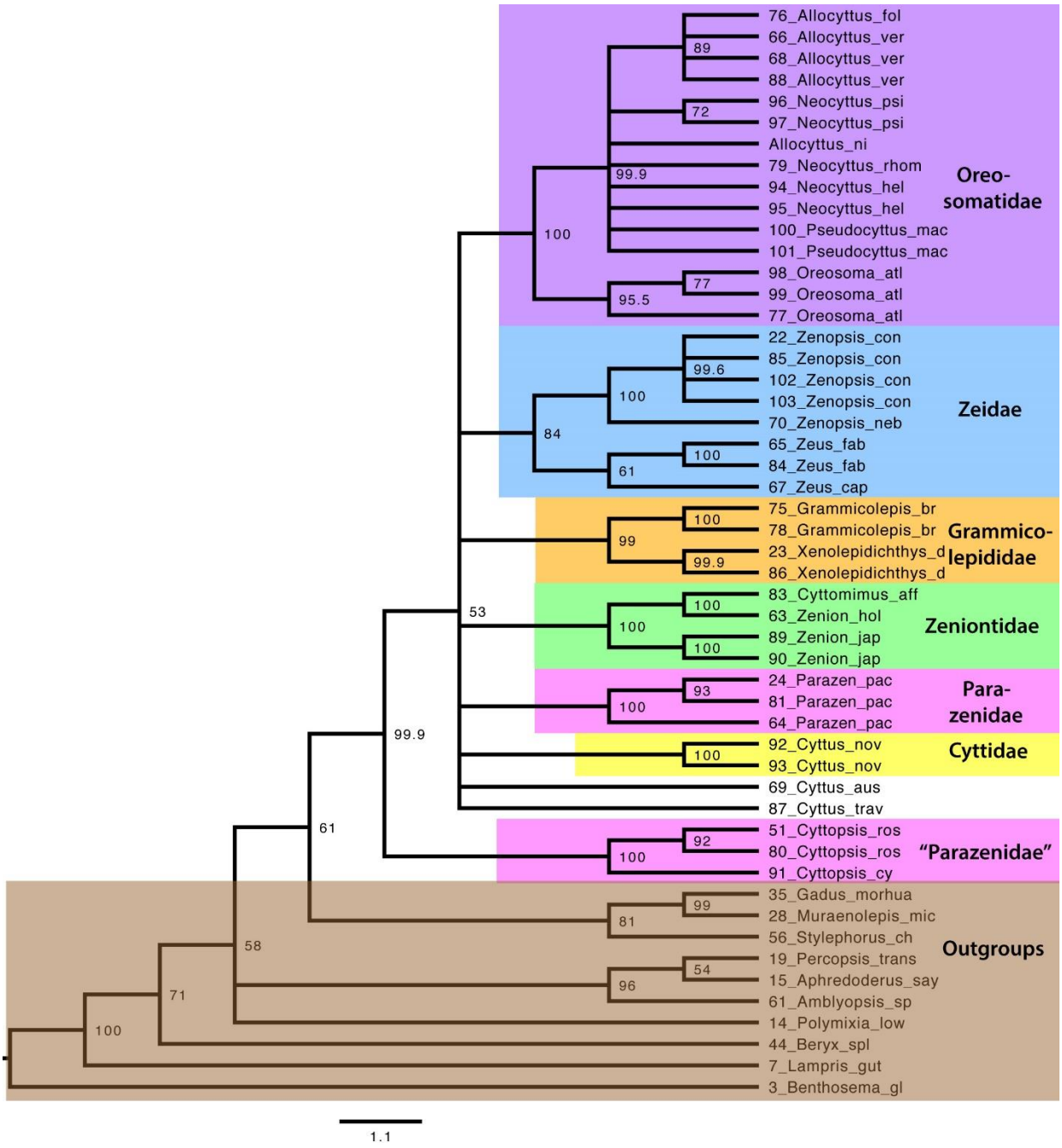

Figure 5. Maximum Parsimony (MP) 12s consensus tree as generated from by PAUP* v4b10 (Swofford, 2002), using sequence data from the non-coding mitochondrial 12s gene. Support values at nodes are bootstrap percentages. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1. Colors indicate zeiform families as recognized by Tyler et al. (2003).

Under MP conditions (Fig. 5), the 12s dataset yielded a tree topology indicating a monophyletic Zeiformes, with genus *Cyttopsis* as sister to all other zeiform fishes, and preserving a non-united "Parazenidae", but with relationships among other families unresolved. Families Grammicolepididae, Oreosomatidae, Zeidae, and Zeniontidae were recovered as monophyletic, as was genus *Parazen*. Cyttidae were not resolved into a discrete clade, nor was there resolution among many genus-level relationships within Oreosomatidae.

The 16s dataset analyzed under all three tree build criteria generated trees exhibiting the least resolution among the two non-coding individual genes. BI criteria, while recovering a monophyletic Zeiformes (pp: 0.69; Fig. 6), only preserved families Cyttidae, Zeidae, and Zeniontidae as monophyletic (pp: 0.66-1.0). The family Zeidae was recovered as most basal and sister to all other zeiform fishes (pp: 1.0), but no other family-level relationships were resolved under BI conditions. Neither Grammicolepididae nor Oreosomatidae was resolved into a discrete family, nor were genera Parazen and Cyttopsis resolved into "Parazenidae" as per Tyler et al. (2003).

ML analyses of the 16s dataset yielded more resolution than the BI tree. Like the BI tree, the ML tree recovered a monophyletic Zeiformes (Fig. 7). Also, families Cyttidae, Grammicolepididae, and Zeniontidae were resolved as monophyletic, as were genera *Cyttopsis* and *Parazen.* Genus *Zeus* was recovered as most basal among zeiforms, but Zeidae did not form a clade. Next most basal was family Zeniontidae and a clade with Grammicolepididae as sister to genus *Cyttopsis*. Higher-order relationships were largely indistinct. Oreosomatidae were not resolved into a discrete family group and instead formed a polytomy also containing the family Cyttidae and genera *Parazen* and *Zenopsis*.

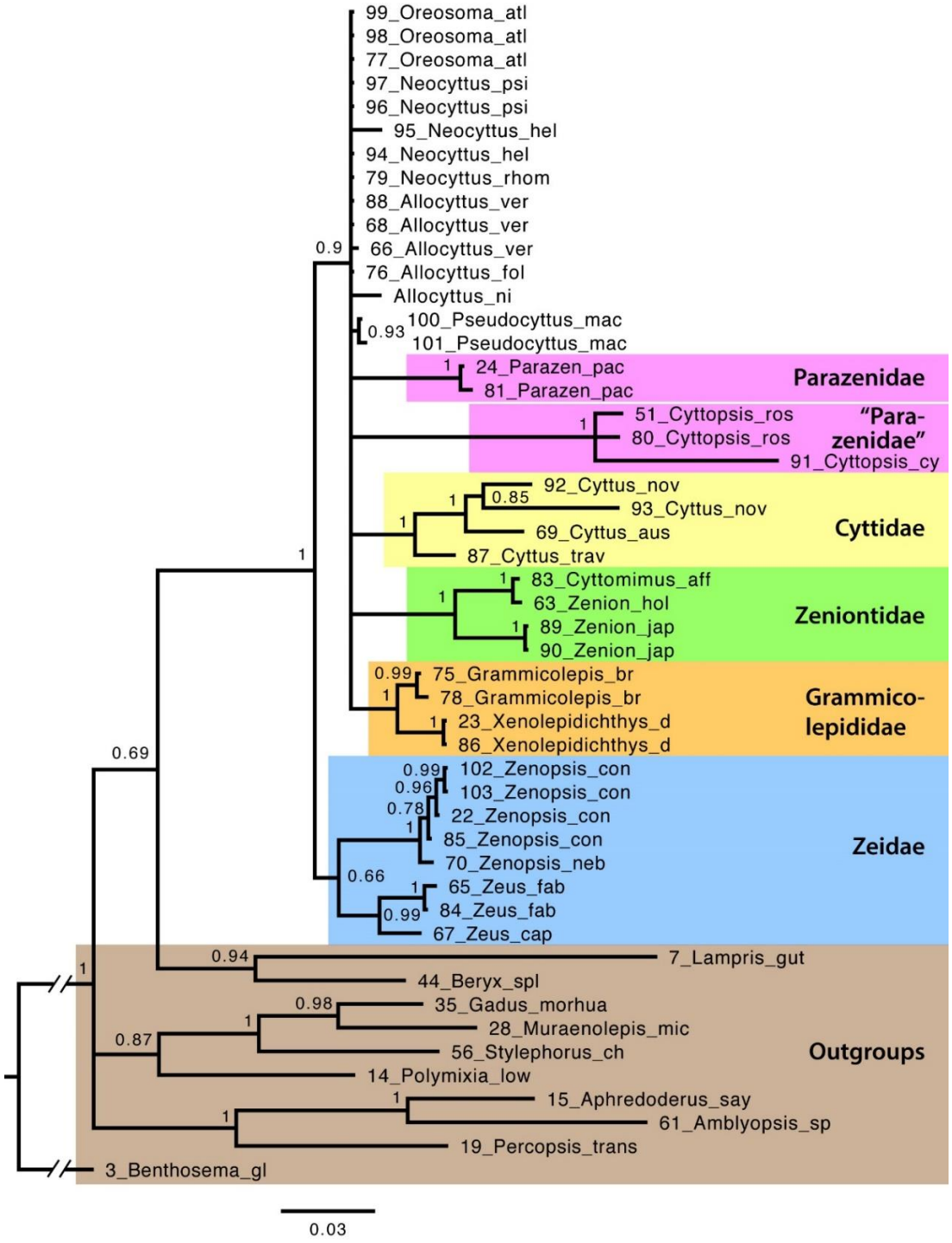

Figure 6. Bayesian Inference (BI) 16s consensus tree as generated by MrBayes v3.2 (Ronquist et al., 2012) using sequence data for the non-coding mitochondrial 16s gene, under the substitution model given in Table 1. Support values at nodes are posterior probabilities. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1. Colors indicate zeiform families as recognized by Tyler et al. (2003).

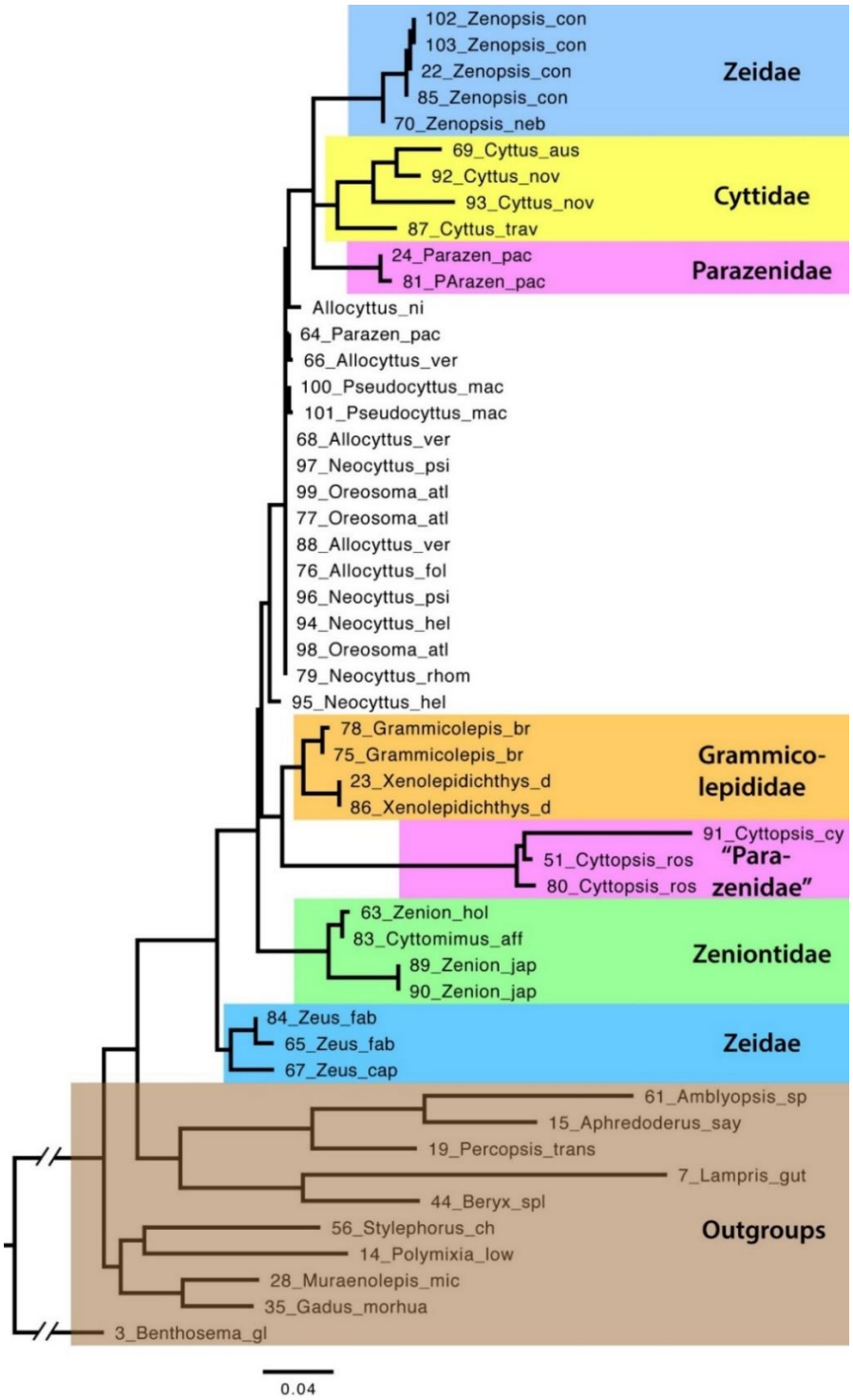

Figure 7. Maximum Likelihood (ML) 16s consensus tree as generated by Garli v2.0 (Zwickl 2006), using sequence data for the non-coding mitochondrial 16s gene, under the substitution model given in Table 1. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1. Colors indicate zeiform families as recognized by Tyler et al. (2003).

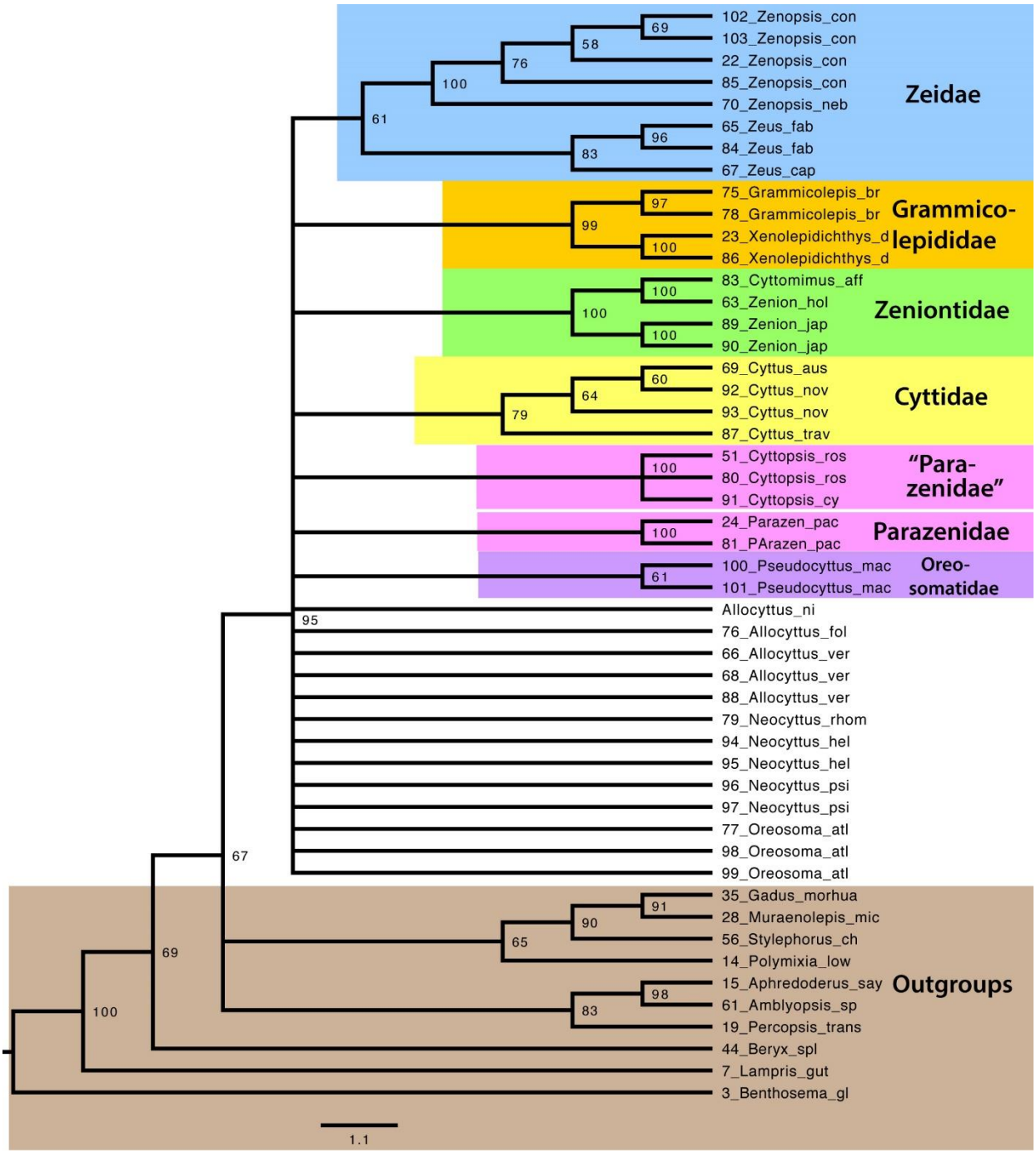

Figure 8. Maximum Parsimony (MP) 16s consensus tree as generated from by PAUP* v4b10 (Swofford, 2002), using sequence data from the non-coding mitochondrial 16s gene. Support values at nodes are bootstrap percentages. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1. Colors indicate zeiform families as recognized by Tyler et al. (2003).

MP analysis of the 16s dataset, while it recovered a monophyletic Zeiformes, did not further resolve any relationships among families within the group (Fig. 8). Cyttidae, Grammicolepididae, Zeidae, and Zeniontidae were recovered as monophyletic families, and both *Cyttopsis* and *Parazen* were recovered as discrete genera in keeping with the expectation of non-monoplyletic "Parazenidae". However, Oreosomatidae did not form a clade, with only the two species of the genus *Pseudocyttus* grouped together.

Section 2: Phylogeny based on Individual Gene Datasets - Coding Gene:

BI analysis of the protein-coding COI dataset yielded a monophyletic Zeiformes (pp: 1.0; Fig. 9), with lower resolution at the family level and higher resolution at the genus level (pp: 0.76-1.0). All families were recovered as monophyletic, as were genera *Cyttopsis* and *Parazen* (pp: 0.98-1.0). However, no family was resolved as most basal, though the BI criterion did resolve sister relationships (*Cyttopsis* + Grammicolepididae) and (Cyttidae + Oreosomatidae), at posterior probabilities of 0.77-0.99. Relationships at the genus level were well resolved, only exhibiting polytomies at the species level where multiple individuals of the same species were included within Oreosomatidae.

ML analysis of the COI dataset also indicated a monophyletic Zeiformes (Fig. 10), as well as all zeiform families and the genera *Cyttopsis* and *Parazen*, though this analysis did not resolve any family as more basal than the others. The genus *Cyttopsis* was recovered as sister to Grammicolepididae, and families Cyttidae and Oreosomatidae were also recovered as sisters. Within each family, genus-level relationships were also well resolved, showing polytomies only at the species level where multiple individuals of the same species were included. This tree topology exhibited the highest resolution within Oreosomatidae of any of the individual gene

trees, resolving the genus-level relationships within the group as *Pseudocyttus* + [*Oreosoma* + (*Allocyttus* + *Neocyttus*)].

The MP criterion for the COI dataset (Fig. 11) was the only single-gene tree that did not recover a monophyletic Zeiformes as sister to (Gadiformes + *Stylephorus*); instead, (Gadiformes + *Stylephorus*) sorted within the clade containing all zeiform fishes. However, despite the erroneously sorting nearest outgroup, all zeiform families were recovered as monophyletic, as were genera *Cyttopsis* and *Parazen*. However, the tree has a basal polytomy among all the families, and thus no family could be determined as the most basal within the group. On the other hand, genus-level relationships within each family are well resolved, as best seen within Oreosomatidae.

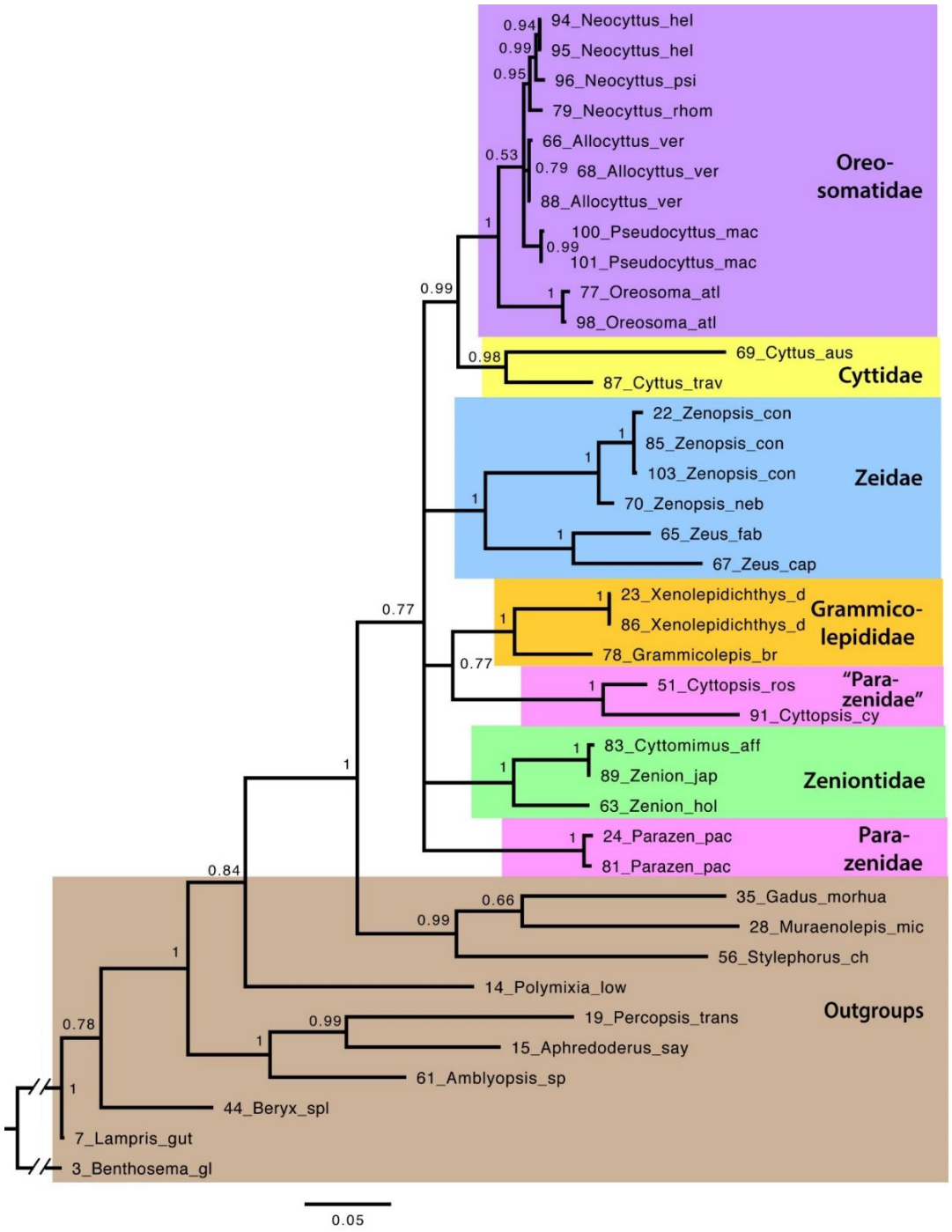

Figure 9. Bayesian Inference (BI) COI consensus tree generated from dataset partitioned by codon position as generated by MrBayes v3.2 (Ronquist et al., 2012) using sequence data for the coding mitochondrial COI gene, partitioned by codon position under the substitution models given in Table 1. Support values at nodes are posterior probabilities. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1.

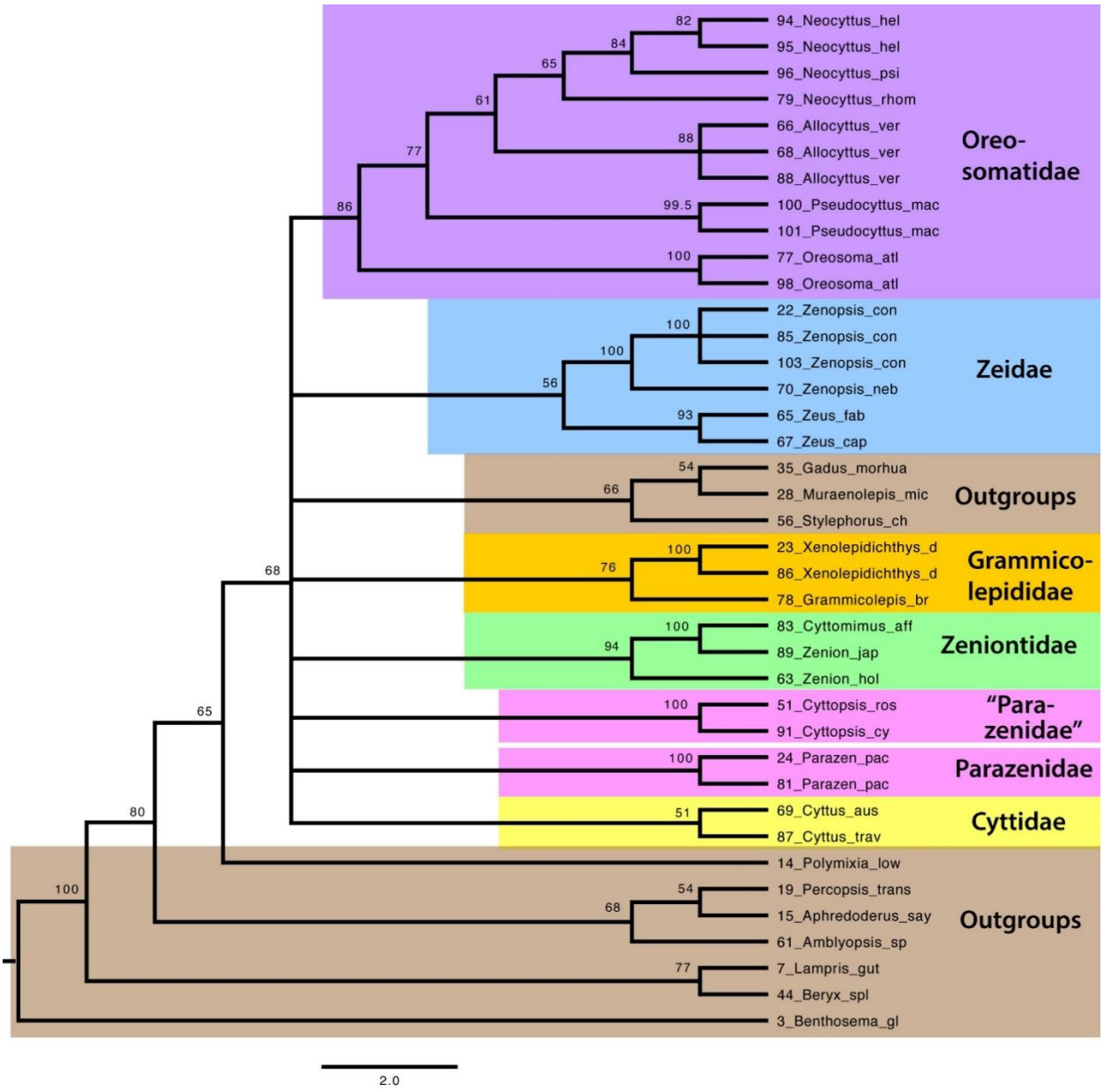

Figure 10. Maximum Likelihood (ML) COI consensus tree generated from dataset partitioned by codon position as generated by Garli v2.0 (Zwickl, 2006) using sequence data for the coding mitochondrial COI gene, partitioned by codon position under the substitution models given in Table 1. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1.

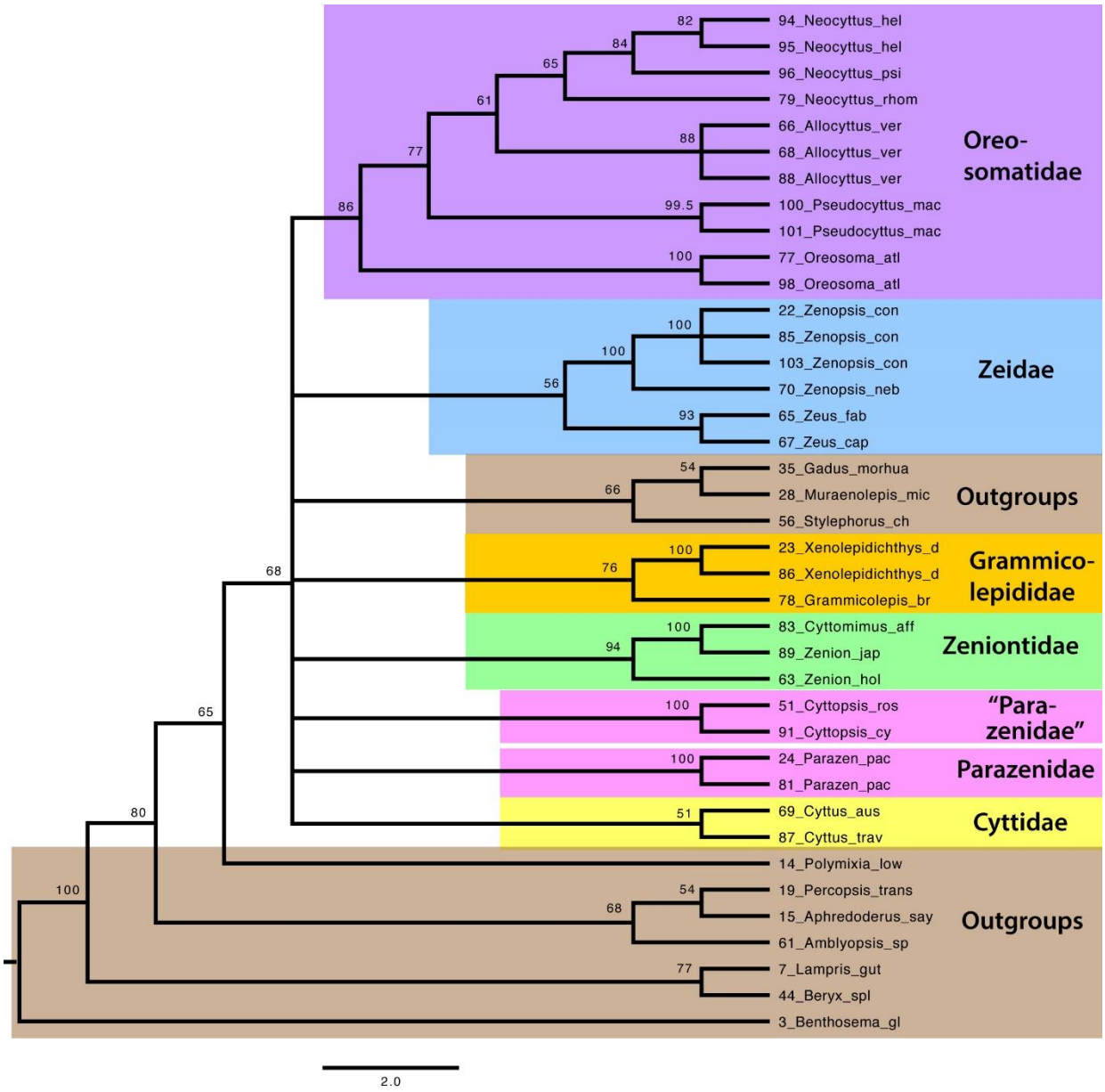

Figure 11. Maximum Parsimony (MP) COI consensus tree generated by PAUP* v4b10 (Swofford, 2002), using unpartitioned sequence data from the coding mitochondrial COI gene. Support values at nodes are bootstrap percentages. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1.

Section 3: Phylogeny based on Combined Coding and Non-Coding Gene Datasets:

Analyzed together under the BI criterion, the combined three-mitochondrial-gene dataset yielded a monophyletic Zeiformes (pp: 1.0; Fig. 12), but with several polytomous relationships at the family level. Though it did not resolve which individual family is most basal, this build criterion showed family Zeniontidae and genus Cyttidae as more basal than the remaining zeiform groups, which were weakly united at a posterior probability of 0.53. However, these other groups formed a polytomy among the genus *Parazen* and two weakly supported clades of sister families (Grammicolepididae + Zeidae) at pp = 0.51 and (Cyttidae + Oreosomatidae) at pp = 0.63. The combined dataset also had higher degrees of resolution at the genus level than any of the individual gene datasets under the BI criterion, with high resolution even between individual specimens of the same taxon. Most notably, the genera within Oreosomatidae exhibited the shortest branch lengths but were well resolved into the following topology at high posterior probabilities for each sister relationship (pp: 1.0): *Oreosoma* + [*Allocyttus* + (*Pseudocyttus* + *Neocyttus*)]. Only *Neocyttus* had a lower posterior probability (e.g., 0.74) and included a single sample with original, possibly erroneous, identification as *Allocyttus niger*.

Under the ML criterion, the three-gene dataset resolved a monophyletic Zeiformes (bt: 100; Fig. 13) with higher resolution than any of the single-gene ML trees. All relationships within the order were resolved down to the genus level, preserving non-monophyletic "Parazenidae" and all other families as monophyletic. The ML tree recovered genus *Cyttopsis* as

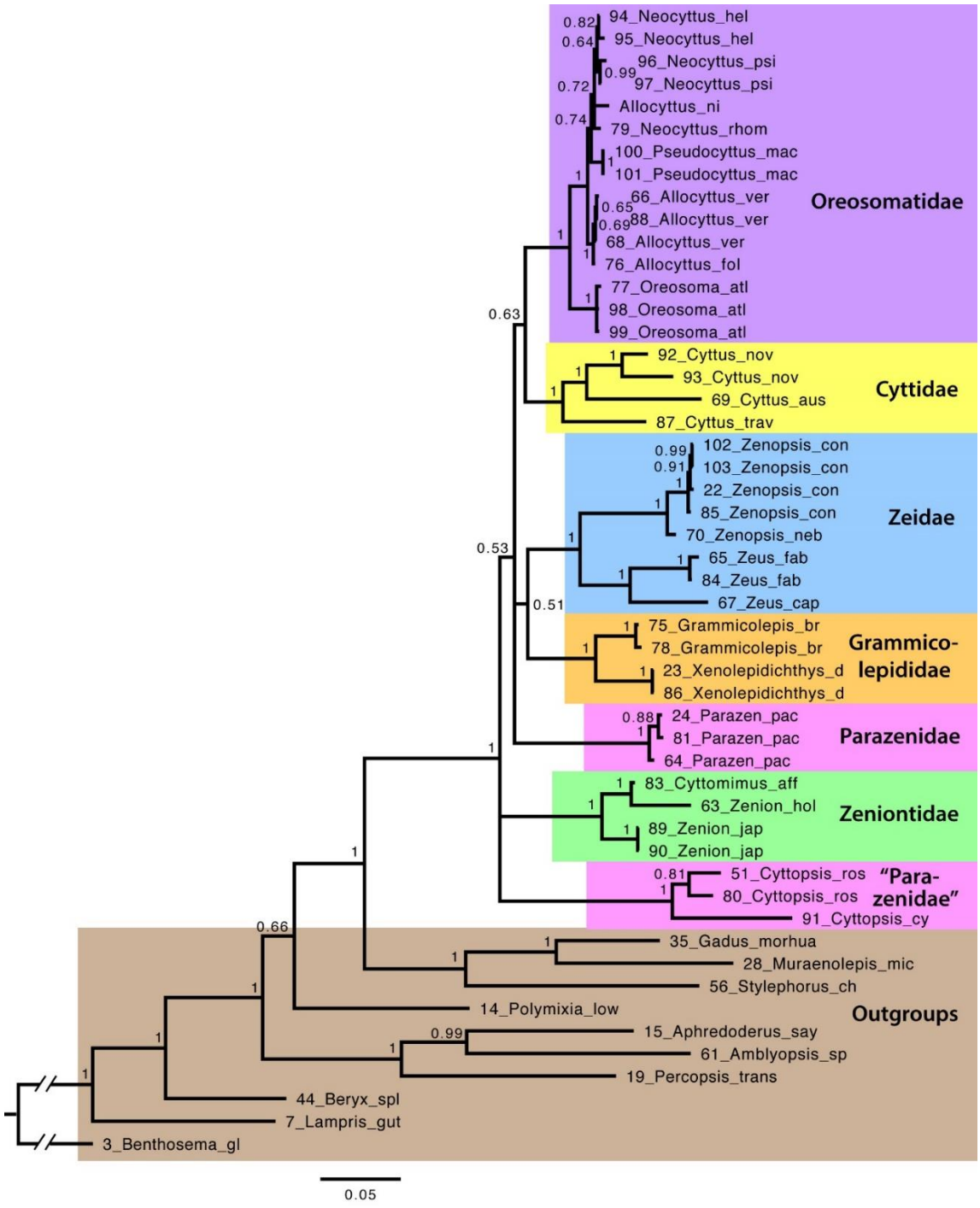

Figure 12. Combined Bayesian Inference (BI) mtDNA consensus tree as generated by MrBayes v3.2 Ronquist et al., 2012) using sequence data for the non-coding mitochondrial 12s and 16s genes and the coding COI gene, partitioned by gene and codon position under the substitution models given in Table 1. Support values at nodes are posterior probabilities. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1.

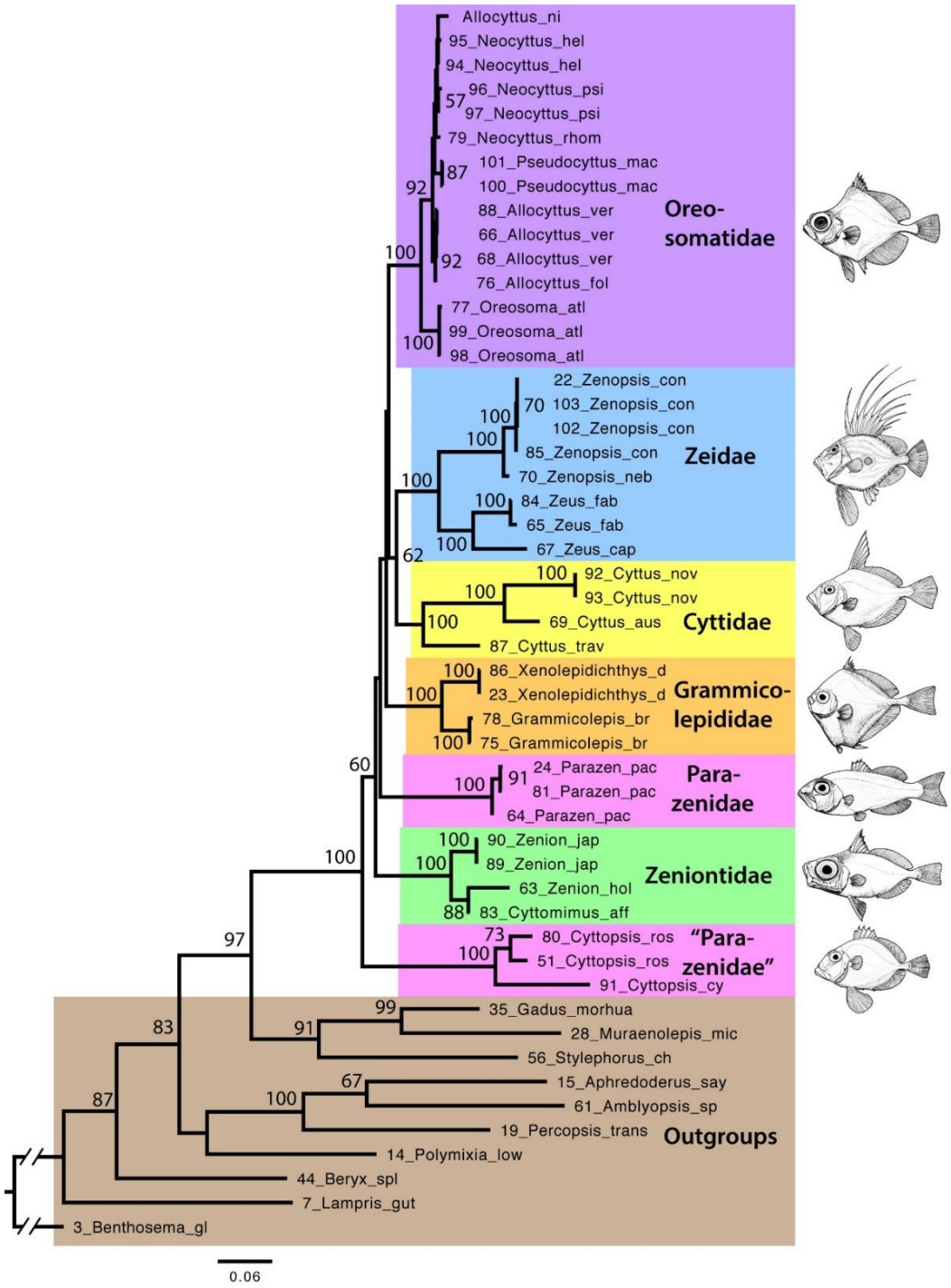

Figure 13. Combined Maximum Likelihood (ML) mtDNA consensus tree as generated by Garli v2.0 (Zwickl, 2006) using sequence data for the non-coding mitochondrial 12s and 16s genes and the coding COI gene, partitioned by gene and codon position under the substitution models given in Table 1. Support values at nodes are bootstrap percentages. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1.

the most basal clade within Zeiformes (bt: 100), and the sister clade (Cyttidae + Zeidae) as most derived. The overall topology of the order was resolved as *Cyttopsis* + Zeniontidae + (*Parazen* + {Grammicolepididae + [Oreosomatidae + (Cyttidae + Zeidae)]}). Genus-level relationships within Oreosomatidae also exhibited high resolution yet short branch lengths and showed a congruent topology to that within Oreosomatidae analyzed under BI.

MP analysis of the combined dataset (Fig. 14) also recovered a monophyletic Zeiformes, with genus *Cyttopsis* as the most basal clade within the group, but did not resolve any other relationships among families. All families were also recovered as monophyletic, as were genera *Cyttopsis* and *Parazen*, and all families except Oreosomatidae exhibit good resolution at the genus level. Within the Oreosomatidae, *Oreosoma* was recovered as the most basal genus in the family, sister to *Pseudocytttus* plus a polytomy. Within the polytomy, all but one sample each of *Allocyttus* and *Neocyttus* were resolved as species-specific clades.

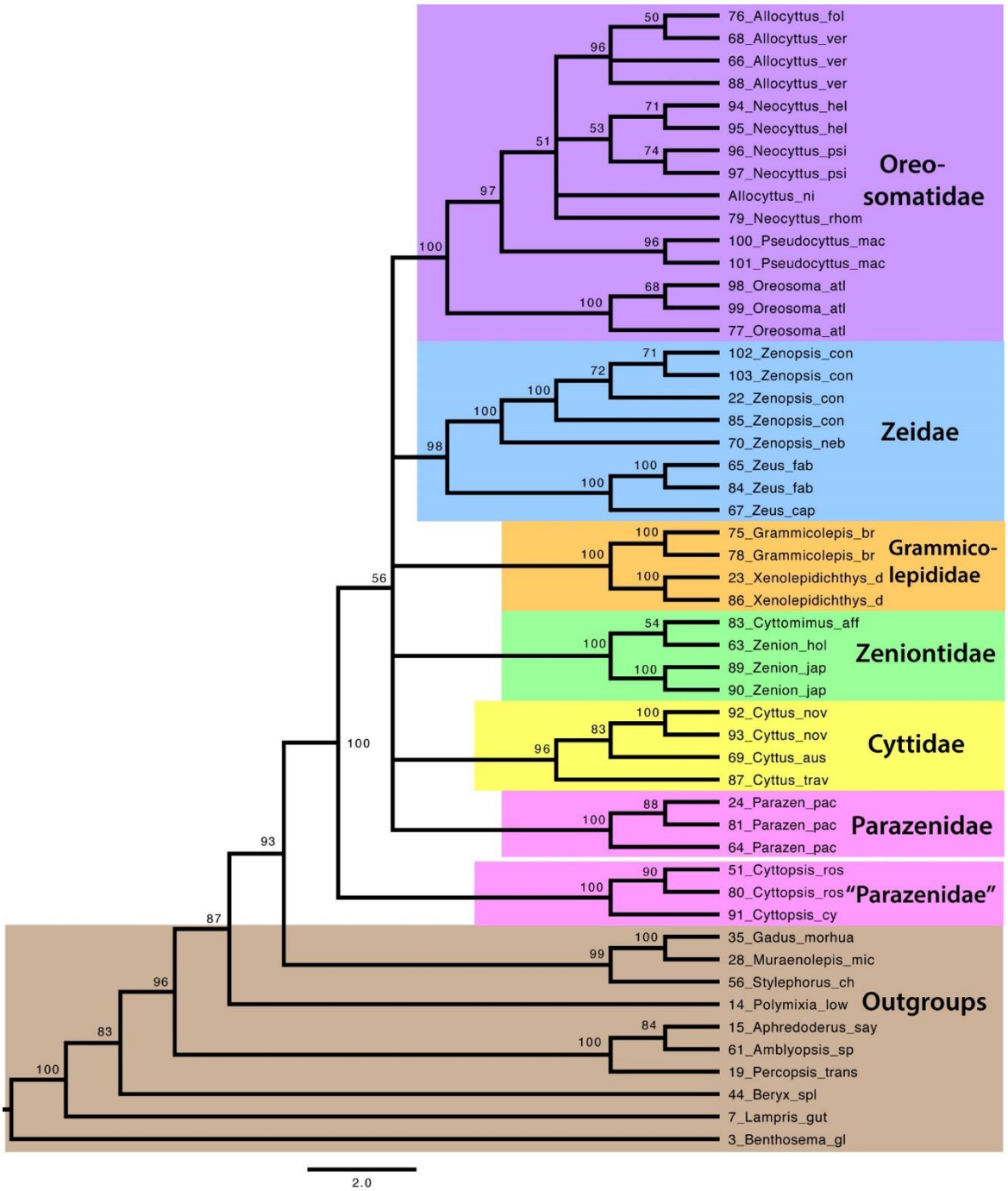

Figure 14. Combined Maximum Parsimony (MP) mtDNA consensus tree as generated by PAUP* v4b10 (Swofford, 2002), using unpartitioned sequence data from the non-coding mitochondrial 12s and 16s genes and the coding mitochondrial COI gene. Support values at nodes are bootstrap percentages. Numbers before scientific names correspond to specimen code numbers listed in Appendix 1.

DISCUSSION

The goals of this study were to a) use mitochondrial sequence data to assess relationships within the Zeiformes, b) to examine whether the use of revised outgroup taxa, different from those used by Tyler et al. (2003) and Tyler and Santini (2005), have contributed to a revised phylogenetic result, and c) to assess whether mitochondrial data resolve zeiform phylogeny differently than the combined nuclear + mitochondrial + morphological data used by Grande et al. (2018).

Use of mitochondrial sequence data to assess zeiform relationships

The three combined-mitochondrial-evidence trees generated by the three different algorithmic methods used in this study differ in their ability to resolve the phylogeny of zeiforms at both the family and genus levels. The rapid rate of mtDNA evolution is particularly well suited to examination of recent evolutionary changes, and thus is capable of resolving relationships among the most closely related taxa at the genus and species level. However, such rapid accumulation of mutations can lead to genetic saturation that masks evolutionary signal between more divergent lineages, which may explain the lower resolution achieved at the family level. The overall trend across all three build criteria was that trees generated from single genes yielded phylogenies with the lowest resolution, and that resolution increased with the addition of genes to the dataset. Inclusion of the coding gene COI was particularly informative with regard to the more derived lineages among zeiforms. The highly conserved first and second codon

positions contribute phylogenetic signal in support of deeper family-level relationships, and the rapidly evolving third codon position is highly informative of relationships among more derived genera. However, across both BI and ML methods, only the full mitochondrial dataset was able to achieve resolution at both family and genus levels, and the BI analyses were unable to completely resolve the family-level topology even using the full 3-mitochondrial-gene dataset.

The combined ML tree (Fig. 13) generated in this study is judged to represent the best fit of the mtDNA data, and thus the "best" hypothesis of zeiform evolution for the purposes of this study, primarily because the trees built using the ML criterion were better able to converge on a resolved phylogeny as more data was added to the character matrix, whereas the trees generated by the BI criterion were unable to resolve the polytomy at the family level within Zeiformes even with the addition of all characters across all three genes in the master matrix. In all, the total-evidence ML tree was based on 408 phylogenetically informative characters out of 1,849 total characters examined.

Comparison with the results of Tyler et al. (2003) and Tyler and Santini (2005)

The studies of Tyler et al. (2003) and Tyler and Santini (2005) were based upon a different hypothesis of the closest relatives of Zeiformes and thus a different selection of outgroups. As a result, analysis of the full three-gene mitochondrial dataset yielded topologies with very different clade composition and relationships from those proposed by Tyler et al. (2003) and Tyler and Santini (2005). The total-mitochondrial-evidence hypotheses consistently recovered a monophyletic Zeiformes, as expected, but consistently split the family "Parazenidae" into two distinct groupings representing genus *Cyttopsis* and genus *Parazen*, differing from Tyler et al. (2003) but in keeping with the findings of Tyler and Santini (2005) (Fig. 15). Tyler et al.

(2003) proposed the family Cyttidae as most basal among the order, a relationship that persisted into the majority-rule MP consensus tree of Tyler and Santini (2005) but was not recovered in any of the total-mitochondrial-evidence trees resulting from this study. The (Cyttidae + Zeidae) sister grouping recovered by all of the total-evidence mtDNA trees was not present in any trees generated by Tyler et al. (2003) or Tyler and Santini (2005), nor was this study's placement of Oreosomatidae as one of the most derived zeiform families recovered in either prior study.

At the genus level, the most obvious difference between the clade composition of the mtDNA total-evidence trees and those of Tyler et al. (2003) and Tyler and Santini (2005) is seen within Oreosomatidae, where both prior hypotheses placed genus *Oreosoma* as more derived than genera *Allocyttus* and *Pseudocyttus*, with *Pseudocyttus* representing the most basal genus within the family. The present mitochondrial analysis instead recovered *Oreosoma* as most basal among oreosomatids, followed by *Pseudocyttus* and *Allocyttus* under both BI and ML criteria; under the MP criterion, genus-level clade composition remained largely unresolved.

While Zeidae were recovered as one of the most derived groups in both this study and in Tyler and Santini (2005), the latter placed Cyttidae, sister to Zeidae in the mtDNA tree, as most basal among Zeiformes. Both of Tyler's studies preserve the association between Oreosomatidae and Cyttidae and hypothesize Oreosomatidae as one of the most basal in the order, whereas the mtDNA evidence presented in this study supports Oreosomatidae as a more derived clade, which we expect to be a result of the differing outgroup taxa chosen for this study based on the shift from the hypothesis of a (Zeiformes + Tetraodontiformes) relationship as per Tyler (2003) and Tyler and Santini (2005) to the currently accepted Zeiformes + (Gadiformes + *Stylephorus*) hypothesis proposed by Grande et al. (2013) among others and elaborated upon by Grande et al.

(2018). The present study shows a stepwise pattern of zeiform evolution that further diverges from Tyler's work; the mtDNA tree supports the following arrangement: *Cyttopsis* + {Zeniontidae + [*Parazen* + (Grammicolepididae + (Oreosomatidae + (Cyttidae + Zeidae)))]}. The morphological evidence presented by Tyler et al. 2003 also supported an arrangement in which each family in ascending order is sister to all other zeiforms, whereas Tyler and Santini (2005) instead recovered several groupings of sister taxa: (*Cyttus* + Oreosomatidae) + ((*Cyttopsis* + *Stethopristes*) + ((Zeniontidae + Parazen) + (Grammicolepididae + Zeidae))).

Tyler and Santini (2005) included several taxa in their morphology-based dataset that inform some relationships that the mtDNA dataset could not examine, because tissues were not available for them, particularly the grouping of *Cyttopsis* with *Stethopristes*, *Macrurocyttus* within Grammicolepididae, and *Capromimus* within Zeniontidae.

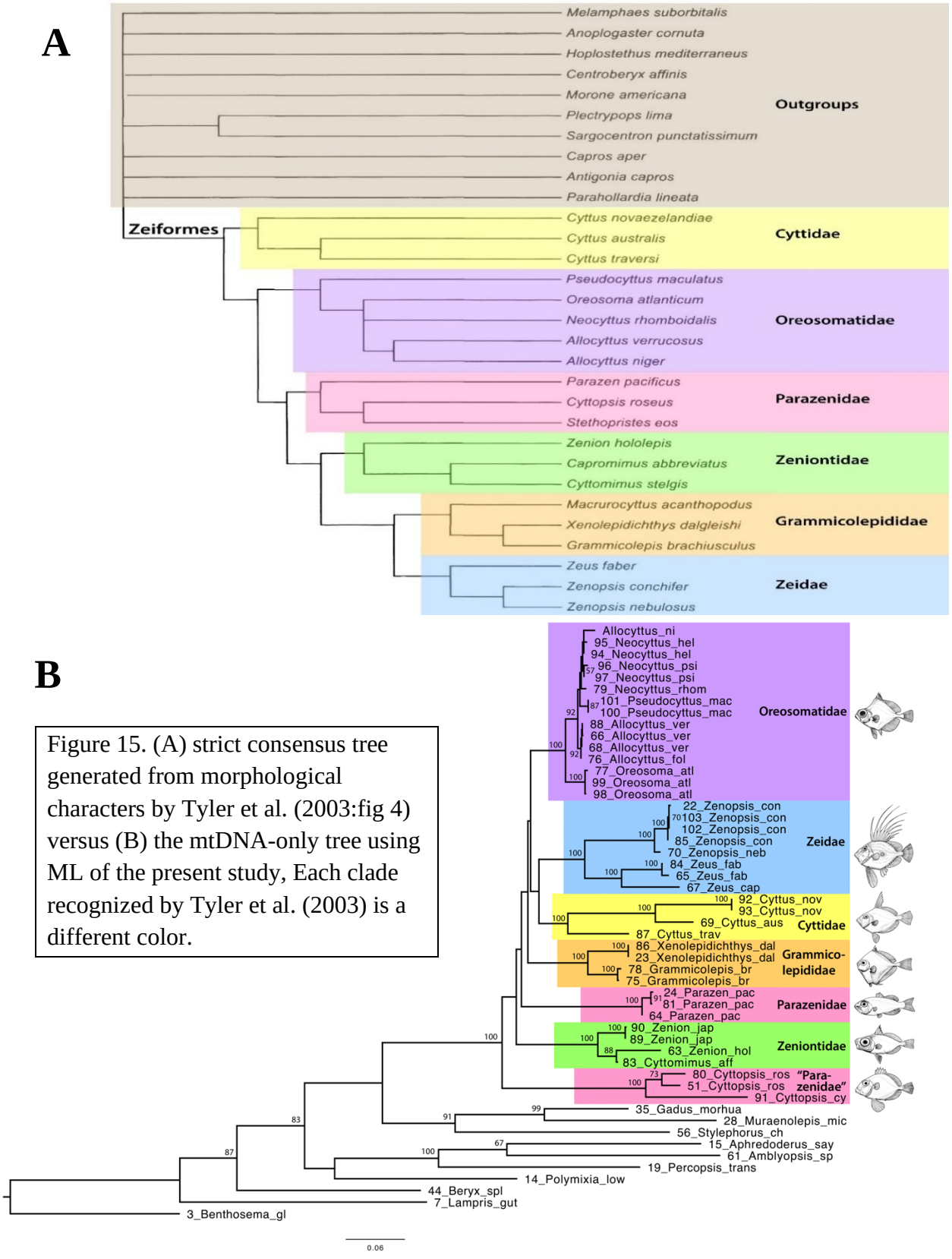

Figure 15. (A) strict consensus tree generated from morphological characters by Tyler et al. (2003:fig 4) versus (B) the mtDNA-only tree using ML of the present study, Each clade recognized by Tyler et al. (2003) is a different color.

Comparison with the results of Grande et al. (2018)

Grande et al. (2018) were also able to place some taxa, for which molecular characters were unavailable, into the total-evidence tree, which included the mtDNA characters of this study, molecular characters from five nDNA genes, morphological characters, and fossil data.

As the mitochondrial-only dataset was included in the molecular dataset under study by Grande et al. (2018), some similarities were expected between the total molecular and morphological evidence tree and the ML mtDNA-only tree (Fig. 16). Both analyses recover a monophyletic Zeiformes as well as congruent arrangements of genera within each united zeiform family. However, where the total-evidence tree resolves the topology of monophyletic Zeiformes into two large clades that further diversify into component families, the present mitochondrial tree instead hypothesizes a stepwise pattern of evolution, hypothesizing each family's divergence individually from the others. Grande et al. (2018) recovered Zeidae as the most basal family, rather than as one of the more derived clades as recovered by the mtDNA-only phylogeny. The total-evidence tree of Grande et al. (2018) recovered a split Zeniontidae (*Zenion* in one clade and *Capromimus* and *Cyttomimus* in another) but a united Parazenidae, as opposed to the split Parazenidae observed across all mtDNA trees generated in this study. Oreosomatidae were recovered as a more derived clade in both trees, but where the mtDNA-only phylogeny recovered clade Oreosomatidae + (Zeidae + Cyttidae), the total-evidence tree instead shows Cyttidae + [Oreosomatidae + (*Capromimus* + *Cyttomimus*)]. Tissue from genus *Capromimus* could not be obtained for the mitochondrial study; a sample was acquired later and added to the combined DNA dataset of Grande et al. (2018) after the benchwork for the present study had already been completed.

Contribution of mtDNA to phylogeny within Zeiformes

The mtDNA sequences contribute most significantly to the overall topology in terms of the resolution achieved at the genus and species levels within Zeiformes. Their influence is especially evident within the four genera and eight extant species of the family Oreosomatidae, which also recovers the shortest branch lengths at the genus and species levels. Both the total-evidence tree and the mtDNA-only tree recover congruent topologies of the component genera within Oreosomatidae, Zeidae, and Cyttidae.

Branch lengths are somewhat representative of relative time between lineage divergences. Longer branch lengths are representative of more substitution events per site, and therefore indicative of longer lengths of evolutionary time between nodes. Shorter branches indicate fewer substitutions, and usually shorter time intervals between divergences. Quickly evolving mitochondrial genes, especially the rapid accumulation of silent mutations within the third codon position of the COI gene, contain more evidence for the evolutionary relationships of younger, more derived lineages, such as those of Oreosomatidae, in which the short branch lengths are indicative of there being a small number of key informative molecular changes that characterize the group and arose over a short period of evolutionary time. Overall, the branch

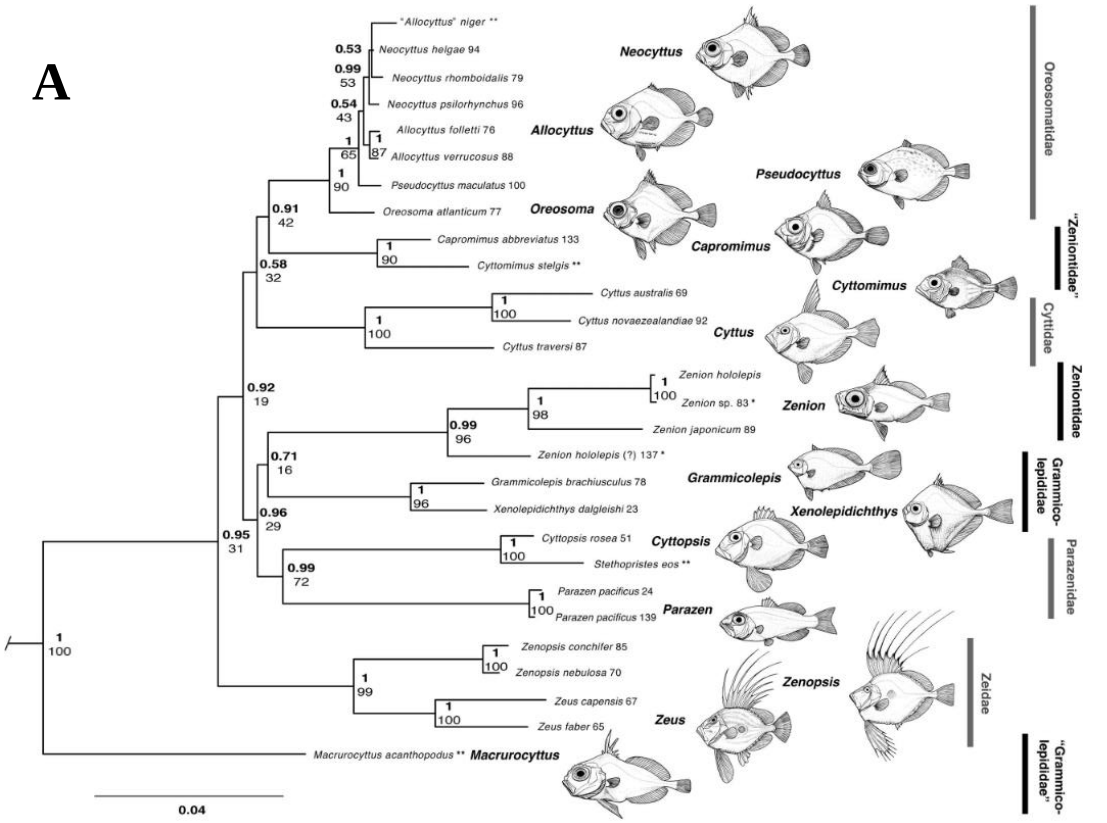

Figure 16. (A) Total molecular and morphological evidence phylogeny of Zeiformes using BI of Grande et al. (2018:fig 7) versus (B) the mtDNA-only phylogeny using ML of the present study.

lengths within Zeiformes are relatively short, even at the family level, and even shorter at the genus level. The few substitutions along short periods of evolutionary time between nodes, giving rise to the 16 genera and 33 extant species within the order, suggest that, once the zeiform family groups formed, diversification within each group occurred rapidly.

This pattern of radiation is especially evident within Oreosomatidae, which have the shortest branch lengths giving rise to the largest number of genera of any zeiform family. Within the Oreosomatidae, the high resolution among genera, achieved by few nucleotide substitutions along very short branch lengths, is likely due to informative molecular characters of the rapidly evolving COI codon position 3. Out of all trees generated across all three build criteria, only those that include the COI codon position 3 dataset in the analyses were able to yield topological resolution within Oreosomatidae. While this topology was not congruently resolved across different tree-building algorithms within this mitochondrial study, the resolution that was achieved specifically resulting from the COI position 3 data partition suggests that only the most rapidly evolving sites within the already rapidly evolving mitochondrial genome contain characters informative enough to resolve clades that formed from rapid diversification events. This phenomenon is also evident between species of *Zenopsis* in family Zeidae, but as Zeidae contains a smaller number of extant species, the effects are less pronounced than in the more diverse Oreosomatidae.

As compared to the analyses containing morphological data conducted by Grande et al. (2018:fig 7), the placement in the present study of *Cyttopsis* as the most basal zeiform rather than Zeidae largely aligns with divergences in body plan observed between these groups. Even in the phylomorphospace analysis of Grande et al. (2018:fig. 9), *Cyttopsis* differs relatively little from

the hypothesized ancestral morphology within the order. The overall body shape observed for genus *Cyttopsis* resembles a more phylogenetically "average" composition, with a shallower body depth, larger mouth, and longer, less oblique jaws as compared to the more derived family Zeidae, in which bodies are more exaggerated in form along several axes. Zeidae exhibit a deeper body shape, smaller mouths as compared to overall size, and more strongly oblique jaws. The shorter branch lengths within the phylogeny lend some support for *Cyttopsis* being one of the earlier-branching lineages within the group.

As more morphologically "average" body shapes may be said to correspond to older lineages within Zeiformes, the ages of zeiform lineages as represented by the branch lengths within the phylogeny lend some support for *Cyttopsis* being one of the older lineages within the group.

The Oreosomatidae are also relatively divergent. The large dorsally situated orbits and shorter, more oblique jaws move the Oreosomatidae away from the "average" body form suggested by *Cyttopsis* as well, suggesting that a more "average fish" body plan like that of basally placed *Cyttopsis* might be closer in shape to the common zeiform ancestor.

This suggests that the resolution achieved within families Oreosomatidae and Zeidae might be linked to phylogenetically informative silent mutations in the COI 3rd codon position, and multiple lines of phylogenetic evidence supporting a more-derived position of clade Oreosomatidae + (Zeidae + Cyttidae) suggest that this group contains some of the youngest lineages within order Zeiformes.

In summary, the evolution of zeiform fishes is complex, and as this study demonstrates, multiple sources of information are needed to disentangle the history of this extraordinary group

of fishes. Zeiforms were a difficult group to classify correctly. All three types of data (nDNA, mtDNA, morphology) made significant contributions to our understanding of the evolution of the zeiform fishes. Molecular data were needed to place them in the correct phylogenetic context (Grande et al., 2013, 2018). Morphological data were able to distinguish most families and genera (Tyler et al., 2003), but their relationships remained obscure. Fossil data indicate that the group evolved in the Late Cretaceous (e.g., Tyler and Santini, 2005; Davesne et al., 2017). Once the individual families evolved, bursts of rapid radiation within each family quickly gave rise to diverse lineages. Nuclear DNA and morphological data provided the backbone and the deeper relationships for the phylogeny (not shown) but mtDNA was needed to illuminate the more recent events giving rise to genera and species. The most rapid bursts of evolutionary radiation could only be revealed with mtDNA.

APPENDIX A

LIST OF TISSUE SAMPLES

AND NCBI GENBANK

SEQUENCE ACCESSION NUMBER

Appendix A. List of tissues and sequences for each of three loci used in this study, with voucher numbers where available. "12S" consists of a large portion of 12s rRNA gene, the complete tRNA-Val gene, and a fragment of the 16S rRNA gene. GenBank accession numbers in bold are new, published under Grande et al. (2018); other accession numbers are from earlier studies. Some new sequences differ from or are longer than sequences available from GenBank. * a shorter sequence by greater than or equal to 30%; ^ tissue has no voucher.

Zeiformes	**Institution**	**code #**	**12S**	**16S**	**COI**
Cyttidae					
Cyttus australis	(AMS I34165)	69	**KY873646**	**KY873690**	EF609340
Cyttus novaezealandiae	(CSIRO GT264 Cyt nov 01)	92	**KY873647**	**KY873691**	**KY873728**
Cyttus novaezealandiae	(CSIRO GT264 Cyt nov 03)	94	**KY873648**	**KY873692**	**KY873729**
Cyttus traversi	(NMV A 25180-001)	87	**KY873649**	**KY873693**	**KY873730**
Grammicolepididae					
Grammicolepis brachiusculus	(ASIZP0915077)	75	**KY873650**	**KY873694**	
Grammicolepis brachiusculus	(NMV A 25107-002)	78	**KY873651**	**KY873695**	**KY873731**
Xenolepidichthys dalgleishi	(R. Hanel)	23	DQ533323	DQ532982	GU804904
Xenolepidichthys dalgleishi	(KUT8348)	86	**KY873665**	**KY873711**	GU805001
Oreosomatidae					
Allocyttus folletti	(SIO.097-120)	76	**KY873637**	JX121802	GU440211
Allocyttus verrucosus	(SAIAB87336)	66	**KY873638**	**KY873681**	DQ108067
Allocyttus verrucosus	(SAIAB87358)	68	**KY873639**	**KY873682**	DQ108068
Allocyttus verrucosus	(NMV A 25102-001)	88	**KY873640**	**KY873683**	DQ108077
Neocyttus helgae	(CSIRO H 5366-01)	94	**KY873652**	**KY873696**	EU148264
Neocyttus helgae	(CSIRO H 5792-04)	95	**KY873653**	**KY873697**	DQ108080
Neocyttus psilorhynchus	(CSIRO H 4509-01)	96	**KY873654**	**KY873698**	EF609417
Neocyttus psilorhynchus	(CSIRO H 4510-01)	97	**KY873655**	**KY873699**	—
Neocyttus rhomboidalis	(NMV A 25149-005)	79	**KY873656**	**KY873700**	DQ108094
Oreosoma atlanticum	(NMV A 21940)	77	**KY873657**	**KY873701**	**KY873732**
Oreosoma atlanticum	(CSIRO H 4430-01)	98	**KY873658**	**KY873702**	DQ108069
Oreosoma atlanticum	(CSIRO H 5386-03)	99	**KY873659**	**KY873703**	DQ108071
Pseudocyttus maculatus	(CSIRO H 3972-01)	100	**KY873663**	**KY873709**	DQ108085
Pseudocyttus maculatus	(CSIRO H 5348-01)	101	**KY873664**	**KY873710**	DQ108086
Parazenidae		91	—	**KY873688**	JQ681446
Cyttopsis cypho	(CSIRO H-2423-01)	51	JX121827	JX121801	JQ774522
Cyttopsis rosea	(KUT8315)	80	**KY873645**	**KY873689**	JQ774523
Cyttopsis rosea	(NMV A 25169-005)	24	JX121808	**KY873704**	AP004433
Parazen pacificus	(FMNH 120982)	64	**KY873660**	**KY873705**	GU804929
Parazen pacificus	(SAIAB82404)	81	**KY873661**	**KY873706**	—

Parazen pacificus	(NMV A 25289-001)				
Zeidae					
Zenopsis conchifer	(AMNH uncat)	22	JX121831	JX121803	KC016043
Zenopsis conchifer	(KUT1074)	85	**KY873670**	**KY873716**	KC016044
Zenopsis conchifer	(FLMNH170556-0214)	102	—	—	—
Zenopsis conchifer	(FLMNH170556-0215)	103	—	—	—
Zenopsis nebulosus	(AMS I34166)	70	**KY873673**	**KY873719**	AP002942
Zeus capensis	(SAIAB87351)	67	**KY873674**	**KY873720**	JF494803
Zeus faber	(SAIAB84189)	65	**KY873675**	**KY873721**	KC501893
Zeus faber	(AMS I37682)	84	**KY873676**	**KY873722**	KC501910
Zeniontidae					
Zenion hololepis	(SAIAB82155)	63	**KY873666**	**KY873712**	**JF718834**
Zenion sp. ‘‘Cyttomimus affinis’’	(ASIZP 0910704)	83	**KY873644**	**KY873687**	**KY873727**
Zenion japonicum	(CSIRO H 7136-19 GT5810)	89	**KY873668**	**KY873714**	**KY873735**
Zenion japonicum	(CSIRO H 7136-19 GT5811)	90	**KY873669**	**KY873715**	**KY873736**
Gadiformes					
Gadus morhua		35	KY873635	JX121817	KC015385
Muraenolepis microps		28	JX121838 R.Hanel	JX121812 R.Hanel	EU326376 *
Stylephoriformes					
Stylephorus chordatus	(KU5228)	56	NC 009948	MCZ 165920	AB280689
Percopsiformes					
Percopsis transmontana		19	**KY873632** KU 29776	**KY873678** KU 29776	AP002928
Aphredoderus sayanus		15	DQ533156 UAIC 14127.03	DQ027910 UAIC 14127.03	JN024807
Amblyopsis spelaea		61	JX121823 WC1	JX121797 WC1	KY873723 WC1
Polymixiiformes					
Polymixia lowei		14	KY873630 AMNH uncat.	AY538966 AMNH uncat.	KC015824 AMNH uncat.
Beryciformes					
Beryx splendens		44	DQ533161	DQ027918	—
Lampriformes					
Lampris guttatus		7	DQ533220	DQ027908	—
Myctophiformes					
Benthosema glaciale		3	DQ533160	DQ532843	—

APPENDIX B

PRIMERS AND PRIMER SEQUENCES USED FOR PCR AMPLIFICATION

Appendix B. List of primers, sources and annealing temperatures for the newly obtained sequences used in this study.

Primer name	Primer sequence	Annealing Temp (°C)	Special Designators
12S	(Titus, 1992; Feller and Hedges, 1998)		
12SL13-L	5'-TTAGAAGAGGCAAGTCGTAACATGGTA-3'	52°	
TitusI-H	5'-GGTGGCTGCTTTTAGGCC-3'	52°	
16S	(Kocher et al., 1989; Palumbi, 1996)		
16Sar-L	5'-CGCCTGTTTATCAAAAACAT-3'	52°	
16Sar-H	5'-CCGGTCTGAACTCAGATCACGT-3'	52°	
COI	(Ward et al., 2005; Ivanova et al., 2007)		
FF2d	5'-TTCTCCACCAACCACAARGAYATYGG-3'	52°	Y wobble (C+T)
FR1d	5'-CACCTCAGGGTGTCCGAARAAYCARAA-3'	52°	R wobble (A+G)
LepF1_t1	5'-ATTCAACCAATCATAAAGATATTGG-3'	50°	
LepR1_t1	5'-TAAACTTCTGGATGTCCAAAAAATCA-3'	54°	

APPENDIX C

PCR CONDITIONS

BY LOCUS

Appendix C. PCR conditions by locus under study.

Locus	Reagent	Volume	PCR temp (°C)	Time (min)	Number of Cycles
12S	Tris-HCl buffer	1 ul	95°	1:30	1
	MgCl2	0.6 ul	95°	0:20	
	TAQ	0.05 ul	52°	0:20	30
	H2O	7.55 ul	72°	0:55	
	dNTPs	0.3 ul	72°	3:00	1
	primer 12SL13-L	0.2 ul	12°	Hold	
	primer TitusI-H	0.2 ul			
	DNA template	0.2 ul			
16S	Tris-HCl buffer	1 ul	95°	1:30	1
	MgCl2	0.6 ul	95°	0:20	
	TAQ	0.05 ul	52°	0:20	30
	H2O	7.55 ul	72°	0:55	
	dNTPs	0.3 ul	72°	3:00	1
	primer 16Sar-L	0.2 ul	12°	Hold	
	primer 16Sar-H	0.2 ul			
	DNA template	0.2 ul			
COI	Tris-HCl buffer	1 ul	94°	2:00	1
	MgCl2	1 ul	94°	0:30	
	TAQ	0.05 ul	52°	0:40	35
	H2O	6.25 ul	72°	1:00	
	dNTPs	0.3 ul	72°	10:00	1
	primer FF2d	0.2 ul	4°	Hold	
	primer FR1d	0.2 ul			
	Tris-HCl buffer	1 ul	94°	1:00	1
	MgCl2	1 ul	94°	0:30	
	TAQ	0.05 ul	50°	0:40	5
	H2O	6.25 ul	72°	1:00	
	dNTPs	0.3 ul	94°	0:30	
	primer LepF1_t1	0.2 ul	54°	0:40	35
	primer LepR1_t1	0.2 ul	72°	1:00	
	DNA template	0.5 ul	72°	10:00	1
			4°	Hold	

APPENDIX D

NEXUS CODE COMMANDS

FOR EACH TREE BUILDER SOFTWARE

USED IN THIS STUDY

Appendix D. NEXUS code commands for each tree builder software used in the present study.

Bayesian Inference - MrBayes v3.2 (Ronquist et al., 2012)

```
#NEXUS

begin taxa;
dimensions ntax=51;
taxlabels

begin characters;
dimensions nchar=[total number of nucleotides across all loci under analysis];
format datatype=dna missing=? gap=- interleave=yes;
matrix

begin mrbayes;
log start file=[save file name].txt;
charset 1=[# of nucleotides in 12S sequence];
charset 2=[# of nucleotides in 16S sequence];
charset 3-codon pos1=[# of nucleotides in COI sequence]\3;
charset 4-codon pos2=[# of nucleotides in COI sequence]\3;
charset 5-codon pos3=[# of nucleotides in COI sequence]\3;
partition favored=5:12S,16S,COI_1st,COI_2nd, COI_3rd;
set partition=favored;
lset applyto=(1,2,5) nst=6 rates=gamma;
lset applyto=(3,4) nst=2 rates=gamma;
prset applyto=(all) brlenspr=unconstrained:exponential(100.00);
unlink shape=(all) pinvar=(all) statefreq=(all) revmat=(all);
mcmc ngen=10000000 nchains=4 samplefreq=1000 printfreq=1000 savebrlens=yes;
sump burnin=25;
sumt burnin=25;
log stop;
```

Maximum Likelihood - GARLI v2.1 (Bazinet et al., 2014)

```
[general]
datafname = [NEXUS file name].nex
constraintfile = none
streefname = stepwise
attachmentspertaxon = 50
ofprefix = [NEXUS file name]
randseed = -1
availablememory = 512
logevery = 10
saveevery = 100
refinestart = 1
outputeachbettertopology = 0
outputcurrentbesttopology = 0
enforcetermconditions = 1
genthreshfortopoterm = 5000
scorethreshforterm = 0.001
significanttopochange = 0.01
outputphyliptree = 0
outputmostlyuselessfiles = 0
writecheckpoints = 0
restart = 0
outgroup = 1
resampleproportion = 1.0
inferinternalstateprobs = 0
outputsitelikelihoods = 0
optimizeinputonly = 0
collapsebranches = 1

searchreps = 1
bootstrapreps = 100
linkmodels = 0
subsetspecificrates = 1

[model1 - GTR + G]
datatype = nucleotide
ratematrix = 6rate
statefrequencies = estimate
ratehetmodel = gamma
numratecats = 4
invariantsites = none

[model2 - GTR + G]
```

```
datatype = nucleotide
ratematrix = 6rate
statefrequencies = estimate
ratehetmodel = gamma
numratecats = 4
invariantsites = none

[model3 - HKY]
datatype = nucleotide
ratematrix = 2rate
statefrequencies = estimate
ratehetmodel = gamma
numratecats = 4
invariantsites = estimate

[model4 - HKY]
datatype = nucleotide
ratematrix = 2rate
statefrequencies = estimate
ratehetmodel = gamma
numratecats = 4
invariantsites = none

[model5 - GTR + G]
datatype = nucleotide
ratematrix = 6rate
statefrequencies = estimate
ratehetmodel = gamma
numratecats = 4
invariantsites = none


[master]
nindivs = 4
holdover = 1
selectionintensity = 0.5
holdoverpenalty = 0
stopgen = 10000000
stoptime = 10000000

startoptprec = 0.5
minoptprec = 0.01
numberofprecreductions = 10
treerejectionthreshold = 50.0
```

topoweight = 1.0
modweight = 0.05
brlenweight = 0.2
randnniweight = 0.1
randsprweight = 0.3
limsprweight = 0.6
intervallength = 100
intervalstostore = 5
limsprrange = 6
meanbrlenmuts = 5
gammashapebrlen = 1000
gammashapemodel = 1000
uniqueswapbias = 0.1
distanceswapbias = 1.0

Maximum Parsimony - PAUP* 4.0 (Swofford, 2002)

#NEXUS

begin taxa;
dimensions ntax=51;
taxlabels

begin characters;
dimensions nchar=[total number of nucleotides across all loci under analysis];
format datatype=dna missing=? gap=- interleave=yes;
matrix

begin SETS;
CHARSET 12S = [# of nucleotides in 12S sequence;
CHARSET 16S = [# of nucleotides in 16S sequence];
CHARSET COI = [# of nucleotides in COQ sequence];end;

begin paup;
outgroup [root taxon name];
set root = outgroup outroot = monophyl;
log file = [save file name].log;
set criterion = parsimony;
set maxtrees=100 increase=auto;
bootstrap nreps=100 search=heuristic /addseq=random;
[savetrees from=1 to=1 savebootp=nodelabels file=filename.tre; to save bootstrap tree]

REFERENCE LIST

Baciu, D. S., A. F. Bannikov, and J. C. Tyler. 2005. Revision of the fossil fishes of the family Zeidae (Zeiformes). Bolletino del Museo Civico di Storia Naturale di Verona 29:95-128.

Bazinet, A. L., D. J. Zwickl, and M. P. Cummings. 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Systematic Biology 63(5):812-818.

Betancur, R., R. E. Broughton, E. O. Wiley, K. Carpenter, J. A. López, C. H. Li, N. I. Holcroft, D. Arcila, M. Sanciangco, J. C. Cureton II, F. F. Zhang, T. Buser, M. A. Campbell, J. A. Ballesteros, A. Roa-Varón, S. Willis, W. C. Borden, T. Rowley, P. C. Reneau, D. J. Hough, G. Q. Lu, T. Grande, G. Arratia, and G. Ortí. 2013. The tree of life and a new classification of bony fishes. PLOS Currents Tree of Life.1-41.

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

Davesne, D., G. Carnevale, and M. Friedman. 2017. *Bajaichthys elegans* from the Eocene of Bolca (Italy) and the overlooked morphological diversity of Zeiformes (Teletaostei, Acanthomorpha). Palaeontology doi: 10.1111/pala.12280: 1-14.

Feller, A. E., and S. B. Hedges. 1998. Molecular evidence for the early history of living amphibians. Molecular Phylogenetics and Evolution 9:509-516.

Feng, D. F. and R. F. Doolittle. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. Journal of Molecular Evolution 25:351–360.

Gissi, C., F. Iannelli, and G. Pesole. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301-320.

Grande, T., W. C. Borden, and W. L. Smith. 2013. Limits and relationships of Paracanthopterygii: A molecular framework for evaluating past morphological hypotheses, p. 385–418. *In:* Mesozoic Fishes 5—Global Diversity and Evolution. G. Arratia, H.-P. Schultze, and M. V. H. Wilson (eds.). Verlag Dr. Friedrich Pfeil, München.

Grande, T., W. C. Borden, M. V. H. Wilson, and L. Scarpitta. 2018. Phylogenetic relationships among fishes in the order Zeiformes based on molecular and morphological data. Copeia 106(1):20-48.

Greenwood, P. H., D. E. Rosen, S. H. Weitzman, and G. S. Meyers. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bulletin of the AMNH 131(4):339-456.

Hagström, E., C. Freyer, B. J. Battersby, J. B. Stewart, and N.-G. Larsson. 2014. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. Nucleic Acids Research 42(2):1111-1116.

Heemstra, P. C. 1980. A revision of the zeid fishes (Zeiformes: Zeidae) of South Africa. Ichthyological Bulletin of the J. L. B. Smith Institute of Ichthyology 41:1-14.

Heemstra, P. C. 1995. Additions and corrections for the 1995 impression. p. v-xv. *In:* M.M. Smith and P.C. Heemstra (eds.) Revised Edition of Smiths' Sea Fishes. Springer-Verlag, Berlin.

Ivanova, N. V., T. S. Zemlak, R. H. Hanner, and P. D. N. Hebert. 2007. Universal primer cocktails for fish DNA barcoding. Molecular Ecology Notes 7:544-548.

James, J. E., G. Piganeu, and A. Eyre-Walker. 2015. The rate of adaptive evolution in animal mitochondria. Molecular Ecology 25:67-78.

Jia, W., and P. G. Higgs. 2008. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. Molecular Biology and Evolution 25(2):339-351.

Johnson, G. D. and C. Patterson. 1993. Percomorph phylogeny: A survey and a new proposal. Bulletin of Marine Science 52:554–626.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, and A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647-1649.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences of the United States of America 86:6196-6200.

Ladoukakis, E. D. and E. Zouros. 2017. Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. Journal of Biological Research-Thessaloniki 24(4):1-7.

Li, C., G. Ortí, G. Zhang, and G. Lu. 2007. A practical approach to phylogenomics: The phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evolutionary Biology 7:44.

Linnæus, C. 1758. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. 1-824. Holmiæ. (Salvius).

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

Miya, M., N. I. Holcroft, T. P. Satoh, M. Yamaguchi, M. Nishida, and E. O. Wiley. 2007. Mitochondrial genome and a nuclear gene indicate a novel phylogenetic position of deep-sea tube-eye fish (Stylephoridae). Ichthyological Research 54:323–332.

Near, T. J., R. I. Eytan, A. Dornburg, K. L. Kuhn, J. A. Moore, M. P. Davis, P. C. Wainwright, M. Friedman, and W. L. Smith. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the National Academy Sciences USA 109(34):13698–13703.

Nelson, J. S., T. C. Grande, and M. V. H. Wilson. 2016. Fishes of the World, 5th Edition. John Wiley & Sons, Inc., New York.

Palumbi, S. R. 1996. Nucleic acids II: the polymerase chain reaction. – *In:* Hillis, D. M., Moritz, C., & Mable, B. K. (eds.): Molecular Systematics, 2nd ed. 205-247; Sunderland, MA (Sinauer).

Philippe, H., H. Brinkmann, D.V. Lavrov, D. Timothy, J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLOS Biology 9(3): e1000602.

Posada, D. 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25(7):1253-1256.

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

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3):539-42.

Rosen, D. E. 1982. Teleostean interrelationships, morphological function and evolutionary inference. American Zoologist 22(2): 261-273.

Rosen, D. E. 1984. Zeiforms as primitive plectognath fishes. American Museum Novitates 2782: 1-45.

Scott, W. B., and M. G. Scott. 1988. Atlantic fishes of Canada. Canadian Bulletin of Fisheries and Aquatic Sciences 219.

Swofford, D. A. 2002. PAUP* 4.b10, Phylogenetic Analysis Using Parsimony. Sinauer Associates, Sunderland, Massachusetts.

Titus, T. A. 1992. A phylogenetic analysis of the Desmognathinae (Caudata: Plethodontidae): Evolutionary patterns inferred from mitochondrial DNA sequences. 122 pp., Unpub. Ph.D. Diss., Univ. Kansas, USA.

Tyler, J. C., P. Bronzi, and A. Ghiandoni. 2000. The Cretaceous fishes of Nardò 11º. A new genus and species of Zeiformes, *Cretazeus rinaldii*, the earliest record for the order. Bolletino del Museo Civico di Storia Naturale di Verona 24:11-28.

Tyler, J. C., B. O'Toole, and R. Winterbottom. 2003. Phylogeny of the genera and families of zeiform fishes, with comments on their relationships with tetraodontiforms and caproids. Smithsonian Contributions to Zoology 618:1-110.

Tyler, J. C., and F. Santini. 2005. A phylogeny of the fossil and extant zeiform-like fishes, Upper Cretaceous to Recent, with comments on the putative zeomorph clade (Acanthomorpha). Zoologica Scripta 34:157-175.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P D. N. Hebert. 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B 360:1847-1857.

Wiley, E. O., G. D. Johnson, and W.W. Dimmick. 2000. The interrelationships of acanthomorph fishes: A total evidence approach using molecular and morphological data. Biochemical Systematics and Ecology 28:319-350.

Yakes, M., l F., and B. Van Houten. 1997. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proceedings of the National Academy of Sciences of the United States of America 94:514-519.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. diss., Univ. Texas Austin. [http://code.google.com/p/garli/].

VITA

Lindsay A. Scarpitta graduated from Merrimack College in North Andover, Massachussetts in May 2011 with a B.S. in Biology, concentrating in Molecular Biology and Bioinformatics. As an undergraduate she studied mate choice in the sailfin molly (*Poecilia latipinna*) using microsatellite sequence elements to determine inheritance patterns in the Genetics lab of Dr. Janine Leblanc-Straceski. Her undergraduate research laid the foundation of a collaborative research initiative between the Genetics and Ecology laboratories and developed the protocols by which future undergraduate genetic research would be conducted for the ongoing project. She also worked as an undergraduate teaching assistant in the Biology Department under Dr. Leblanc-Straceski, where she assisted in facilitating Introduction to Biology practical laboratories for incoming students.

Upon graduating from Merrimack College, Scarpitta moved to the Midwest to begin her Master of Science in Biology at Loyola University Chicago in 2011, focusing on evolution and phylogenetics in the Zeiformes order of paracanthopterygian fishes under Dr. Terry Grande. She obtained a Lerner-Gray research grant from the American Museum of Natural History to carry out her work, and completed this thesis as part of a larger study conducted by the Grande Lab and published in the journal *Copeia* (now called *Ichthyology & Herpetology*) in 2018. During her time at Loyola, Scarpitta enjoyed teaching Evolution BIOL 319 under Dr. Grande as a teaching assistant, and Introduction to Biology Laboratory BIOL 111 and 112 as Instructor of Record.

In 2015, Scarpitta accepted a full-time position as a Research Specialist under Dr. Pieter Faber at the University of Chicago Functional Genomics Facility, where she now manages a team facilitating sample preparation and processing for client-submitted Next-Generation Sequencing experiments. She credits her time at Loyola for developing the scientific knowledge and experimental rigor that have earned her two promotions at the Genomics Facility since beginning her tenure in the lab.