

Systematics of Clupeomorpha (Osteichthyes:Teleostei)
with methodological considerations for morphological phylogenetics

by

Oksana V Vernygora

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Systematics and Evolution

Department of Biological Sciences
University of Alberta

© Oksana V Vernygora, 2020

Abstract

Here I conduct a systematic revision of Clupeomorpha, a group of bony fishes that includes over 500 living and extinct species of herrings, sardines, anchovies, and their relatives. Despite a long history of research on clupeomorph fishes, evolutionary relationships within the group remain poorly understood. To provide an updated phylogeny and classification for Clupeomorpha, I begin by addressing methodological issues pertinent to morphological phylogenetics and species delimitation in palaeontological and neontological studies and conclude by using a combined evidence approach to analyze morphological, molecular, and concatenated data sets employing multiple methods of phylogenetic inference. In the chapters discussing methodological issues in morphological phylogenetics, I assess the performance of widely used probabilistic algorithms of phylogenetic inference (Bayesian inference and maximum likelihood), and also develop and test a new model of discrete morphological trait evolution (FreqMorph) that uses empirical character state frequencies. Results of simulation studies indicate that Bayesian inference applications are the most consistent and robust to variations in taxonomic sampling and amount of missing data. Additionally, using empirical character state frequencies, as implemented in FreqMorph, can improve accuracy of phylogenetic reconstructions in cases when the number of taxa is large and branch lengths are of particular importance. Delimiting taxonomic units is another pertinent issue in systematic studies. In a case study of the extant genus *Alosa*, I show that an integrative approach combining genetic and morphological data provides a reliable framework for delimiting closely related and morphologically similar species. In palaeontological studies, species delimitation relies heavily on morphological data. In a case study of the extinct genus †*Armigatus*, I conduct a comparative examination of multiple specimens to seek morphological traits that allow more reliable species

delimitation within the genus. Lastly, I bring together morphological and molecular data to perform a combined evidence phylogenetic analysis of the Clupeomorpha. Synthesis of the multiple phylogenetic results suggests that Clupeomorpha can be subdivided into three orders, Denticipitiformes, †Ellimmichthyiformes, and Clupeiformes, with Denticipitiformes being the basalmost clupeomorph lineage. Interrelationships within †Ellimmichthyiformes and Clupeiformes are overall consistent with the current classification. I assign a family rank to previously recognized clupeid subfamilies that have been consistently recovered as monophyletic groups, Clupeidae (new usage), Dorosomatidae, and Alosidae, and provide a list of morphological features that can be used to diagnose members of these groups.

Preface

This thesis contains my original work as well as results of collaborative research that has been published or accepted for publication in peer-reviewed journals.

A version of Chapter 2 of the thesis was published as Vernygora, O.V., T. R. Simões, and E. O. Campbell. 2020. Evaluating the performance of probabilistic algorithms for phylogenetic analysis of big morphological datasets: a simulation study. *Systematic Biology*: 10.1093/sysbio/syaa020. I conducted all simulations for the study and performed half of the phylogenetic analyses. Drs. Campbell and Simões contributed toward designing the study, manuscript writing, and running the other half of the analyses.

A version of Chapter 4 was published as Vernygora, O.V., C. S. Davis, A. M. Murray, and F. A. H. Sperling. 2018. Delimitation of *Alosa* species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches. *Journal of Fish Biology* 93(6):1216–1228. I was responsible for data collection and analysis as well as drafting the manuscript. Drs. Murray and Sperling provided guidance in the project and access to laboratory equipment and comparative material. Dr. Davis performed high throughput sequencing of samples. All authors contributed to writing the manuscript.

A version of Chapter 5 was accepted for publication as Vernygora, O. V. and A. M. Murray. 2020. Morphological variation among the species of †*Armigatus* (Teleostei, Clupeomorpha, Ellimmichthyiformes) and new material of †*Armigatus alticorpus* from the Late Cretaceous (Cenomanian) of Hakel, Lebanon. *Cretaceous Research*. In Press. I was responsible for examination and description of study material. Dr. Murray provided supervisory and editorial contribution to manuscript composition.

To my family, for their endless love and support.

Acknowledgements

I express my endless thanks and gratitude to everyone who has supported and encouraged me throughout these years. It has been an exciting journey that would be impossible without all the wonderful people that have been a part of it!

First, I would like to thank my supervisor, Alison Murray for her mentorship and for providing me with an opportunity to pursue my research interests. Over the past seven years that I have been a member of your lab, I learnt so much about palaeoichthyology and was very fortunate to work with you on amazing fossils of not just clupeomorphs but also sturgeons and dercetids; these studies are unfortunately beyond the scope of this thesis but they are a prominent experience in my scientific career and I am very grateful for that.

I am sincerely thankful to the members of my supervisory committee, Eva Koppelhus and Felix Sperling. Your guidance and feedback on my research helped me stay on track and improve my science throughout these years and your support and encouragement helped me power through even the most overwhelming times.

Thank you to my examining committee, Jocelyn Hall and Jason Anderson, for your feedback and input that help me develop as a scientist. I am also thankful to John Acorn and Maya Evenden for serving on my candidacy committee.

Many people have contributed to this thesis work and my academic progress over the years. I would also like to express my gratitude to Philip Currie, Rob Holmes, Michael Caldwell, John Acorn, Corwin Sullivan, Heather Proctor, Al Lindoe, Clive Coy, and Corey Davis for providing an inspiring and challenging academic environment that promoted my learning and curiosity.

I would also like to thank everyone who collected and prepared specimens that I studied and described for my research project as well as everyone who provided access to the museum collections, specimens, and equipment, particularly, Don Brinkman, Brandon Strilisky, Kieran Shepherd, Margaret Currie, and Barbara Brown.

I have been extremely fortunate to work with a great group of brilliant fellow researchers who have been an excellent influence and great inspiration. I would like to thank my paleo grad fellows Thodoris Argyriou, Paulina Huidobro, Tiago Simões, Aaron LeBlanc, Alessandro Palci, Julien Divay, Ilaria Paparella, Angelica Torices, Stephanie Blais, Tetsuto Miyashita, Aaron Van Der Reest, Meghan Dueck, Victoria Arbour, Hallie Street, Betsy Kruk, Michelle Campbell, Katherine Bramble, Takuya Konishi, Gavin Bradley, Rebekah Vice, Michael Hudgins, Sydney Mohr, Catie Strong, Yan-yin Wong, Luthfur Rahman, Sinjini Sinha, Jasdeep Kaur, Matthew Rhodes, Mark Powers, Aaron Dyer, Greg Funston, Sam Hamilton, Annie McIntosh, Scott Persons, Michael Burns, Javier Luque, and Fernando Garberoglio.

I owe my thanks to everyone in the Sperling Lab and affiliates. Thank you, Erin Campbell, Janet Sperling, Victor Shegelski, Brittany Wingert, Tyler Nelson, Zac MacDonald, Federico Riva, Rowan French, Bryan Brunet, Giovanni Fagua, Pasan Lebunasin Arachchige, Stephen Trevoy, Kyle Snape, and Marnie Wright, you have been a great influence on me and pushed the limits of my knowledge to a great extent.

And more than anything, I owe thanks to my family for they have always been there for me with their unconditional love and support! No matter how far apart we are, you are always with me, my strongest supporters and critics. I love you and can never thank you enough for all you have done for me.

Table of Contents

Abstract.....	ii
Preface.....	iv
Acknowledgements.....	vi
List of tables.....	xi
List of figures.....	xii
Institutional abbreviations.....	xv
CHAPTER 1: Introduction	1
1.1 General introduction	1
1.2 Brief taxonomic history of Clupeomorpha	4
1.3 Challenges in clupeomorph systematics	7
1.4 Thesis objectives.....	9
References.....	11
CHAPTER 2: Evaluating the performance of probabilistic algorithms for phylogenetic analysis of big morphological data sets: a simulation study.....	24
2.1 Introduction.....	24
2.2 Materials and methods	27
2.2.1 True tree simulations.....	27
2.2.2 Dataset simulations	28
2.2.3 Phylogenetic analyses	29
2.2.4 Performance assessment	34
2.3 Results.....	36
2.3.1 Effect of the number of taxa	36
2.3.2 Effect of missing data	38
2.3.3 Branch lengths	39
2.4 Discussion.....	40
2.4.1 The performance of different probabilistic methods and software.....	40
2.4.2 Effects of increased taxon sampling on phylogenetic performance	41
2.4.3 The impact of different distance metrics	42
2.5 Conclusions.....	47
References.....	49

CHAPTER 3: Using empirical character state frequencies in morphological phylogenetics under Bayesian inference	74
3.1 Introduction.....	74
3.2 Materials and methods	76
3.2.1 Description and rationale of the FreqMorph model.....	76
3.2.2 Empirical data sets	80
3.2.3 Simulated data sets.....	80
3.2.4 Phylogenetic analyses	82
3.2.4 Performance assessment	83
3.3 Results.....	84
3.3.1 Empirical data sets	84
3.3.2 Simulated data set	85
3.4 Discussion.....	86
3.5 Conclusions.....	91
References.....	92
CHAPTER 4: Delimitation of <i>Alosa</i> species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches	111
4.1 Introduction.....	111
4.2 Material and methods.....	114
4.2.1 DNA extraction and mitochondrial gene sequencing	115
4.2.2 Next generation ddRAD sequencing	116
4.2.3 ddRAD data processing	117
4.2.4 Phylogenetic analyses and divergence time estimation.....	118
4.3 Results.....	119
4.3.1 PCA results	120
4.3.2 Phylogenetic results	121
4.3.3 Linear discriminant analysis results.....	122
4.4 Discussion.....	123
4.4.1 Divergence time estimation	126
4.5 Conclusions.....	129
References.....	131

CHAPTER 5: Morphological variation among the species of † <i>Armigatus</i> (Teleostei: Clupeomorpha: † <i>Ellimmichthyiformes</i>) and new material of † <i>Armigatus alticarpus</i> from the Late Cretaceous (Cenomanian) of Hakel, Lebanon	157
5.1 Introduction.....	157
5.2 Materials and methods	159
5.3 Systematic paleontology	162
5.3.1 Description.....	163
5.4 Phylogenetic results	168
5.5 Discussion.....	170
5.5.1 Phylogenetic affinities of † <i>Armigatus alticarpus</i>	170
5.5.2 Updated phylogeny of † <i>Ellimmichthyiformes</i>	171
5.6 Conclusions.....	176
References.....	178
CHAPTER 6: Phylogenetic analysis of Clupeomorpha	203
6.1 Introduction.....	203
6.2 Materials and methods	210
6.2.1 Data set construction.....	210
6.2.2 Phylogenetic analyses	212
6.3 Results.....	216
6.3.1 Morphological data set analyses	216
6.3.2 Molecular data set analyses	219
6.3.3 Combined data set time-calibrated analysis.....	219
6.4 Discussion.....	220
6.4.1 Ingroup relationships	223
6.5 Conclusions.....	246
References.....	248
CHAPTER 7	330
7.1 General conclusions	330
7.2 Future research.....	333
References.....	336
LITERATURE CITED	339

List of tables

TABLE 2-1. Summary of simulation parameters and phylogenetic analyses settings.....	57
TABLE 2-2. Median and standard deviation values of the normalized Robinson-Foulds (nRF)distances for the simulated datasets grouped into five bins.....	58
TABLE 2-3. Median and standard deviation values of the normalized Matching Splits (nMS) distances for the simulated datasets grouped into five bins.....	58
TABLE 2-4. Normalized RF (uniform) by number of taxa (median values).....	59
TABLE 2-5. Normalized MS (uniform) by number of taxa (median values).....	60
TABLE 2-6. Normalized RF (uniform) by number of taxa (STDEV.S values).....	61
TABLE 2-7. Normalized MS (uniform) by number of taxa (STDEV.S values).....	62
TABLE 2-8. Results of the rank test using the nRF values for each data set replicate.....	63
TABLE 2-9. Results of the rank test using the nMS values for each data set replicate.....	64
TABLE 2-10. Results of the rank test using the KF values for each data set replicate.....	65
TABLE 3-1. Results of the model fit analyses for empirical data sets.....	99
TABLE 3-2. Relative resolution and posterior support values for the summary topologies of the empirical data set runs recovered under the Mk and FreqMorph models.....	100
TABLE 3-3. Results of the two-tailed pairwise t-test comparing performance of the Mk and FreqMorph models. Statistically significant values ($p < 0.05$) are indicated in bold font.....	101
TABLE 4-1. Measurements and counts of <i>Alosa</i> spp. specimens from the Sea of Azov.....	144
TABLE 4-2. Morphological data recorded for the sampled individuals of <i>Alosa</i> spp.....	145
TABLE 4-3. Data sets used for the phylogenetic analyses.....	146
TABLE 5-1. Meristics and measurements (in mm) for the holotype and TMP 1998.65.11 specimens of † <i>Armigatus alticorpus</i>	184

List of figures

Figure 1-1. Representatives of the major extant clupeomorph lineages.....	21
Figure 1-2. Representatives of the extinct order †Ellimmichthyiformes.....	22
Figure 1-3. Phylogenetic hypothesis of Clupeomorpha after Grande (1985).....	23
FIGURE 2-1. Flow chart of study design.....	66
FIGURE 2-2. Homoplasy levels in simulated data sets.....	67
FIGURE 2-3. Normalized Robinson-Foulds and Matching Splits values for all methods across three categories of missing data.....	68
FIGURE 2-4. Normalized Robinson-Foulds and Matching Splits metric values by individual phylogenetic software and three categories of missing data.....	69
FIGURE 2-5. Contour plots showing density distribution of the results of phylogenetic analyses in the space of normalized Robinson-Foulds distances.....	70
FIGURE 2-6. Contour plots showing density distribution of the results of phylogenetic analyses in the space of normalized Matching Splits distances.....	71
FIGURE 2-7. Results of rank tests for normalized Robinson-Foulds, Matching Splits, and Kuhner-Felsenstein tree comparison metrics.....	72
FIGURE 2-8. Calculation of the Matching Splits metric for fully resolved topologies and topologies containing polytomies.....	73
FIGURE 3-1. Assymmetric beanplots showing distribution of normalized Robinson-Foulds distances for the maximum clade credibility trees and consensus topologies reconstructed under FreqMorph for individual categories of data sets.....	102
FIGURE 3-2. Stacked column charts showing results in a pooled sampled of 1100 simulated data sets.....	104

FIGURE 3-3. Comparison of phylogenetic accuracy of Mk and FreqMorph under the RFL metric per individual categories of data sets.....	105
FIGURE 3-4. Tree length skewness distributions.....	106
FIGURE 3-5. Proportion of missing data per character in the three empirical data sets.....	107
FIGURE 4-1. Species of <i>Alosa</i> in the Sea of Azov.....	147
FIGURE 4-2. Map showing sampling localities in the Sea of Azov.....	148
FIGURE 4-3. Morphometric measurements.....	149
FIGURE 4-4. Principal component analysis biplot showing distribution of the 20 sampled individuals in the PC1 – PC2 morphospace.....	150
FIGURE 4-5. Maximum likelihood COI/cyt b and SNPs topologies.....	151
FIGURE 4-6. Linear discriminant analysis scatter plot showing separation of the two genetic clusters in the morphospace.....	152
FIGURE 4-7. Phylogenetic chronogram based on the maximum clade credibility tree from a time-calibrated Bayesian analysis.....	153
FIGURE 5-1. † <i>Armigatus alticorpus</i> , NHMUK P.63134 (holotype).....	185
FIGURE 5-2. TMP 1998.65.11, limestone slab preserving multiple clupeomorph specimens.....	186
FIGURE 5-3. TMP 1998.65.11, four † <i>Armigatus alticorpus</i> specimens referred in this study.....	187
FIGURE 5-4. Close up of the head of the TMP 1998.65.11 specimen A.....	188
FIGURE 5-5. Caudal fin of † <i>Armigatus alticorpus</i>	189
FIGURE 5-6. Dorsal series of scutes of † <i>Armigatus alticorpus</i>	190
FIGURE 5-7. Majority rule consensus tree of 738 most parsimonious trees.....	191

FIGURE 6-1. Some diagnostic features of Clupeomorpha.....	272
FIGURE 6-2. Flow chart summarizing the analytical pipeline of this study	273
FIGURE 6-3. Effects of weighting strength function (K) on fit score	274
FIGURE 6-4. Equal weights maximum parsimony analysis of morphological.....	275
FIGURE 6-5. Implied weights parsimony analyses of morphological data.....	277
FIGURE 6-6. Maximum likelihood analysis of morphological data.....	278
FIGURE 6-7. Bayesian inference analysis of morphological data.....	280
FIGURE 6-8. Maximum likelihood analysis of molecular data.....	282
FIGURE 6-9. Bayesian inference analysis of molecular data.....	283
FIGURE 6-10. Time-calibrated phylogeny of Clupeomorpha.....	284
FIGURE 6-11. ‘Beryciform’ foramen in clupeomorph fishes.....	286
FIGURE 6-12. Simplified cladogram of Clupeomorpha based on combined evidence analysis.....	287

Institutional abbreviations

AMNH, American Museum of Natural History, New York, NY, USA;

CMN, Canadian Museum of Nature, Ottawa, Ontario, Canada;

CNHM, Department of Geology and Paleontology, Croatian Natural History Museum, Zagreb, Croatia;

DPC, Duke University Primate Centre;

JFBM, J. F. Bell Museum of Natural History, St. Paul, MN, USA;

MNHN, Muséum national d'Histoire naturelle, Paris, France;

NHMUK, Natural History Museum, London, United Kingdom;

TCWC, Texas A&M University, Texas Cooperative Wildlife Collections, College Station, TX, USA;

TMP, Royal Tyrell Museum, Drumheller, Alberta, Canada;

UALVP, University of Alberta Laboratory for Vertebrate Paleontology, Edmonton, Alberta, Canada;

UAMZ, University of Alberta Museum of Zoology, Edmonton, Alberta, Canada;

USNM, National Museum of Natural History, Washington, DC, USA;

UWFC, University of Washington, Seattle, WA, USA;

VIMS, Virginia Institute of Marine Science, Gloucester Point, VA, USA;

WM, Wembere-Manonga Palaeontological Expedition 1996 collections catalogued in the National Museum of Tanzania.

CHAPTER 1

Introduction

1.1 General introduction

The Clupeomorpha are a diverse group of teleost fishes commonly known as herrings, anchovies, sardines, and relatives (Figs 1-1, 1-2). The group includes over 400 extant and 150 extinct species distributed worldwide and occupying a wide range of ecological niches (Grande, 1985; Whitehead, 1985; Whitehead et al., 1988; Lavoué et al., 2013; Nelson et al., 2016; Fricke et al., 2020). Most living clupeomorph species are marine and euryhaline fishes that inhabit coastal waters in tropical regions. Fewer than a quarter of the species are primarily freshwater fishes (Whitehead, 1985; Whitehead et al., 1988; Froese and Pauly, 2019). Some clupeomorphs show remarkable salinity tolerance, with distinct marine and freshwater ecomorphs occurring within a single species. Moreover, some members of the group are diadromous, undertaking regular migrations between marine, freshwater, and brackish habitats throughout their life (Blaxter and Hunter, 1982; Whitehead, 1985; Whitehead et al., 1988; Lavoué et al., 2014). Overall distribution and species richness of clupeomorph fishes follow latitudinal and longitudinal gradients, with the highest diversity and abundance found in the Indo-West Pacific region; however, the geographical range of Clupeomorpha has a broad latitudinal extent with members of the group found worldwide between 70°N and 60°S latitude (Blaxter and Hunter, 1982; Lavoué et al., 2013; Froese and Pauly, 2019).

In marine ecosystems, clupeomorphs play an important role as forage species. Herrings and their relatives are shoaling and schooling fish that form large aggregations serving a perfect target for larger predatory fish, marine mammals, and sea birds (e.g., Blaxter and Hunter, 1982;

Parrish, 1993; Nøttestad and Axelsen, 1999; Fennessy et al. 2010; O'Donoghue et al., 2010a, 2010b, 2010c). Moreover, some clupeomorphs found in tropical waters are oceanodromous reef-associated fishes that serve an important role in nutrient transfer between open waters and reef ecosystems (Beukers-Stewart and Jones, 2004; Froese and Pauly, 2019; Robertson et al., 2019).

Clupeomorphs are among some of the most commercially valuable fishes. Their great abundance together with their remarkable schooling behaviour and high nutritional value make clupeomorphs one of the most heavily exploited groups of fish. Herrings, anchovies, and sardines significantly outnumber other commercial groups of fish and consistently lead in fisheries statistics both at local and global scales. In the most recent statistics yearbook of the Food and Agriculture Organization of the United Nations (FAO), eighteen clupeomorph species are listed among the top 70 species of commercial fish, crustaceans, and molluscs, annual capture production of which exceeds 150 000 tonnes (FAO, 2019). The most heavily exploited clupeomorph species are the Peruvian anchoveta (*Engraulis ringens* Jenyns, 1842), Atlantic herring (*Clupea harengus* Linnaeus, 1758), European pilchard (*Sardina pilchardus* (Walbaum, 1792)), and Japanese anchovy (*Engraulis japonicus* Temminck and Schlegel, 1846); these species contribute over one million tonnes to global fisheries production annually. In Canada, clupeomorph fishes contribute over 25% of the country's total fisheries production, with the most commercially important species being the Atlantic and Pacific herrings (*Clupea harengus* and *C. pallasii* Valenciennes in Cuvier and Valenciennes, 1847, respectively) and Alewife (*Alosa pseudoharengus* Wilson, 1811) (FAO, 2019; Department of Fisheries and Ocean, 2020).

Considering the high ecological and economic importance of clupeomorph fishes, it is of primary importance to have a strong understanding of the evolutionary history and interrelationships of these fishes. Special attention should be given to establishing a classification

system that can accurately identify natural relationships within the group and provide useful information for identification of members of this group of fishes. This knowledge can then be translated into biologically sound conservation and stock management measures. However, despite a long history of research on Clupeomorpha, interrelationships within the group are still not well-understood. Current taxonomic research on the group is focused primarily on using molecular data to resolve ingroup relationships (e.g., Li and Orti, 2007; Lavoué et al., 2007, 2013; Wilson et al., 2008; Bloom and Lovejoy, 2012, 2014; Bloom et al., 2018). While molecular phylogenetic approaches offer high resolution and support for new or previously established evolutionary lineages within the Clupeomorpha, they preclude inclusion of fossil taxa except for using them to inform divergence date estimation in the phylogenetic analysis; therefore, it is impossible to infer phylogenetic placement of the fossil members of the group using molecular data alone, which leads to a cascade of challenges associated with the accurate estimation of evolutionary processes throughout the long history of the group. Another issue associated with using only molecular data to infer evolutionary interrelationships is that this approach does not provide an answer to distinguishing species visually. Both problems can be solved by using an integrative approach combining molecular and morphological data to infer phylogenetic relationships. The combined evidence approach has been successfully implemented in phylogenetic studies of various groups of organisms (e.g., Lee et al., 2013; Brusatte et al., 2014; Simões et al., 2017; Halliday et al., 2019); however, a comprehensive study of Clupeomorpha is still lacking.

1.2 Brief taxonomic history of Clupeomorpha

The systematics of clupeomorph fishes has a long history that has largely been shaped by their commercial value. Even before the advent of the Linnaean taxonomic system, herring-like fishes have been set apart from other fishes due to recognition of their economic importance in Northern Europe, where taxonomy and fisheries science took on their modern framework (Whitehead, 1985).

In the first edition of *Systema Naturae* (1735), Carl Linnaeus recognized four species within the single genus *Clupea*. The genus included well-recognized species of primary economic importance in Europe – European herring (*Clupea harengus*), anchovy (*Engraulis encrasicolus* (Linnaeus, 1758)), sprat (*Sprattus sprattus* (Linnaeus, 1758)), and shad (*Alosa alosa* (Linnaeus, 1758)). By the time the tenth edition of his *Systema Naturae* was published in 1758, the genus contained ten species including two clupeomorph species from the Indo-West Pacific region, gizzard shad (*Clupanodon thrissa* (Linnaeus, 1758)) and grenadier anchovy (*Coilia mystus* (Linnaeus, 1758)). The other four species added to the group in the tenth edition were later re-classified and are now recognized as either ostariophysans or *nomina dubia* (Stiassny et al., 1996). Linnaeus' diagnosis of the group, however, was based largely on plesiomorphic characters, such as position of the pelvic fins and number of fin rays, that are not unique to clupeomorph fishes. This problem became more apparent as more species were being added to the group, making it difficult, if not impossible, to draw a line between clupeomorphs and other similar basal teleosts.

For over two centuries clupeomorphs continued to be classified broadly as the most basal teleosts and were placed in the order Isospondyli (Cope, 1871; Gill, 1872; Woodward, 1895) together with osteoglossomorphs, elopomorphs, and other teleosts of uncertain affinities. Some

of later authors at the beginning of the XX century classified clupeomorphs with salmoniforms, esociforms, albuliforms and other difficult to classify groups of fishes (e.g., Regan, 1929; Berg, 1940; Svetovidov, 1952). The artificial grouping of these fishes was apparent to many researchers, prompting a rigorous revision of Clupeomorpha and other basal teleost lineages.

Greenwood et al. (1966) were the first to provide a thorough reassessment of Clupeomorpha and establish a diagnosis for the group based on unique derived characteristics. These characteristics were divided into three anatomical complexes (cranial system, sensory system, and caudal skeleton) and included the following features: 1) intracranial connection between the swim bladder and the inner ear, temporal foramina, pre-epiotic fossae, and the auditory fenestrae; 2) reduced lateral line system and presence of the recessus lateralis; 3) fusion of the second hypural and the first ural centrum, separation of the first hypural from the first ural centrum, and fusion between the first uroneural and the first preural centrum.

This classification and diagnosis of Clupeomorpha was widely adopted in subsequent studies by other authors who added more morphological features to the diagnosis and expanded taxonomic revision of Clupeomorpha by including fossil taxa in their studies (Patterson and Rosen, 1977; Grande, 1982, 1985). Among these studies, the systematic revision of Clupeomorpha by Grande (1982, 1985) was a turning point in our understanding of the group's relationships. Grande's classification system (Fig. 1-3) was based on extensive morphological examination of over 150 extant and fossil clupeomorph species and he provided clear lists of diagnostic features for the major clupeomorph groups. He expanded the superorder to include extinct species of 'double-armoured' herrings such as †*Diplomystus*, †*Ellimmichthys*, and †*Armigatus*, which were placed in a new order †Ellimmichthyiformes, except for the genus †*Armigatus* that was left outside of the order in a polytomy (Fig. 1-3A). Like previous authors

(Whitehead et al., 1966; Patterson and Rosen, 1977), Grande diagnosed Clupeomorpha based on the presence of an otophysic connection between the swimbladder and the inner ear, fusion between the second hypural and the first ural centrum, and the presence of a well-defined pre-epiotic fossa. He also added presence of abdominal scutes (at least one) and a series of dorsal scutes to the list of diagnostic clupeomorph characteristics (Grande, 1985).

Grande (1985) subdivided Clupeomorpha into two Divisions. Division 1 contained a single enigmatic taxon, †*Erichalsis arcta* Forey, 1975, and Division 2 included all the rest of the traditionally recognized clupeomorph taxa divided into two orders, the †Ellimmichthyiformes comprised of the fossil genera †*Diplomytus* and †*Ellimmichthys*, and Clupeiformes containing all extant clupeomorph species and some derived fossil taxa. The fossil genus †*Armigatus* was not classified within either of the clupeomorph orders because members of this group did not have subrectangular lateral wings of the dorsal scutes as in †*Diplomytus* or †*Ellimmichthys* nor any of the more derived traits diagnostic of Clupeiformes. Within Clupeiformes, Grande (1985) recognized two suborders, Denticipitoidei and Clupeoidei; the latter included four families, the Engraulidae (anchovies), Pristigasteridae (longfin herrings), Chirocentridae (wolf herrings), and Clupeidae (herrings). The herring family Clupeidae included subfamilies Clupeinae, Alosinae, Dorosomatinae, Pellonulinae, and Dussumieriinae (Fig. 1-3B).

Subsequent works on clupeomorph taxa showed that †*Erichalsis arcta* is not a valid clupeomorph taxon (Arratia 1997; Hermus and Wilson, 2001); therefore there is no difference between Grande's Division 2 and Clupeomorpha as a whole. The phylogenetic relationships of †*Armigatus* were also reassessed and the genus was placed within †Ellimmichthyiformes (e.g., Chang and Maisey, 2003; Zaragüeta-Bagils, 2004; Alvarado-Ortega et al., 2008; Murray and Wilson, 2013). However, apart from these changes, the overall framework of clupeomorph

interrelationships established by Grande (1985) is still widely used and recognized in systematic research (e.g., Sato, 1994; Di Dario, 2002; Lavoué et al., 2007, 2013; Li and Orti, 2007; Wilson et al., 2008; Bloom and Lovejoy, 2012, 2014; Bloom and Egan, 2018).

1.3 Challenges in clupeomorph systematics

Many problems in clupeomorph systematics persist today. Interrelationships as well as the composition of major lineages of clupeomorph fishes are still a subject of debate (e.g., Di Dario, 2002; Lavoué et al., 2013; Bloom and Egan, 2018). Very few morphological studies of Clupeomorpha have been conducted since Grande's revision of the group; these studies almost invariably focus on extant members of the group and investigate a relatively small suite of phenotypic traits to infer interrelationships of the major family groups (e.g., Di Dario, 2002; Miyashita, 2010). A more detailed species level assessment of clupeomorph fishes was performed by Sato (1994) who compiled an extensive list of morphological traits to infer evolutionary relationships within Clupeoidei. His work did not include fossil taxa but provided a sound basis for future morphological revision of the group. More importantly, the most recent phylogenetic studies of clupeomorph fishes are based on molecular data alone and do not investigate phylogenetic placement of fossil taxa (e.g., Li and Orti, 2007; Wilson et al., 2008; Lavoué et al., 2013; Bloom and Lovejoy, 2014; Bloom and Egan, 2018). These molecular studies recover topologies that suggest alternative arrangements of clupeomorph lineages compared to the traditional classification established by Grande (1985) and used in subsequent morphological studies (e.g., Sato, 1994; Di Dario, 2002; Miyashita, 2010). This growing body of evidence calls for a new revision of Clupeomorpha that would incorporate morphological and molecular data to

clarify interrelationships of clupeomorph lineages including fossil taxa. This is here addressed as the higher rank relationships within Clupeomorpha studied in detail in Chapter 6 of my thesis.

The problem of delimiting clupeomorph taxa is also pertinent at the species level. Interspecific variation within a genus often exhibits a high degree of overlap of phenotypic and ecological traits that makes species delimitation a challenging task (e.g., McDowal, 2001, 2003; Alexandrino et al., 2006; Thomas et al., 2014). This problem is amplified by their high degree of intraspecific variation, which further blurs species boundaries in closely related species. In extant taxa, this issue is commonly addressed with an integrative approach that combines different available lines of evidence to delimit species boundaries (Schlick-Steiner et al., 2010; Stern et al., 2016). In palaeontological research, the species delimitation problem is usually addressed in a framework with the morphological species concept and consideration of the temporal occurrence of the taxa (Drooger, 1954). Despite differences in approaches, both neontological and palaeontological research aim to answer the same question of delineating boundaries of species. Two case studies highlighting these two different approaches are provided in Chapters 4 (extant species) and 5 (fossil species) of the thesis.

Finally, reconstructing evolutionary relationships requires accurate and objective methods of phylogenetic inference. The choice of a method depends on the type of data used in the evolutionary analysis and the individual characteristics of a data set, such as consistency. While the field of molecular phylogenetics has rapidly adopted probabilistic methods of phylogenetic inference (maximum likelihood and Bayesian inference), morphological phylogenetics has traditionally been dominated by parsimony methods and non-cladistic approaches to the reconstruction of evolutionary relationships. It was not until recently that the performance of phylogenetic inference methods using morphological data was assessed in simulation studies

(Wright et al. 2014, O'Reilly et al. 2016, Goloboff et al. 2018, Puttick et al. 2017a, 2017b, O'Reilly et al. 2018a, 2018b). However, it remains unclear whether probabilistic methods can provide an advantage for morphological phylogenetics that is similar to what they do for molecular phylogenetics. Additionally, no formal assessment has been performed to determine which probabilistic method provides more accurate phylogenetic reconstructions. Assessing the best way to analyse morphological data is essential to evolutionary biology in order to accurately infer evolutionary relationships of fossil taxa. Integrating available morphological and molecular data for extant and fossil taxa in combined evidence studies provides a reliable way to assess evolutionary relationships and estimate divergence times across various branches of the tree of life (Guillerme et al. 2016; Ronquist et al. 2016; Zhang et al. 2016; Simões et al., 2018). A performance assessment of probabilistic methods for morphological phylogenetics is provided in Chapter 2. A new approach to analysis of morphological data in a Bayesian inference framework is presented in Chapter 3 of my thesis.

1.4 Thesis objectives

My doctoral dissertation is focused on a comprehensive study of the teleost group Clupeomorpha including both extinct and extant members. The overarching aim of the thesis is to provide an updated classification of the superorder Clupeomorpha and to improve our understanding of the relationships and evolutionary history of the members of this commercially important group of fishes. In this interdisciplinary research, I combine traditional comparative anatomy techniques with molecular analyses to address the issues in clupeomorph systematics listed above.

The main objectives of my thesis include: 1) to assess the performance of probabilistic methods for phylogenetic analysis of morphological data sets; 2) to introduce a new approach to morphological data analysis under the Bayesian inference framework; 3) to investigate interspecific variation and species delimitation in closely related clupeomorph species in extant and fossil case studies; 4) to conduct an extensive phylogenetic analysis of the Clupeomorpha using a combined evidence data set.

References

- Alexandrino, P., R. Faria, D. Linhares, F. Castro, M. Le Corre, R. Sabatié, J. L. Bagliniere, and Weiss, S. 2006. Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. *Journal of Fish Biology* 69:242–259.
- Alvarado-Ortega, J., E. Ovalles-Damián, and G. Arratia. 2008. A review of the interrelationships of the order Ellimmichthyiformes (Teleostei: Clupeomorpha); pp. 257–278 in G. Arratia, H.-P. Schultze, and M.V.H. Wilson (eds.), *Mesozoic Fishes 4 – Homology and Phylogeny*. Verlag Dr. Friederich Pfeil, Munich, Germany, 502 pp.
- Arratia, G. 1997. Basal teleosts and teleostean phylogeny. *Palaeo Ichthyologica* 7:1–168.
- Berg, L. S. 1940. Sistema ryboobraznykh i ryb, nye zhivushchikh i iskopaemykh: Classification of fishes, both recent and fossil. *Izd-vo Akademii nauk SSSR*, 517 pp.
- Beukers-Stewart, B. D. and G. P. Jones. 2004. The influence of prey abundance on the feeding ecology of two piscivorous species of coral reef fish. *Journal of Experimental Marine Biology and Ecology* 299:155–184.
- Blaxter, J. H. S. and J. R. Hunter. 1982. The biology of the clupeoid fishes. *Advances in Marine Biology* 20:1–223.
- Bloom, D. D. and J. P. Egan. 2018. Systematics of Clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. *Neotropical Ichthyology* 16:p.e180095.

- Bloom, D. D. and N. R. Lovejoy. 2012. Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *Journal of Evolutionary Biology* 25:701–715.
- Bloom, D. D. and N. R. Lovejoy. 2014. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proceedings of the Royal Society B: Biological Sciences* 281:2013–2081.
- Bloom, D. D., M. D. Burns, and T. A. Schriever. 2018. Evolution of body size and trophic position in migratory fishes: a phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies). *Biological Journal of the Linnean Society* 125:302–314.
- Brusatte, S. L., G. T. Lloyd, S. C. Wang, and M. A. Norell. 2014. Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. *Current Biology* 24:2386–2392.
- Chang, M.-M. and J. Maisey. 2003. Redescription of *Ellimma branneri* and *Diplomystus shengliensis*, and relationships of some basal clupeomorphs. *American Museum Novitates* 3404:1–35.
- Cope, E. D. 1877. A contribution to the knowledge of the ichthyological fauna of the Green River Shales. *Bulletin of United States Geological and Geographical Survey* 3:807–819.
- Department of Fisheries and Ocean. Canada. 2020. Seafisheries landed quantity by province, 2018. Preliminary statistics. Electronic version 10/02/2020 accessed 20 May 2020.
- Di Dario, F. 2002. Evidence supporting a sister-group relationship between Clupeoidea and Engrauloidea (Clupeomorpha). *Copeia* 2002: 496–503.

- Drooger, C. W. 1954. Remarks on the species concept in paleontology. *The Micropaleontologist* 8:23–26.
- FAO Fisheries department. 2019. FAO yearbook. Fishery and Aquaculture Statistics 2017/FAO annuaire. Statistiques des pêches et de l'aquaculture 2017/ FAO anuario. Estadísticas de pesca y acuicultura 2017. Rome/Roma, 109 pp.
- Fennessy, S. T., P. Pradervand, and P. A. de Bruyn. 2010. Influence of the sardine run on selected nearshore predatory teleosts in KwaZuluNatal. *African Journal of Marine Science* 32:375–382.
- Forey, P. L. 1975. A fossil clupeomorph fish from the Albian of the Northwest Territories of Canada, with notes on cladistic relationships of clupeomorphs. *Journal of Zoology* 175:151–177.
- Fricke, R., W. N. Eschmeyer, and J. D. Fong. 2020. Eschmeyer's catalog of fishes: species by family/subfamily. <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>. Electronic version accessed 28 May 2020.
- Froese, R. and D. Pauly. 2019. FishBase. World Wide Web electronic publication. www.fishbase.org. Electronic version 12/2019 accessed 28 May 2020.
- Gill, T. 1872. Arrangement of the families of fishes, or Classes Pisces, Marsipobranchii, and Leptocardii. *Smithsonian Miscellaneous Collections* 247:1–49.
- Goloboff, P. A., A. Torres, and J. S. Arias. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* 34:407–437.
- Grande, L. 1982. A revision of the fossil genus *Diplomystus*, with comments on the interrelationships of clupeomorph fishes. *American Museum Novitates* 2728:1–34.

- Grande, L. 1985. Recent and fossil clupeomorph fishes, with materials for the revision of the subgroups of clupeoids. *Bulletin of the American Museum of Natural History* 181:231–372.
- Greenwood, P. H., D. E. Rosen, S. H. Weitzman, and G. S. Myers. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History* 131:339–456.
- Guillerme T. and N. Cooper. 2016. Effects of missing data on topological inference using a Total Evidence approach. *Molecular Phylogenetics and Evolution* 94:146–158.
- Halliday, T. J., M. dos Reis, A. U. Tamuri, H. Ferguson-Gow, Z. Yang, and A. Goswami. 2019. Rapid morphological evolution in placental mammals post-dates the origin of the crown group. *Proceedings of the Royal Society B* 286:p.20182418.
- Hermus, C. R. and M. V. H. Wilson. 2004. A new species of *Erichalcis* from Arctic Canada, with a revised diagnosis of the genus; pp. 405–427 in G. Arratia, and A. Tintori (eds.), *Mesozoic Fishes 3—Systematics, Paleoenvironments and Biodiversity*. Verlag Dr. Friedrich Pfeil, Munich, 649 pp.
- Jenyns, L. 1842. Fish; Part IV in C. Darwin (ed.), *The zoology of the voyage of H. M. S. Beagle, under the command of Captain Fitzroy, R. N., during the years 1832 to 1836*. Smith, Elder, and Co, London, 172 pp.
- Lavoué, S., M. Miya, K. Saitoh, and N. B. Ishiguro. 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Molecular Phylogenetics and Evolution* 43:1096–1105.

- Lavoué, S., M. Miya, P. Musikasinthorn, W.-J. Chen, and M. Nishida. 2013. Mitogenomic evidence for an Indo-West Pacific origin of the Clupeoidei (Teleostei: Clupeiformes). *PLoS ONE* 8:e56485.
- Lavoué, S., P. Konstantinidis, and W. J. Chen. 2014. Progress in clupeiform systematics; pp. 3–42 in K. Ganas (ed.), *Biology and ecology of sardines and anchovies*, CRC Press, 394 pp.
- Lee, M. S. Y., J. Soubrier, and G. D. Edgecombe. 2013. Rates of phenotypic and genomic evolution during the Cambrian explosion. *Current Biology* 23:1889–1895.
- Li, C. and G. Orti. 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 44:386–398.
- Linnaeus, C. 1735. *Systema Naturae, sive Regna Tria Naturae Systematice Proposita per Classes, Ordines, Genera, and Species*. Theodorum Haak, Lugduni Batavorum, Leiden, 13 pp.
- Linnaeus, C. 1758. *Systema Naturae per Regna Tria Naturae, secundum Classes, Ordines, Genera, Species, cum Characteribus, Differentiis, Synonymis, Locis*. Tonus I. Editio Decima, Reformata. Laurentii Salvii, Holmiae, 824 pp.
- McDowall, R. M. 2001. Diadromy, diversity and divergence: implications for speciation process. *Fish and Fisheries* 2:278–286.
- McDowall, R. M. 2003. Shads and anadromy: implications for ecology, evolution, and biogeography; pp. 11–23 in K. E. Limburg and J. R. Waldman (eds.), *Biodiversity, Status, and Conservation of the World's Shads*. Baltimore, MD: American Fisheries Society, 370 pp.

- Miyashita, T. 2010. Unique occipital articulation with the first vertebra found in pristigasterids, chirocentrids, and clupeids (Teleostei: Clupeiformes: Clupeoidei). *Ichthyological Research* 57:121–132.
- Murray, A. M., and M. V. H. Wilson. 2013. Two new paraclupeid fishes (Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco; pp. 267–290 in G. Arratia, H.-P. Schultze, and M. V. H. Wilson (eds.), *Mesozoic Fishes 5 – Global Diversity and Evolution*. Verlag Dr. Friederich Pfeil, Munich, Germany, 560 pp.
- Nelson, J. S., T. C. Grande, and M. V. H. Wilson. 2016. *Fishes of the World*, fifth edition. John Wiley and Sons, Toronto, 752 pp.
- Nøttestad, L. and B. E. Axelsen. 1999. Herring schooling manoeuvres in response to killer whale attacks. *Canadian Journal of Zoology* 77:1540–1546.
- O'Donoghue, S. H., L. Drapeau, and V. M. Peddemors. 2010b. Broad-scale distribution patterns of sardine and their predators in relation to remotely sensed environmental conditions during the KwaZuluNatal sardine run. *African Journal of Marine Science* 32:279–291.
- O'Donoghue, S. H., L. Drapeau, S. F. J. Dudley, and V. M. Peddemors. 2010a. The KwaZulu-Natal sardine run: shoal distribution in relation to nearshore environmental conditions, 1997–2007. *African Journal of Marine Science* 32:293–307.
- O'Donoghue, S. H., P. A. Whittington, B. M. Dyer, and V. M. Peddemors. 2010c. Abundance and distribution of avian and marine mammal predators of sardine observed during the 2005 KwaZuluNatal sardine run survey. *African Journal of Marine Science* 32:361–374.
- O'Reilly J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018a. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* 61:105–118.

- O'Reilly J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018b. Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff et al. *Palaeontology* 61:631–635.
- O'Reilly J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biological Letters* 12:20160081.
- Parrish, J. K. 1993. Comparison of the hunting behavior of four piscine predators attacking schooling prey. *Ethology* 95:233–246.
- Patterson, C. and D. E. Rosen. 1977. Review of ichthyodectiform and other Mesozoic teleost fishes, and the theory and practice of classifying fossils. *Bulletin of the American Museum of Natural History* 158:81–172.
- Puttick, M. N., J. E. O'Reilly, D. Oakley, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, and D. Pisani. 2017b. Parsimony and maximum-likelihood phylogenetic analyses of morphology do not generally integrate uncertainty in inferring evolutionary history: a response to Brown et al. *Proceedings of the Royal Society B: Biological Sciences* 284:20171636.
- Puttick, M. N., J. E. O'Reilly, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, and P. C. J. Donoghue. 2017a. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B* 284:p.20162290.
- Regan, C. T. 1929. Fishes. *Encyclopaedia Britannica* 9:305–329.

- Robertson, D. R., O. Domínguez-Domínguez, Y. M. L. Aroyo, R. M. Mendoza, and N. Simões. 2019. Reef-associated fishes from the offshore reefs of western Campeche Bank, Mexico, with a discussion of mangroves and seagrass beds as nursery habitats. *ZooKeys* 843:71–115.
- Ronquist F., N. Lartillot, and M. J. Phillips. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B* 371:20150136.
- Sato, Y. 1994. Cladistic-phenetic relationships and the evolutionary classification of the suborder Clupeoidei (Pisces, Teleostei). Unpublished PhD thesis. Tokyo Metropolitan University, Tokyo, 186 pp.
- Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of entomology* 55:421–438.
- Simões, T. R., M. W. Caldwell, M. Tałanda, M. Bernardi, A. Palci, O. Vernygora, F. Bernardini, L. Mancini, and R. L. Nydam. 2018. The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature* 557:706–709.
- Stern, N., B. Rinkevich, and M. Goren. 2016. Integrative approach revises the frequently misidentified species of *Sardinella* (Clupeidae) of the Indo-West Pacific Ocean. *Journal of Fish Biology* 89:2282–2305.
- Stiassny, M. L., L. R. Parenti, and G. D. Johnson. 1996. *Interrelationships of Fishes*. Academic Press, London, 496 pp.
- Svetovidov, A. N. 1952. *Fauna of the USSR, Fishes. Herrings (Clupeidae)*. Academy of Sciences of USSR publishing, Moscow-Leningrad, 331 pp.

- Temminck, C. J. and H. Schlegel. 1846. Pisces; pp. 173–269 in: Siebold, P. F. de (ed.), Fauna Japonica, sive descriptio animalium, quae in itinere per Japoniam suscepto annis 1823-1830 collegit, notis, observationibus et adumbrationibus illustravit Ph. Fr. de Siebold. Lugduni Batavorum, 323 pp.
- Thomas Jr, R. C., D. A. Willette, K. E. Carpenter, and M. D. Santos. 2014. Hidden diversity in sardines: genetic and morphological evidence for cryptic species in the goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849). PLoS ONE 9:p.e84719.
- Valenciennes, A. 1847. In Cuvier, G. and A. Valenciennes. Histoire naturelle des poissons. Tome vingtième. Livre vingt et unième. De la famille des Clupéoïdes. Histoire Naturelle Des Poissons, 472 pp.
- Walbaum, J. J. 1792. Petri Artedi sueci genera piscium. In quibus systema totum ichthyologiae proponitur cum classibus, ordinibus, generum characteribus, specierum differentiis, observationibus plurimis. Redactis speciebus 242 ad genera 52. Ichthyologiae pars III. Ant. Ferdin. Rose, Grypeswaldiae, 723 pp.
- Whitehead, P. J. P. 1985a. King herring: his place amongst the clupeoids. Canadian Journal of Fisheries and Aquatic Sciences 42:3–20.
- Whitehead, P. J. P. 1985b. FAO species catalogue. vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part I. Chirocentridae, Clupeidae and Pristigasteridae. FAO Fisheries Synopsis 125:1–303.
- Whitehead, P. J. P., G. J. Nelson, and T. Wongratana. 1988. FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part 2. Engraulididae. FAO Fisheries Synopsis 125:305–579.

- Wilson, A. 1811. [Article on Clupea.] *The Cyclopedic; or, universal dictionary of arts, sciences, and letters*. v. 9: No pagination.
- Wilson, A. B., G. G. Teugels, and A. Meyer. 2008. Marine incursion: The freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. *PLoS ONE* 3:e1979.
- Woodward, A. S. 1895. *Catalogue of the Fossil Fishes in the British Museum (Natural History), Cromwell Road, S. W. Part III. Containing the Actinopterygian Teleostomi of the Orders Chondrostei (concluded), Protospondyli, Aetheospondyli, and Isospondyli (in part)*. British Museum (Natural History), London, 706 pp.
- Wright A. M., and D. M. Hillis. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* 9:e109210.
- Zaragüeta-Bagils, R. 2004. Basal clupeomorphs and ellimmichthyiform phylogeny; pp. 391–404 in G. Arratia, G. Tintori, and A. Tintori (eds.), *Mesozoic Fishes 3 – Systematics, Paleoenvironment and Biodiversity*. Verlag Dr. Friederich Pfeil, Munich, Germany, 703 pp.
- Zhang C., T. Stadler, S. Klopstein, T. A. Heath, and F. Ronquist. 2016. Total-Evidence dating under the fossilized birth–death process. *Systematic Biology* 65:228–249.

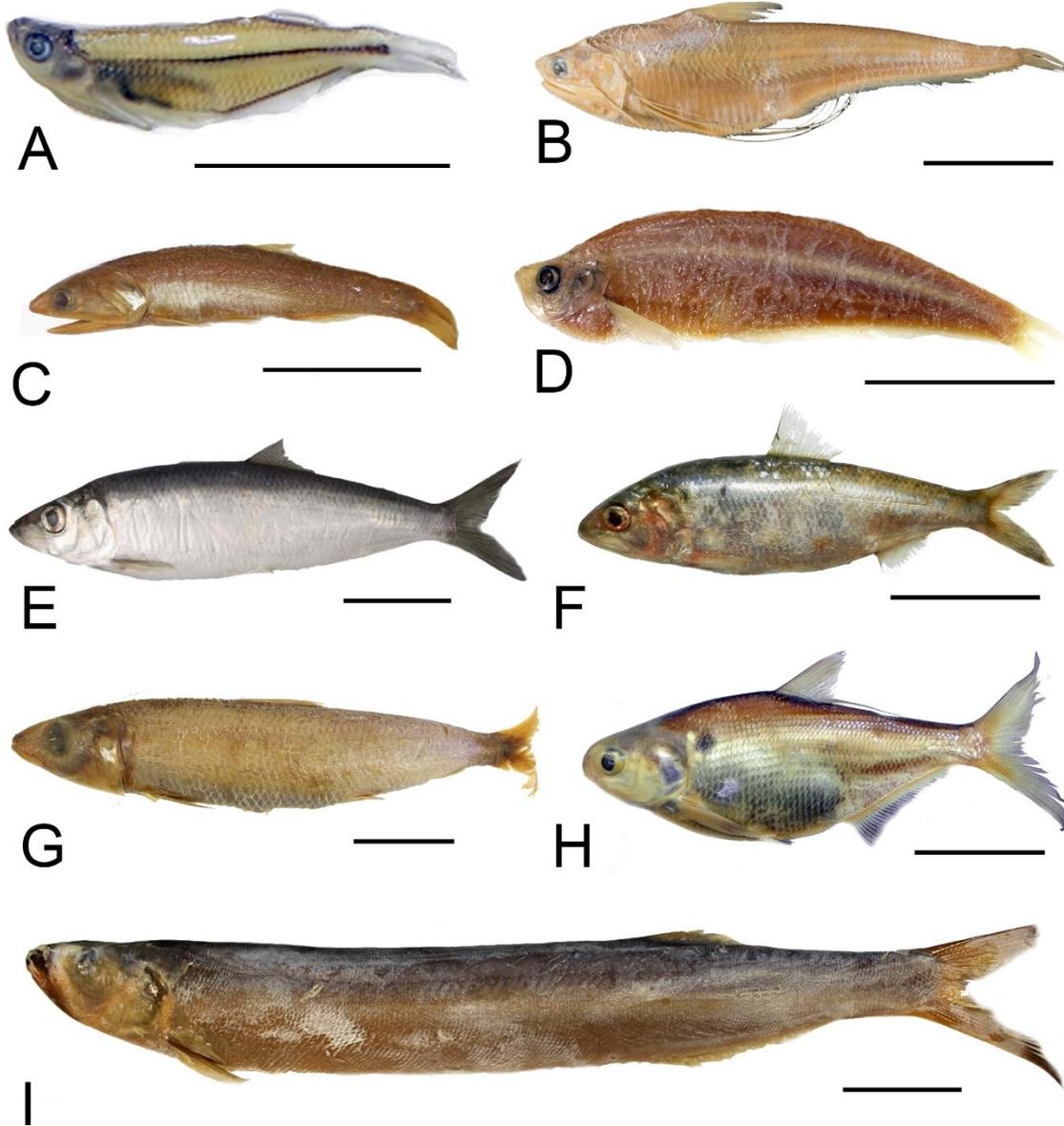


FIGURE 1-1. Representatives of major extant clupeomorph lineages. A – *Denticeps clupeoides* AMNH 235843 (Denticipitidae); B – *Coilia lindmani* MNHN 23-205 (Engraulidae); C – *Engraulis mordax* CMN 78-0075 (Engraulidae); D – *Odonthognathus panamensis* (Pristigasteridae); E – *Clupea harengus* MNHN 2004-1490 (Clupeidae); F – *Alosa caspia*, uncatalogued material (Alosinae); G – *Etrumeus sadina* CMN 54-629 (Dussumieriidae); H – *Dorosoma cepedianum* AMNH 243379 (Dorosomatinae); I – *Chirocentrus dorab*, MNHN 1966-0194 (Chirocentridae). Scale bar equals 2 cm in (A) and 5 cm in (B – I).

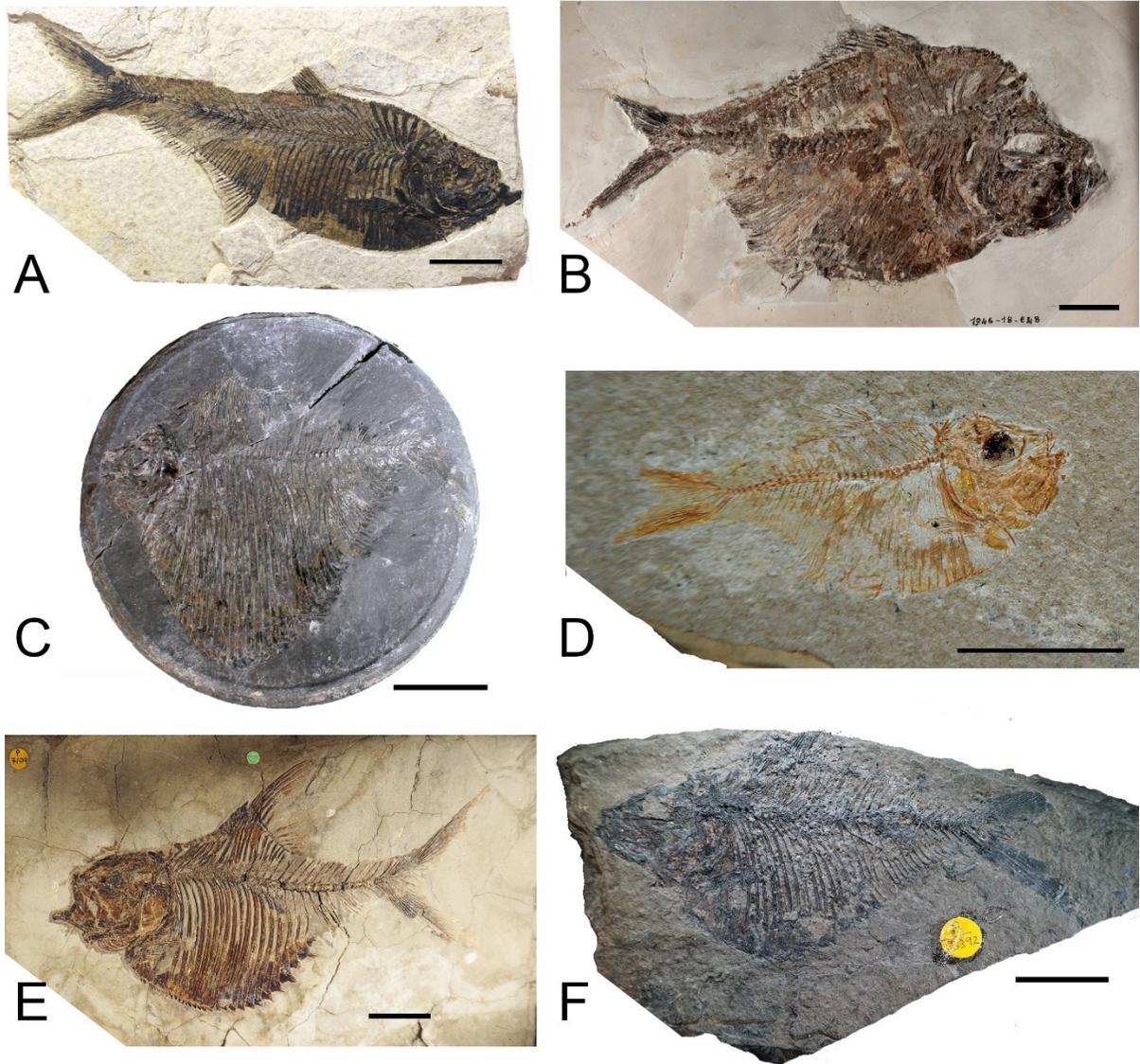


FIGURE 1-2. Representatives of the extinct order †Ellimmichthyiformes. A – †*Diplomystus dentatus* CMN 41639; B – †*Diplomystus dubertreti* MNHN F.SHA2052; C – †*Tychoeroichthys dunveganensis* CMN 52730; D – †*Armigatus alticorpus* NHMUK P.63134 (photo credit Alison Murray); E – †*Ellimmichthys longicostatus* NHMUK P.7109 (photo credit Alison Murray); F – †*Ellimmichthys goodi* NHMUK P.13392 (photo credit Alison Murray). Scale bar equals 2 cm.

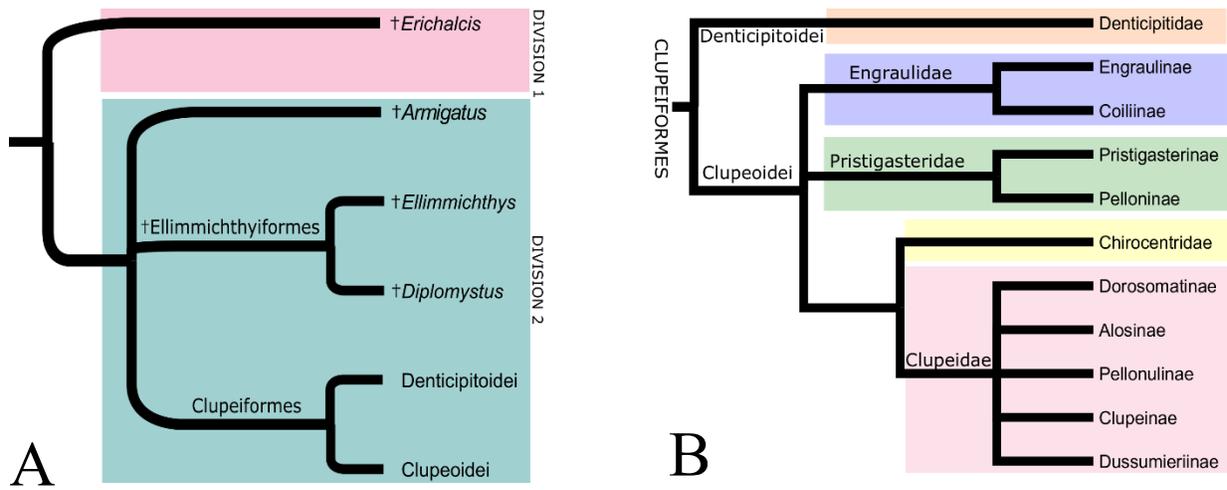


FIGURE 1-3. Phylogenetic hypothesis of Clupeomorpha after Grande (1985). A – interrelationships of major clupeomorph lineages; B – interrelationships of Clupeiformes.

CHAPTER 2

Evaluating the performance of probabilistic algorithms for phylogenetic analysis of big morphological datasets: a simulation study

A version of this chapter has been published as Vernygora, O.V., T. R. Simões, and E. O. Campbell. 2020. Evaluating the performance of probabilistic algorithms for phylogenetic analysis of big morphological datasets: a simulation study. *Systematic Biology*: 10.1093/sysbio/syaa020.

2.1 Introduction

Reconstructing the tree of life is essential to evolutionary biology. The past two decades have been marked by great advances in the development of phylogenetic methods including faster and more accurate algorithms for the most widely used optimality criteria: maximum parsimony (MP) (Goloboff, 1999; Goloboff et al., 2016), as well as maximum likelihood (ML) (Stamatakis, 2014; Nguyen et al., 2015) and Bayesian inference (BI) (Ronquist et al., 2012; Bouckaert et al., 2014; Höhna et al., 2016). These improvements are particularly important for morphological phylogenetics that allows inclusion of fossil taxa in evolutionary analyses. Because most of the diversity of life that has ever existed is now extinct (Novacek et al., 1992), it is essential to incorporate fossil taxa in large scale phylogenetic analyses to reconstruct the tree of life. Integrating both extant and fossil taxa in total evidence studies provides a more reliable way to assess phylogenetic relationships and divergence times across various branches of the tree of life (Guillerme et al., 2016; Ronquist et al., 2016), such as estimating the origin of placental

mammals (Ronquist et al., 2016), squamates (Simões et al., 2018b) and hymenopterans (Zhang et al., 2016).

The use of probabilistic methods in morphological phylogenetics has lagged behind that of molecular phylogenetics, and MP remains the most commonly used approach for analyzing morphological data. Only recently have simulation studies been used to test the performance of phylogenetic inference methods using morphological data (Wright and Hillis, 2014; O'Reilly et al., 2016; Puttick et al., 2017; Goloboff et al., 2018; O'Reilly et al., 2018a; Goloboff et al., 2019; Puttick et al., 2019). Almost invariably, these simulation studies have found that BI outperforms MP for discrete morphological characters under different scenarios of rate heterogeneity, tree shape, and even when data are simulated under a non-probabilistic framework (i.e., without an assumption of shared branch lengths across characters) (Wright and Hillis, 2014; O'Reilly et al., 2016; Puttick et al., 2017; O'Reilly et al., 2018a; Puttick et al., 2019). This difference in performance seems to be particularly evident when data for slow-evolving characters are missing and when the number of morphological characters is relatively low (Wright and Hillis 2014; O'Reilly et al., 2016; Puttick et al., 2017; O'Reilly et al., 2018a).

While previous morphological studies have compared the performance of MP to probabilistic methods, a benchmark assessment using simulated morphological datasets across different probabilistic software is currently lacking. Recently developed ML applications have been designed to improve phylogenetic inference of big taxonomic datasets (Nguyen et al., 2015), however their performance has only been tested using empirical molecular datasets (Nguyen et al., 2015; Höhna et al., 2016; Zhou et al., 2018), which cannot assess phylogenetic accuracy because the true trees are not known. Furthermore, there has been no formal performance comparison of different ML and BI software for the analysis of morphological data

in a single study. It is therefore unclear whether software implementing either BI or ML differ substantially in their performance and the accuracy of reconstructed topologies.

Furthermore, increased taxon sampling has been found to improve accuracy using molecular data (Hillis, 1996, 1998; Pollock et al., 2002; Zwickl et al., 2002; Hillis et al., 2003) despite the fact that this exponentially increases the number of potential trees obtained in any given analysis (Felsenstein, 1978). However, the impact of increased dataset size on the accuracy of phylogenetic analysis using morphological data is so far unknown. As “big” morphological datasets become increasingly common in systematics (Simões et al., 2018a), it is important to directly assess how dataset size impacts phylogenetic inference under probabilistic approaches.

Previous studies comparing the performance of alternate phylogenetic approaches have almost invariably used only a single tree comparison metric, Robinson-Foulds (RF) distance, to assess the accuracy of reconstructed topologies. However, there are many other metrics, including Matching Splits (MS), computationally effective Subtree Prune-and-Regraft distance (SPR, Whidden and Matsen, 2018), tree alignment (‘Align’, Nye et al., 2006), triplet distance (Critchlow et al., 1996), branch lengths-informed triplet distance (Kuhner and Yamato, 2015), and Kuhner-Felsenstein (KF). As different tree metrics are likely to have distinct strengths and weaknesses, metric choice should be considered when making inferences about the accuracy of phylogenetic reconstruction (Kuhner and Yamato, 2015).

In this chapter, I seek to answer some of the fundamental questions raised by the problems mentioned above, including: when implemented under typical user conditions (i.e.: by employing consensus tree construction and branch support estimation methods that are commonly used for each program), do different software implementations of the same optimality criterion (e.g. different ML software, IQ-TREE vs RAxML) result in considerable differences in performance?

Does performance improve with increased taxonomic sampling as it does for molecular datasets, despite increasing computational demands? If so, is there a threshold after which all methods converge towards the true tree? How does missing morphological data affect the performance of distinct probabilistic phylogenetic programs? Finally, how do different metrics impact overall assessments of phylogenetic performance? To answer these questions, I analyse simulated morphological data under various taxonomic sampling and missing data conditions, as implemented in four of the most widely used ML and BI software implementations for phylogenetic inference—RAxML (Stamatakis, 2014), IQ-TREE (Nguyen et al., 2015), MrBayes (Ronquist et al., 2012), and RevBayes (Höhna et al., 2016).

2.2 Materials and methods

The general approach consisted of generating true trees, then simulating morphological datasets that were analyzed using different phylogenetic methods, and finally evaluating the performance using the various metrics as outlined below. See Fig. 2-1 and Table 2-1 for a summary of the procedures and comparisons.

2.2.1 True tree simulations

I generated true tree topologies under the birth-death model using the *diversitree* (FitzJohn, 2012) package for R (RC Team, 2018). Speciation and extinction rates were sampled randomly from a uniform distribution conditioned on speciation rate being higher than extinction rate. To investigate the effects of taxon number on the performance of each method, I generated ten different true trees with the number of terminal taxa ranging from 50 to 500, using increments of 50. Using several simulated true tree topologies to assess phylogenetic performance reduces

biases that may be introduced by choosing a single tree topology to generate all dataset replicates (Rannala et al., 1998), and allows generation of trees with distinct shapes (such as various degrees of tree asymmetry) and branch lengths. Variation in the number of taxa also allows for an assessment of the impact of increased taxon sampling, which generally leads to more complex phylogenetic problems and is becoming increasingly common in the literature.

2.2.2 Dataset simulations

For each of the ten true trees, I simulated ten binary morphological datasets using the *phangorn* (Schliep, 2011) package for R (RC Team, 2018). I set the number of characters in each matrix to be 2.5 times the number of taxa (e.g., 125 characters for a 50-taxon topology), based on the average number of characters per taxon in empirical datasets (Scotland et al., 2003). Each character was simulated independently with a unique and randomly generated set of substitution rate and state frequency parameters. I generated variable characters to reflect the compositional bias of empirical morphological datasets, which usually include only traits that vary among taxa, while invariable traits are omitted.

I generated data matrices under the all-rates-different (ARD) model with substitution rates randomly sampled from an exponential distribution ($\lambda \in [0.05; 500]$) which allows for asymmetrical substitution rates violating the assumptions of a time-reversible Mk model (Lewis, 2001). Character state frequencies were randomly sampled from a uniform distribution for each character. To ensure that the datasets met empirically plausible levels of homoplasy, I only used datasets with ensemble consistency index (eCI) values between 0.26 and 1.0 (Sanderson et al., 1989). The eCI differs from individual character consistency index (cCI) in that it is an overall measure of character homoplasy in a dataset (Goloboff et al., 2018); while different datasets may

have similar eCI values, the distribution of individual character homoplasy may vary considerably between them. Thus, generating datasets with very different numbers of taxa and character composition will also introduce an important variation in both eCI and cCI values among datasets. Distributions of homoplasy across characters and the proportion of characters with $cCI < 1$ for the simulated datasets are shown in Figure 2-2.

Finally, I generated two additional categories of datasets that contained 20% and 50% missing data by randomly sampling and removing data entries from the original simulated matrices while ensuring that characters remained variable after data removal. This strategy does not necessarily reflect the empirically observed distribution of missing data in empirical datasets and alternative approaches have been used elsewhere (e.g., Wright and Hillis, 2014; Guillerme et al., 2016), but it is a necessary simplification for testing a large number of simulated datasets with variable numbers of taxa.

2.2.3 Phylogenetic analyses

For the phylogenetic analyses, I used settings commonly reported in the empirical analysis of morphological datasets using probabilistic methods and applied those settings to all datasets with minimal variation between them. This approach is a necessary simplification in order to produce representative sampling of results for each software; therefore, empirical analyses with customized or fine-tuned settings may behave differently from the patterns observed in this study. However, the most important parameters to maintain empirical realism were taken into account and adjusted if required. For instance, with every increase in the number of taxa under BI, I increased the number of generations to values commonly used in empirical morphological datasets (between 10-20 million generations) to ensure that all analyses would reach the

stationarity phase and attain convergence between runs. I used two independent runs, each with multiple chains, to assess convergence between runs, as is commonly done with empirical datasets using MrBayes and RevBayes. All phylogenetic analyses were conducted on the Cedar and Graham computing clusters available through Compute Canada.

Maximum likelihood analyses

Simulated datasets were analyzed in IQ-TREE v. 1.5.5 (Nguyen et al., 2015) and RAxML v.8.2.11 (Stamatakis, 2014). IQ-TREE is a time efficient ML-based software designed specifically to overcome the problem of multiple local optima during the tree search process. The software uses multiple starting trees (100 parsimony trees and a BIONJ tree by default) and iterative stochastic Nearest-Neighbour-Interchange (NNI) rearrangements of candidate trees in a greedy hill-climbing tree search algorithm (Nguyen et al., 2015). By retaining and updating the pool of candidate trees, IQ-TREE should achieve a more thorough exploration of tree space compared to other programs that use a single starting tree.

I used the Mk+R+ASC model in IQ-TREE analyses, which is a time-reversible model that assumes equal rates of character change and equal state frequencies (Lewis, 2001), but allows for free rate heterogeneity across sites, (+R option with four discrete rate categories, Yang 1995, Soubrier et al. 2012). The ASC flag corrects for any artificial increase in branch lengths due to the use of exclusively variable characters, which is known to cause an overestimation of divergence in phylogenetic analyses (Leache et al., 2015). I specified 1000 ultrafast bootstrap replicates (Minh et al., 2013) and 1000 SH-aLRT branch test replicates (Guindon et al., 2010). All bootstrap replicates were saved (-wbt command option) and used to construct 50% majority rule consensus trees in the ape package implemented in R. The use of bootstrapped consensus

trees minimizes bias across performance comparisons introduced by the difference in resolution of reconstructed topologies (Brown et al., 2017).

RAxML is another popular software for ML inference of phylogenetic trees with a special focus on large phylogenetic datasets (Stamatakis, 2006, 2014). The RAxML search algorithm uses a single randomized stepwise sequence addition parsimony tree (default function) as a starting tree for the subsequent heuristic hill-climbing tree search. The software employs a subtree pruning-and-regrafting (SPR) algorithm to rearrange candidate trees. This approach is more computationally intensive compared to the NNI approach implemented in IQ-TREE; therefore, to accelerate analyses, RAxML uses “lazy subtree rearrangements” with a limited rearrangement distance and omits likelihood calculations for the subtree branches if re-grafting results in a poor likelihood score at the insertion point (Stamatakis et al., 2005, 2007).

I specified the binary model for morphological data with gamma-distributed rate heterogeneity across sites and ascertainment bias correction (-m ASC_BINGAMMA --asc-corr lewis). I conducted a simultaneous ML search and a rapid bootstrap analysis with 1000 replicates in a single RAxML run (-f a). As with the IQ-TREE analyses, I summarized RAxML bootstrap replicates by constructing 50% majority rule consensus trees using the ape package for R.

Note that the bootstrap strategies implemented in IQ-TREE and RAxML use distinct algorithms, which may create a considerable disparity in the resulting consensus trees. The ultrafast bootstrapping implemented in IQ-TREE is a widely-used feature of this software which significantly speeds up the bootstrapping procedure (Minh et al., 2013). This gain in computational time efficiency, however, may come at a cost of overestimating branch support under certain conditions (Minh et al., 2013; Hoang et al., 2017). Although there is an option to use a standard bootstrapping approach in IQ-TREE, comparing the ultrafast method in IQ-TREE

to the standard approach implemented in RAxML represents an important variable that should be assessed using the RF, MS, and KF metrics so that it can be adequately considered by potential users of either software.

Bayesian inference analyses

Bayesian inference is a statistical approach that estimates posterior distributions of parameter values given the data, probabilistic model of evolution, and prior distributions of parameter values. In phylogenetics, BI uses Markov Chain Monte-Carlo (MCMC) algorithms to sample parameter space including a phylogenetic tree landscape. Unlike the hill-climbing algorithms used in ML software, MCMC sampling allows for “downhill” moves if the likelihood score of a proposed move is above a certain threshold. This non-greedy algorithm, implemented over many iterations, allows for a very thorough exploration of parameter space and avoids the problem of getting stuck on local optima. Furthermore, an even more efficient search is possible with the implementation of the Metropolis Coupled MCMC (MC3) approach in which the parameter value landscape is explored using additional “heated” chains that help in moving over the “valleys” of low likelihood scores while searching for a global optimum.

BI analyses were conducted using MrBayes v.3.2.6 (Ronquist et al., 2012) and RevBayes v.1.0.9 (Höhna et al., 2016). In MrBayes, the MC3 search is initiated from a random tree (unless a starting tree is specified) and proceeds using NNI, SPR (random and parsimony-biased) and tree-bisection-and-reconnection (TBR) by default for non-clock analyses. The rearrangement algorithm and how much time the program dedicates to each procedure can be modified by the user, but the default parameters were used here for better comparison with most published results using morphological data. Rate heterogeneity among characters was sampled from a gamma

distribution with four discrete categories, and the number of generations varied between 10 and 20 million generations. Two independent runs were performed for each analysis with either four or eight chains per run.

Like MrBayes, RevBayes starts with a random tree that is subsequently rearranged using NNI, SPR, or both algorithms, as specified by the user in a RevBayes script. The software is very flexible and allows users to specify how much weight the program should assign to each of the rearrangement algorithms if both are used in the analysis. Although RevBayes does not implement a TBR branch swapping algorithm, the iterative nature of the MCMC sampling using guided tree proposals (Höhna et al., 2012) and the slice-sampling incorporated in RevBayes (Besag et al., 1993) allow a thorough and efficient exploration of the phylogenetic tree space. In the benchmark assessment of RevBayes against two other BI phylogenetic software packages (BEAST and MrBayes), Höhna et al. (2016) demonstrated that RevBayes performed on par or better than basic implementations of BEAST and MrBayes in terms of computational time. I used the Mkv model for morphological data with four discrete categories of gamma distribution of the across-site rate heterogeneity. Each analysis had two independent runs with 10-20 million iterations.

For both MrBayes and RevBayes analyses, the convergence of independent runs was assessed across randomly selected outputs using the average standard deviation of split frequencies (ASDSF); all analyses that were assessed had reached convergence (ASDSF < 0.01). It is important to note that in order to focus on evaluating topology reconstructions both within and between optimality criteria, I opted to conduct non-clock BI analyses. Due to this, BEAST and BEAST2 (Bouckaert et al., 2014, Suchard et al., 2018) were not included in this study as they implement clock-based analyses only, and estimate time-calibrated trees before branch

lengths (total number of substitutions per branch) are calculated (Drummond and Rambaut, 2009; Drummond and Bouckaert, 2015; Bouckaert et al., 2019). Consequently, BEAST trees are not directly comparable to the non-clock analyses conducted using MrBayes and RevBayes.

In summary, I included 300 datasets in the study. I simulated ten true trees with ten replicates per tree, and then simulated morphological data matrices under three categories of missing data for each true tree. Each simulated dataset was analysed using the four phylogenetic programs described above, totalling 1200 individual analyses.

2.2.4 Performance assessment

To assess performance across different software, I evaluated: i) accuracy, based on the distance between the true tree and the estimated trees (measured with two different tree distance metrics, see below); ii) precision, based on the total range of tree distance values for each taxon sampling category (i.e., concentration of tree distance results around a single tree distance value); and iii) resolution, based on the proportion of fully resolved nodes in a tree relative to the number of nodes in a fully resolved topology with the same number of terminal taxa.

To calculate accuracy and precision, I used two distinct metrics implemented in TreeCmp (Bogdanowicz et al., 2012b): Robinson-Foulds (RF) distance (Robinson et al., 1981), and the Matching Splits (MS) metric (Bogdanowicz et al., 2012a). The RF metric measures the number of different bifurcations in the compared topologies (in the TreeCmp implementation, it is the total number of different splits in two trees divided by two). The MS metric is a measure of the minimum cost of matching bifurcations between two topologies. For both metrics, higher values (closer to 1) indicate a decrease in accuracy. The use of these two approaches allows assessment of whether different metrics affect the interpretation of phylogenetic accuracy and precision.

Because absolute tree distance increases with an increase in the number of taxa, comparing tree distance results across datasets with different numbers of taxa requires normalizing both the RF and MS metrics (Bogdanowicz et al., 2012b). Therefore, I report here only the normalized RF (nRF) and MS (nMS) metric values calculated in TreeCmp (raw distance values divided by pre-computed average distance values for random trees generated under the uniform model).

I additionally calculated KF (Kuhner and Felsenstein, 1994) distance implemented in the *phangorn* package (Schliep, 2011) for R. The KF metric calculates the sum of squares of the differences between the length of each branch in the two compared trees. Branches that are not shared between compared topologies are set to length 0. For trees with identical branch lengths, the KF distance is 0. Because there is no maximum possible value for the branch length metric to normalize across different tree topologies with variable numbers of taxa, I measured accuracy of the branch length reconstruction using a rank test; for each individual dataset replicate, phylogenetic inference programs were ranked according to the accuracy of the branch length reconstruction (1st rank = most accurate, 4th rank = least accurate). I then compared total rank scorings for each program across all dataset replicates in each of the three categories of missing data. To compare the performance assessment across different metrics (nRF, nMS, KF), I also performed rank tests for the nRF and nMS metric results. All comparative analyses used unrooted trees, and since resolution has a direct effect on the accuracy of metric estimations (Brown et al., 2017), I also calculated the relative resolution of each reconstructed topology.

2.3 Results

2.3.1 *Effect of the number of taxa*

Overall, accuracy and precision of tree reconstruction for each software implementation improved as the size of the dataset increased, irrespective of the amount of missing data or the metric used to measure tree distances (Fig. 2-3). When measured using nRF distance, BI software applications were more accurate than ML programs, regardless of the amount of missing data (Fig. 2-3; Table 2-2). MrBayes and RevBayes performed similarly and were the most accurate methods overall (Fig. 2-3; Table 2-2). IQ-TREE was substantially less accurate than RAxML in smaller datasets (50-100 taxa), although the accuracy of both ML applications improved considerably as taxon number increased.

Contrary to the results suggested by the nRF metric, IQ-TREE was always more accurate than other tree inference methods when measured by nMS (Fig. 2-3; Tables 2-3, 2-4, 2-5). Under this metric, RAxML was the least accurate, and BI software applications had an intermediate performance between that of IQ-TREE and RAxML. For complete datasets with 0% missing data, MrBayes had more accurate results than RevBayes. Disparity between the results of the two Bayesian software packages decreased as the proportion of missing data increased to 20% and 50%.

The median nRF values for the complete (0% missing data) small-sized datasets (50—100 taxa) varied from 0.223 ± 0.036 [MrBayes] to 0.341 ± 0.064 [IQ-TREE] (Table 2-2), while nRF values for the complete, large-sized datasets (450—500 taxa) ranged from 0.109 ± 0.007 [MrBayes] to 0.175 ± 0.014 [RAxML]; these results indicate a twofold increase in accuracy across methods as dataset size increased. In the 20% and 50% missing data categories, median nRF values were higher than for the corresponding complete datasets, indicating a relative

decrease in performance across all methods (Fig. 2-4). Similar to the nRF metric results, the median nMS values were the highest and had the greatest disparity among software for the small complete datasets (50—100 taxa). These values ranged from 0.165 ± 0.042 [IQ-TREE] to 0.305 ± 0.065 [RAxML], but decreased significantly as dataset size increased. For large matrices (450—500 taxa), nMS values ranged from 0.038 ± 0.007 [IQ-TREE] to 0.133 ± 0.022 [RAxML]. According to the nMS metric, datasets with 20% missing data were similar in performance to datasets with no missing data, and datasets with 50% missing data generally had the highest nMS values and were the least accurate (Fig. 2-4).

The precision of each method, as measured by the standard deviation of the nRF and nMS values, also improved as dataset size increased (Figs. 2-5 and 2-6, Tables 2-2 and 2-3). The nRF metric results indicate that RAxML had the highest overall precision for datasets with 50—200 taxa, while for larger datasets (250—500 taxa), MrBayes and RevBayes had the highest precision. However, nMS distances instead suggest that IQ-TREE was the most precise implementation in most cases. The standard deviation across all methods decreased up to 13 times (RAxML, 0% missing data) with an increase in dataset size from 50 to 500 taxa (Table 2-6). Across 20% and 50% missing data categories, RAxML was the most precise according to the nRF values (Fig. 2-5), although the performance of MrBayes and RevBayes was very similar to RAxML. IQ-TREE had the lowest precision according to the nRF values across all missing data categories (Table 2-6). Precision of the nMS metric was generally higher for the IQ-TREE results, followed by MrBayes. RAxML was less precise across all missing data categories (Fig. 2-6).

2.3.2 *Effect of missing data*

As expected, increasing the percentage of missing data resulted in reduced accuracy and precision of the tree topology reconstruction across all methods under both nRF and nMS metrics. This decrease in performance was the greatest when the proportion of missing data reached 50%, while the difference in performance was not as drastic between the 0% and 20% missing data categories. The nRF metric was more sensitive to missing data and had a greater decrease in accuracy and precision across the three classes of missing data than the nMS metric (Fig. 2-4). In general, the difference in both accuracy and precision among datasets with different proportions of missing data decreased with greater taxon sampling. Under the nMS metric, the accuracy of phylogenetic reconstructions using datasets with 50% missing data greatly improved as the taxon sampling increased from 50 to 150 taxa, and approximated the level of performance of datasets with 0% and 20% missing data for large datasets (200-500 taxa) for all software (Fig. 2-4). Under the nRF metric, however, no increase in taxonomic sampling for datasets with 50% missing data provided an equally accurate phylogenetic estimate as datasets with 0-20% missing data (Fig. 2-4).

For high amounts of missing data (50%), BI programs outperformed ML software in accuracy and precision when measured by nRF across all categories of taxonomic sampling (Fig. 2-3, Table 2-2). In datasets with no missing data, MrBayes was ranked as the best performing software in 85% of the analyses, with RevBayes being the second best program in 84% of the analyses (Fig. 2-7). In datasets with 20% and 50% missing data, BI programs remained the best performing; however, there was no clear distinction in the rank scoring between RevBayes and MrBayes (Fig. 2-7; Table 2-8). In contrast, IQ-TREE constantly had the worst performance among all software at 50% missing data, especially at smaller dataset sizes (with 50—150 taxa).

Finally, RAxML had a very similar performance to IQ-TREE for datasets with 0% and 20% missing data; however as the amount of missing data increased to 50%, RAxML outperformed IQ-TREE (Figs. 2-3, 2-7).

Under the nMS metric and as amount of missing data increased, BI programs no longer consistently outperformed the ML programs (Fig. 2-4, Tables 2-3, 2-9). Instead, IQ-TREE was ranked as the best performing program in over 80% of analysed datasets across all categories of missing data (Fig. 2-7). When missing data was high (50%), IQ-TREE performed so well under nMS that the 50th percentile mark around the median did not overlap with the distributions of any other method (Fig. 2-3). MrBayes and RevBayes were ranked as the second and third best performing programs, respectively, over 70% of the time and performed similarly regardless of the number of taxa sampled. RAxML was consistently the worst performing method according to the nMS metric and ranked as the least accurate method in over 90% of analyses (Figs. 2-3, 2-5, 2-7; Table 2-9).

2.3.3 Branch lengths

Accuracy of the branch length reconstructions (KF metric) showed a different pattern than that observed for the topology-only metrics (nRF and nMS). Under this metric, the ML programs ranked as the best performing in all categories of the missing data (Fig. 2-7; Table 2-10); RAxML was ranked as the most accurate program in over 90% of analyses of datasets with 0% and 20% missing data, and IQ-TREE was the best performing program in over 70% of the analyses in the 50% missing data category. BI programs performed consistently as the second (MrBayes) and fourth (RevBayes) best performing programs across all categories of missing data. In the 0% and 20% missing data categories, the performance of MrBayes was split nearly

evenly among the second, third, and fourth ranks, whereas in the 50% missing data category, branch length accuracy of MrBayes was ranked as either first or second in 28% and 72% of the analyses respectively. RevBayes branch lengths estimations were most frequently ranked as the least accurate across all categories of missing data (Fig. 2-7; Table 2-10).

2.4 Discussion

2.4.1 The performance of different probabilistic methods and software

This study provides the first test of the performance of different probabilistic phylogenetic software packages for morphological datasets with variable degrees of missing data and a variety of tree topologies. The results indicate that even within a single optimality criterion method (e. g. ML), there can be a great disparity in results across different software packages. There was notably less disparity between results of the two BI programs tested in the study compared to the ML software. MrBayes and RevBayes generally performed with very similar accuracy under both nRF and nMS tree distance metrics.

The nRF metric results show that BI programs had an overall consistently better performance than ML programs across all categories of missing data and taxonomic sampling. This suggests that the Metropolis-Coupled Markov chain Monte Carlo method implemented in MrBayes (MC3), and even the regular Markov chain Monte Carlo method used in RevBayes, allow Markov chains to efficiently explore multiple local optima of different parameter values.

As previously observed by increasing the number of characters in RAxML (Wright and Hillis, 2014; O'Reilly et al., 2016; Puttick et al., 2017b; O'Reilly et al., 2018a), increasing the number of taxa while keeping a constant character-to-taxon ratio also brings the performance of ML programs closer to the performance of BI programs. However, this is dependent on the ML

software used and the amount of missing data. Whereas under the nRF metric, RAxML results were generally closer to the performance of the BI programs, particularly for small datasets with 50 taxa, IQ-TREE results never reached the same level of accuracy, especially for datasets with 50% missing data. In sharp contrast, IQ-TREE had the best overall performance using the nMS metric.

2.4.2 Effects of increased taxon sampling on phylogenetic performance

Previous studies investigating the effect of taxonomic sampling on phylogenetic performance have suggested that increasing taxonomic sampling in molecular datasets has an overall positive effect on the accuracy of phylogenetic methods (Hillis, 1996, 1998; Pollock et al., 2002; Zwickl et al., 2002; Hillis et al., 2003). Results of this simulation study unambiguously indicate that increased taxon sampling in morphological datasets also leads to improved accuracy and precision in all four of the software implementations assessed in this study. I note that these findings hold true when taxonomic sampling is increased together with the number of characters (a condition tested in the present study). Considerable improvement occurs until the addition of 200 taxa, after which slower but still detectable improvement continues up to the highest levels of taxonomic sampling tested in the study (500 taxa).

Despite the overall advantages of increasing taxon sampling, it is important to acknowledge that different approaches to taxon sampling may have quite distinct effects on the overall performance of phylogenetic analyses. For example, previous studies using molecular data suggest that phylogenetic accuracy may be improved by increasing taxon sampling within each monophyletic group, rather than adding stem taxa that increase the age of the root (Lecointre et al., 1993; Rannala et al., 1998). Therefore, it is not possible to choose a single taxon that can be used as a representative of an entire monophyletic group, and more intensive

taxonomic sampling of each group of interest is preferred. Limited taxon sampling may also obscure results due to long branch attraction (Swofford et al., 1996; Poe et al., 2000; Wiens, 2006), especially when combined with large amounts of missing data (Poe et al., 2000). The effects of increased taxon sampling might be especially beneficial if newly added taxa contribute to reducing branch lengths at the tips of the tree rather than breaking internal branches (Kim, 1998).

2.4.3 The impact of different distance metrics

I used three distinct tree distance metrics (KF, RF, and MS) to assess accuracy of topological reconstructions. In an extensive performance study of different tree comparison metrics, Kuhner and Yamato (2015) noted that branch-length metrics such as KF are extremely vulnerable to saturation when comparing highly dissimilar trees and should instead be used to compare trees with very similar topologies. Trees with very distinct topologies are better assessed with metrics focused on topological differences, such as RF and MS (Kuhner and Yamato, 2015). In the present study, I was comparing true trees with the inferred consensus topologies that, in the majority of cases, contained polytomies and were considerably different from the fully resolved true trees, and contained additional differences stemming from the simulations using various eCIs. Therefore, considering the low accuracy of the branch length metrics when comparing dissimilar trees and the problem with standardizing this type of metric across trees with various number of taxa, I focused this study and discussion primarily on assessing accuracy using the topology only-based nRF and nMS metrics, as these are more appropriate metrics for a broad-scale performance study.

Many implementations of the RF metric (including the one in TreeCmp, used in the present study) assign a distance value of 1 for different bipartitions and a value of 0 for identical

bipartitions. When one of the trees has a polytomy that is consistent with a taxon bipartition on the true tree, a distance value of 0.5 is given to that polytomy. Polytomies are treated as half correct, thus benefiting methods which provide a consensus tree that does not try to resolve poorly supported nodes. This is why comparisons of majority rule consensus trees from BI analyses with fully resolved maximum likelihood trees from ML analyses will tend to favor the results of BI as more accurate under the RF metric (Brown et al., 2017). The fully resolved ML trees will necessarily have nodes that are fully resolved despite having very low support, and probably will not match taxon bifurcations on the true tree. Therefore, although RF distances can measure both Type I and Type II errors, the treatment of both types of error is only comparable for equally resolved topologies. When topologies have very different resolutions, the RF metric will penalize topologies with false positives more heavily than topologies with false negatives, as previously discussed by Zwickl et al. (2002) and Holder et al. (2008).

This behaviour of the RF metric explains why estimated trees that had generally higher resolution (e. g. IQ-TREE bootstrap majority rule consensus trees) were consistently farther away from the true trees compared to all other phylogenetic methods when accuracy was assessed using nRF. The higher resolution of the IQ-TREE consensus trees, followed by their relatively poor accuracy under the nRF metric, indicate those trees had more poorly supported but well-resolved nodes that are absent in the true trees (false positives). Those trees with incorrect but fully resolved nodes were more severely penalized by the nRF metric than the polytomies in the less resolved consensus trees of all other methods.

Interestingly, the accuracy of RevBayes and MrBayes was overall better than the accuracy of RAxML despite the resolution of the BI trees being slightly higher than the resolution of the consensus trees obtained by RAxML (Fig.2-5). This indicates that the resolved

nodes in the BI analyses were correctly estimated, yielding better nRF values compared to the results from the ML analyses.

The results obtained from the nMS metric indicate that the treatment of Type I and Type II error is opposite to that of the RF metric. The results from the IQ-TREE estimation, which have higher resolution than the other results, always have lower nMS values (higher accuracy; Fig. 2-6). This indicates that trees with a higher number of polytomies tend to have lower accuracy under nMS, which is a previously unacknowledged behavior of this metric. To exemplify the treatment of polytomies under the MS metric, consider the tree comparisons in Figure 2-8, which illustrates an example based on Bogdanowicz et al. (2012a, 2012b). The MS metric will always calculate the minimum-weight perfect matching between any given trees. When trees are fully resolved (tree comparisons T1 and T2 in Fig. 2-8), the matching costs (h_s) and subsequent MS distance (dMS) value are straightforward. In the example in Figure 2-8, the matching costs are: $h_s(ab|cde, ac|bde) = 2$, $h_s(abc|de, acd|be) = 2$, $h_s(ab|cde, acd|be) = 2$; $h_s(abc|de, ac|bde) = 1$. In this situation, the minimum-weight perfect matching between those two trees (dMS) = 3. In the second example, containing a polytomy (tree comparisons T1 and T3), the metric is modified to include a “dummy” element (O), so that the number of splits in the smaller set matches the number of splits in the largest set. The cost of the matching splits with the addition of the dummy element is based on the minimum possible cost to match the split that it is being compared to: $h_s(A|B, O) = \min(A|B)$. Therefore, the perfect matching costs for the second example are: $h_s(ab|cde, acd|be) = 2$, $h_s(ab|cde, O) = 2$, $h_s(abc|de, acd|be) = 2$, $h_s(abc|de, O) = 2$. In this case, the dMS = 4, as any matching pairs will have a minimum cost of 4 regardless of the matching splits that are chosen. In a third situation, if the polytomy on the second tree involved taxa b, d, and e, (tree

comparisons T1 and T4) then the matching splits would be: $h_s(ab|cde, ac|bde) = 2$, $h_s(ab|cde, O) = 2$, $h_s(abc|de, ac|bde) = 1$, $h_s(abc|de, O) = 2$; and $dMS = 3$. Therefore, depending on which node is collapsed in tree T2, the MS distance can be equal to or higher than the MS distance of the original tree before collapsing any of its nodes. Under the MS metric, tree comparisons in which one of the trees includes polytomies and another tree is fully resolved will always yield MS values that are equal to or higher than the MS value that would be obtained if the polytomic tree was in fact fully resolved. This is exactly the scenario in studies that compare a fully resolved true tree to a series of consensus trees, which will always contain some polytomies that differ between datasets. When several consensus trees are compared to fully resolved true trees, the overall balance is that at least some of the trees with polytomies will have a dMS that is higher than the dMS would have been if that tree was fully resolved, thus explaining the results obtained herein.

Overall, results of the study indicate that the RF metric penalizes trees containing well-resolved but incorrectly inferred nodes (higher false positives), while the MS metric penalizes trees that contain a higher number of polytomies (higher false negatives). It has been argued that the MS distance is a refinement of the RF distance, as it takes into account more subtle differences between trees (Bogdanowicz et al., 2012a, 2012b). While this seems to be true in some situations in which trees are, for example, fully resolved (Bogdanowicz et al., 2012b), this is not the case when a fully resolved tree is compared to a tree with polytomies. Type I and Type II error have different implications from a systematic point of view. Making false assumptions of species relationships is generally considered worse than making no assumptions (with a risk of not detecting potentially true sister-group relationships) (Zwickl et al., 2002; Holder et al., 2008; O'Reilly and Donoghue, 2018). Therefore, false positives are considered a bigger problem than

false negatives in systematics, an assumption that has also led to recommendations on the use of consensus trees over fully resolved trees (e.g., maximum clade credibility trees) (Holder et al., 2008; O'Reilly and Donoghue, 2018). Because no existing metric can completely avoid the biases created by tree resolution, methods that penalize Type I errors more than Type II errors should be preferred, making the RF metric preferable to MS in this particular context.

2.4.4 Considerations for morphological data simulations

As with any simulation study, the present simulations have limitations in how close they approach empirical morphological datasets. My simulation design does not necessarily represent a realistic scenario of morphological trait evolution, but rather represents a complex scenario in which assumptions of the Mk model (used to analyze morphological phylogenetic datasets) are violated. This approach tests the performance of the four phylogenetic inference programs assessed under model misspecification in which the true model (ARD) is more complex than the inference model (Mk). I argue that this approach approximates a realistic scenario commonly encountered in morphological phylogenetics: because there is no available evolutionary model that adequately captures the intricacies of morphological evolutionary processes in nature, there is always some degree of model misspecification during empirical phylogenetic analysis. Several recent studies have published new morphological models (Goloboff et al., 2017; Pyron, 2017; Puttick et al., 2019), and additional ones are forthcoming (e.g., Vernygora, in review; Simões et al., in prep.), that will improve both analysis and simulation of morphological datasets. Yet, it will likely take many years until models come close to the complexities embedded in phenotypic evolution (e.g. integrating the effect of phenotypic and epigenetic change). Thus, having some degree of model mismatch between data generation and subsequent tree inference provides some

empirical realism in the absence of sufficiently sophisticated morphological evolutionary models.

Performance results based on simulations are not deterministic indicators regarding the utilization of different methodological approaches. Ongoing refinements to simulation procedures and evolutionary models will continue to improve the quality of simulation-based performance studies themselves. Further testing various analytical conditions such as different dataset simulation strategies, tree comparison metrics, and different types of the summary trees are necessary to investigate more specific behaviour of the phylogenetic programs under various conditions but are beyond the scope of this study.

2.5 Conclusions

This study explores the performance of four commonly used probabilistic programs (IQ-TREE, RAxML, MrBayes, and RevBayes) under a wide variety of taxonomic sampling and missing data conditions using different tree distance metrics. Simulation-based approaches to performance evaluation offer the only way to assess the accuracy and error of phylogenetic methods because the true tree is known, allowing a straightforward assessment of the accuracy. Therefore, simulation studies provide necessary guidance to practicing systematists.

Results of the study indicate that increased taxonomic sampling unambiguously improves accuracy and precision of phylogenetic analyses of morphological data for up to 500 taxa using probabilistic methods, despite increased computational burden. The greatest improvement in performance was observed as the number of taxa increased from 50 to 200. The BI phylogenetic programs tested in the study had very similar performances, and yielded more consistent and accurate results compared to ML programs across a broad range of dataset sizes under the

condition of minimizing the recovery of false positive sister-group relationships. More importantly, BI was more robust to increases in missing data, which is a realistic condition for empirical datasets that are almost never complete, especially datasets containing fossil taxa. The Bayesian inference implementations of the simple Mk model outperforms all other methods, even the ML algorithms designed specifically to perform a more thorough search of the parameter space.

References

- Besag, J. and P. J. Green. 1993. Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society B* 55:25–37.
- Bogdanowicz, D. and K. Giaro 2012a. Matching split distance for unrooted binary phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9:150–160.
- Bogdanowicz, D., K. Giaro, and B. Wróbel. 2012b. TreeCmp: comparison of trees in polynomial time. *Evolutionary Bioinformatics* 8:EBO.S9657.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10:e1003537.
- Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, and M. Matschiner. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15:p.e1006650.
- Brown, J. W., C. Parins-Fukuchi, G. W. Stull, O. M. Vargas, and S. A. Smith. 2017. Bayesian and likelihood phylogenetic reconstructions of morphological traits are not discordant when taking uncertainty into consideration: a comment on Puttick et al. *Proceedings of the Royal Society B* 284:p.20170986.
- Conrad, J. L. 2008. Phylogeny and systematics of Squamata (Reptilia) based on morphology. *Bulletin of the American Museum of Natural History* 310:1–182.
- Critchlow, D. E., D. K. Pearl, and C. Qian. 1996. The triples distance for rooted bifurcating phylogenetic trees. *Systematic Biology* 45:323–334.

- Felsenstein, J. 1978. The number of evolutionary trees. *Systematic Zoology* 27:27–33.
- FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3:1084–1092.
- Goloboff, P. A. 1999. Analyzing large datasets in reasonable times: solutions for composite optima. *Cladistics* 15:415–428.
- Goloboff, P. A. and J. S. Arias. 2019. Likelihood approximations of implied weights parsimony can be selected over the Mk model by the Akaike information criterion. *Cladistics* 35:695–716.
- Goloboff, P. A. and S. A. Catalano. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* 32:221–238.
- Goloboff, P. A., A. Torres, and J. S. Arias. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* 34:407–437.
- Guillerme, T. and N. Cooper. 2016. Effects of missing data on topological inference using a total evidence approach. *Molecular Phylogenetics and Evolution* 94:146–158.
- Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59:307–321.
- Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* 383:130.
- Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47:3–8.
- Hillis, D. M., D. D. Pollock, J. A. McGuire, and D. J. Zwickl. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Systematic Biology* 52:124–126.

- Höhna, S. and A. J. Drummond. 2012. Guided tree topology proposals for Bayesian phylogenetic inference. *Systematic Biology* 61:1–11.
- Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65:726–736.
- Holder, M. T., J. Sukumaran, and P. O. Lewis. 2008. A justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. *Systematic Biology* 57:814–821.
- Huelsenbeck, J. P. 1995. Performance of phylogenetic methods in simulation. *Systematic Biology* 44:17–48.
- Kim, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Systematic Biology* 47:43–60.
- Kitching, I. J., P. L. Forey, C. J. Humphries, and D. D. William. 1998. *Cladistics: the theory and practise of parsimony analysis* (2nd edition). Oxford University Press, USA, 228 pp.
- Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* 11:459–468.
- Kuhner, M. K. and J. Yamato. 2015. Practical performance of tree comparison metrics. *Systematic Biology* 64:205–214.
- Leache, A. D., B. L. Banbury, J. Felsenstein, A. N. M. de Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP Phylogenies. *Systematic Biology* 64:1032–1047.
- Lecointre, G., H. Philippe, H. L. V. Lê, and H. Le Guyader. 1993. Species sampling has a major impact on phylogenetic inference. *Molecular Phylogenetics and Evolution* 2:205–224.

- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Minh, B. Q., M. A. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30:1188–1195.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274.
- Novacek, M. and Q. Wheeler. 1992. Extinct taxa: Accounting for 99.999% of the earth's biota; pp. 1–16 in M. Novacek and Q. Wheeler (eds.), *Extinction and Phylogeny*. New York, Columbia University Press, 253 pp.
- Nye, T. M. W., P. Lio, and W. R. Gilks. 2006. A novel algorithm and web based tool for comparing two alternative phylogenetic trees. *Bioinformatics* 22:117–119.
- O'Reilly, J. E. and P. C. J. Donoghue. 2018. The efficacy of consensus tree methods for summarizing phylogenetic relationships from a posterior sample of trees estimated from morphological data. *Systematic Biology* 67:354–362.
- O'Reilly, J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* 61:105–118.
- O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biological Letters* 12:20160081.

- Poe, S., and J. J. Wiens. 2000. Character selection and the methodology of morphological phylogenetics; pp. 20–36 in J. J. Wiens (ed.), *Phylogenetic analysis of morphological data*. Washington, D.C., Smithsonian Institution Press, 232 pp.
- Pollock, D. D., D. J. Zwickl, J. A. McGuire, D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51:664–671.
- Puttick, M. N., J. E. O'Reilly, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, and P. C. J. Donoghue. 2017. Uncertainty: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B: Biological Sciences* 284:p.20162290.
- Puttick, M. N., J. E. O'Reilly, D. Pisani, and P. C. J. Donoghue. 2019. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology* 62:1–17.
- Pyron, R. A. 2017. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Systematic Biology* 66:38–56.
- Pyron, R. A., F. T. Burbrink, and J. J. Wiens. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology* 13:10.1186/1471-2148-1113-1193.
- Rannala, B., J. P. Huelsenbeck, Z. Yang, and R. Nielsen. 1998. Taxon sampling and the accuracy of large phylogenies. *Systematic Biology* 47:702–710.
- Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53:131–147.

- 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* 61:539–542.
- Ronquist, F., N. Lartillot, and M. J. Phillips. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society. B* 371:20150136.
- Sanderson, M. J. and M. J. Donoghue. 1989. Patterns of Variation in Levels of Homoplasy. *Evolution* 43:1781–1795.
- Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27:592–593.
- Schrago, C. G., B. O. Aguiar, and B. Mello. 2018. Comparative evaluation of maximum parsimony and Bayesian phylogenetic reconstruction using empirical morphological data. *Journal of Evolutionary Biology* 31:1477–1484.
- Scotland, R. W., R. G. Olmstead, and J. R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology* 52:539–548.
- Simões, T. R., M. W. Caldwell, A. Palci, and R. L. Nydam. 2018a. Giant taxon-character matrices II: a response to Laing et al. (2017). *Cladistics* 34:702–707.
- Simões, T. R., M. W. Caldwell, M. Tañanda, M. Bernardi, A. Palci, O. Vernygora, F. Bernardini, L. Mancini, and R. L. Nydam. 2018b. The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature* 557:706–709.
- Soubrier, J., M. Steel, M. S. Y. Lee, C. Der Sarkissian, S. Guindon, S. Y. W. Ho, and A. Cooper. 2012. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution*. 29:3345–3358.

- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stamatakis, A., F. Blagojevic, D. S. Nikolopoulos, and C. D. Antonopoulos. 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM Cell. *J VLSI Signal Process* 48:271–286.
- Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.
- Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4:vey016.
- Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference; pp. 407–514 in D. M. Hillis, C. Moritz, B. K. Mable (eds.), *Molecular systematics*, 2nd edition. Sinauer Associates, Sunderland, Massachusetts, 655 pp.
- Team, R. C. 2018. R: A Language and Environment for Statistical Computing.
- Whidden, C. and F. A. Matsen IV. 2018. Efficiently inferring pairwise subtree prune-and-regraft adjacencies between phylogenetic trees; pp. 77–91 in 2018 Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO). Society for Industrial and Applied Mathematics, 177 pp.
- Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39:34–42.

- Wright, A. M. and D. M. Hillis. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* 9:e109210.
- Yang, Z. 1995. A space-time process model for the evolution of DNA sequences. *Genetics* 139:993–1005.
- Zhang, C., T. Stadler, S. Klopstein, T. A. Heath, and F. Ronquist. 2016. Total-evidence dating under the fossilized birth–death process. *Systematic Biology* 65:228–249.
- Zhou, X., X. X. Shen, C. T. Hittinger, and A. Rokas. 2018. Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic datasets. *Molecular Biology and Evolution* 35:486–503.
- Zwickl, D. J. and D. M. Hillis. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51:588–598.

TABLE 2-1. Summary of simulation parameters and phylogenetic analyses settings.

Abbreviations: ARD, all rates different model; ASC, ascertainment bias correction; NNI, nearest-neighbour-interchange; SPR, subtree-pruning-and-grafting; TBR, tree-bisection-and-reconnection.

Parameter	Simulation	Analysis			
		IQ-TREE	RAxML	MrBayes	RevBayes
Tree Topologies	Variable: 10 topologies	Uniform + Topology search: NNI	Uniform + Topology search: SPR	Uniform + Topology search: NNI+SPR+TBR	Uniform + Topology search: NNI+SPR
Tree model/Branch Lengths (V)	Birth-Death $b-d \in U(0,1)$; $b>d$	V optimized to maximize tree likelihood	V optimized to maximize tree likelihood	Unconstrained: $V \in \Gamma$ -Dirichlet (1.0,0.100,1.0,1.0)	Unconstrained: $V \in \exp(\lambda = \exp(0.2))$
Character variability model	Variable-only	Variable-only + ASC	Variable-only + ASC	Variable-only + ASC	Variable-only + ASC
Character substitution rates (μ)	Variable: $\mu \in \exp[0.05;500]$	Fixed: 1.0	Fixed: 1.0	Fixed: 1.0	Fixed: 1.0
Rate variation among characters ($C\Delta\mu$)	Variable: ARD model	Variable: $C\Delta\mu \in \Gamma$ 4 discrete categories	Variable: $C\Delta\mu \in \Gamma$ 4 discrete categories	Variable: $C\Delta\mu \in \Gamma$ [mean $\in \exp(1.0)$] 4 discrete categories	Variable: $C\Delta\mu \in \Gamma$ [mean $\in U(0.0, 1E6)$] 4 discrete categories
Character state frequencies (π)	Variable: $\pi \in U(0,1)$	Fixed: 0=0.5, 1=0.5	Fixed: 0=0.5, 1=0.5	Fixed: 0=0.5, 1=0.5	Fixed: 0=0.5, 1=0.5

TABLE 2-4. Normalized RF (uniform) by number of taxa (median values)

Taxa	0% missing data				20% missing data				50% missing data			
	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes
50	0.320	0.267	0.230	0.219	0.416	0.331	0.288	0.294	0.683	0.448	0.438	0.432
100	0.369	0.276	0.222	0.235	0.408	0.344	0.287	0.284	0.661	0.454	0.413	0.398
150	0.266	0.232	0.186	0.201	0.300	0.281	0.226	0.242	0.517	0.415	0.366	0.356
200	0.216	0.230	0.161	0.175	0.272	0.271	0.206	0.213	0.439	0.384	0.323	0.328
250	0.166	0.235	0.132	0.148	0.211	0.271	0.173	0.176	0.383	0.364	0.291	0.289
300	0.227	0.228	0.161	0.171	0.285	0.272	0.207	0.180	0.470	0.388	0.336	0.343
350	0.226	0.222	0.158	0.162	0.275	0.278	0.204	0.205	0.453	0.394	0.321	0.323
400	0.202	0.227	0.152	0.159	0.265	0.275	0.195	0.192	0.446	0.380	0.319	0.318
450	0.143	0.190	0.109	0.120	0.185	0.231	0.139	0.138	0.311	0.323	0.245	0.250
500	0.147	0.170	0.110	0.119	0.180	0.207	0.144	0.146	0.322	0.321	0.255	0.244

TABLE 2-5. Normalized MS (uniform) by number of taxa (median values)

Taxa	0% missing data				20% missing data				50% missing data			
	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes
50	0.176	0.319	0.204	0.231	0.216	0.411	0.230	0.248	0.297	0.541	0.425	0.468
100	0.139	0.305	0.175	0.179	0.150	0.331	0.193	0.200	0.190	0.429	0.284	0.296
150	0.073	0.192	0.112	0.159	0.078	0.197	0.118	0.172	0.113	0.292	0.178	0.175
200	0.069	0.200	0.084	0.118	0.070	0.203	0.092	0.128	0.098	0.252	0.125	0.137
250	0.047	0.213	0.066	0.106	0.050	0.213	0.076	0.084	0.074	0.257	0.102	0.107
300	0.072	0.209	0.093	0.145	0.078	0.217	0.112	0.142	0.095	0.261	0.140	0.169
350	0.051	0.161	0.076	0.087	0.050	0.172	0.080	0.083	0.073	0.203	0.105	0.119
400	0.052	0.171	0.079	0.097	0.055	0.181	0.082	0.082	0.074	0.224	0.102	0.115
450	0.039	0.149	0.046	0.090	0.044	0.154	0.052	0.050	0.053	0.177	0.061	0.084
500	0.035	0.123	0.047	0.089	0.039	0.124	0.050	0.054	0.051	0.151	0.062	0.069

TABLE 2-6. Normalized RF (uniform) by number of taxa (STDEV.S values)

Taxa	0% missing data				20% missing data				50% missing data			
	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes
50	0.077	0.047	0.045	0.047	0.097	0.042	0.062	0.063	0.086	0.029	0.025	0.020
100	0.051	0.022	0.025	0.027	0.045	0.017	0.022	0.019	0.055	0.019	0.028	0.029
150	0.030	0.013	0.014	0.020	0.041	0.019	0.021	0.019	0.055	0.021	0.026	0.023
200	0.024	0.013	0.017	0.014	0.025	0.012	0.021	0.021	0.033	0.016	0.012	0.011
250	0.022	0.018	0.008	0.012	0.034	0.013	0.016	0.021	0.034	0.017	0.020	0.018
300	0.024	0.015	0.009	0.008	0.023	0.014	0.021	0.010	0.018	0.017	0.013	0.009
350	0.021	0.020	0.012	0.012	0.013	0.016	0.011	0.012	0.026	0.014	0.016	0.011
400	0.019	0.014	0.012	0.015	0.018	0.011	0.014	0.014	0.019	0.015	0.009	0.022
450	0.006	0.012	0.008	0.011	0.019	0.017	0.012	0.014	0.026	0.013	0.017	0.019
500	0.014	0.004	0.006	0.005	0.016	0.007	0.006	0.008	0.013	0.008	0.009	0.056

TABLE 2-7. Normalized MS (uniform) by number of taxa (STDEV.S values)

Taxa	0% missing data				20% missing data				50% missing data			
	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes	IQ-Tree	RAXML	MrBayes	RevBayes
50	0.032	0.087	0.062	0.063	0.039	0.075	0.081	0.092	0.064	0.016	0.055	0.056
100	0.047	0.031	0.012	0.012	0.029	0.020	0.024	0.027	0.043	0.029	0.032	0.044
150	0.023	0.031	0.019	0.043	0.024	0.033	0.018	0.026	0.023	0.023	0.032	0.032
200	0.018	0.025	0.016	0.032	0.020	0.029	0.015	0.035	0.023	0.025	0.020	0.020
250	0.008	0.012	0.014	0.025	0.010	0.016	0.014	0.018	0.008	0.017	0.013	0.011
300	0.011	0.025	0.015	0.029	0.015	0.026	0.012	0.019	0.012	0.021	0.013	0.022
350	0.008	0.017	0.008	0.009	0.009	0.016	0.009	0.007	0.009	0.013	0.010	0.012
400	0.009	0.016	0.011	0.015	0.010	0.017	0.011	0.009	0.010	0.012	0.012	0.016
450	0.006	0.019	0.011	0.022	0.007	0.017	0.012	0.011	0.006	0.012	0.015	0.018
500	0.008	0.012	0.007	0.017	0.007	0.015	0.008	0.010	0.007	0.010	0.006	0.023

TABLE 2-8. Results of the rank test using the nRF values for each data set replicate.

Phylogenetic inference programs were ranked from the best (1st rank) to the worst performing (4th rank) based on the accuracy of the reconstructed topologies. Numbers indicate percentage of the total number of analyses in which programs scored at each rank.

	0% missing data				20% missing data				50% missing data			
	1	2	3	4	1	2	3	4	1	2	3	4
MrBayes	85	12	2	1	52	44	3	1	47	52	1	0
RevBayes	14	84	2	0	47	51	2	0	52	46	2	0
IQ-TREE	1	2	58	39	0	2	48	50	0	0	16	84
RAxML	0	2	38	60	1	3	47	49	1	2	81	16

TABLE 2-9. Results of the rank test using the nMS values for each data set replicate.

Phylogenetic inference programs were ranked from the best (1st rank) to the worst performing (4th rank) based on the accuracy of the reconstructed topologies. Numbers indicate percentage of the total number of analyses in which programs scored at each rank.

	0% missing data				20% missing data				50% missing data			
	1	2	3	4	1	2	3	4	1	2	3	4
MrBayes	13	83	4	0	8	70	22	0	5	71	24	0
RevBayes	1	6	88	5	2	22	71	5	1	23	76	0
IQ-TREE	86	11	3	0	90	8	2	0	94	6	0	0
RAxML	0	0	5	95	0	0	5	95	0	0	0	100

TABLE 2-10. Results of the rank test using the KF values for each data set replicate.

Phylogenetic inference programs were ranked from the best (1st rank) to the worst performing (4th rank) based on the accuracy of the reconstructed topologies. Numbers indicate percentage of the total number of analyses in which programs scored at each rank.

	0% missing data				20% missing data				50% missing data			
	1	2	3	4	1	2	3	4	1	2	3	4
MrBayes	0	46	26	28	0	35	33	32	28	72	0	0
RevBayes	0	31	1	68	9	22	15	54	0	0	3	97
IQ-TREE	0	23	73	4	0	34	52	14	72	28	0	0
RAxML	100	0	0	0	91	9	0	0	0	0	97	3

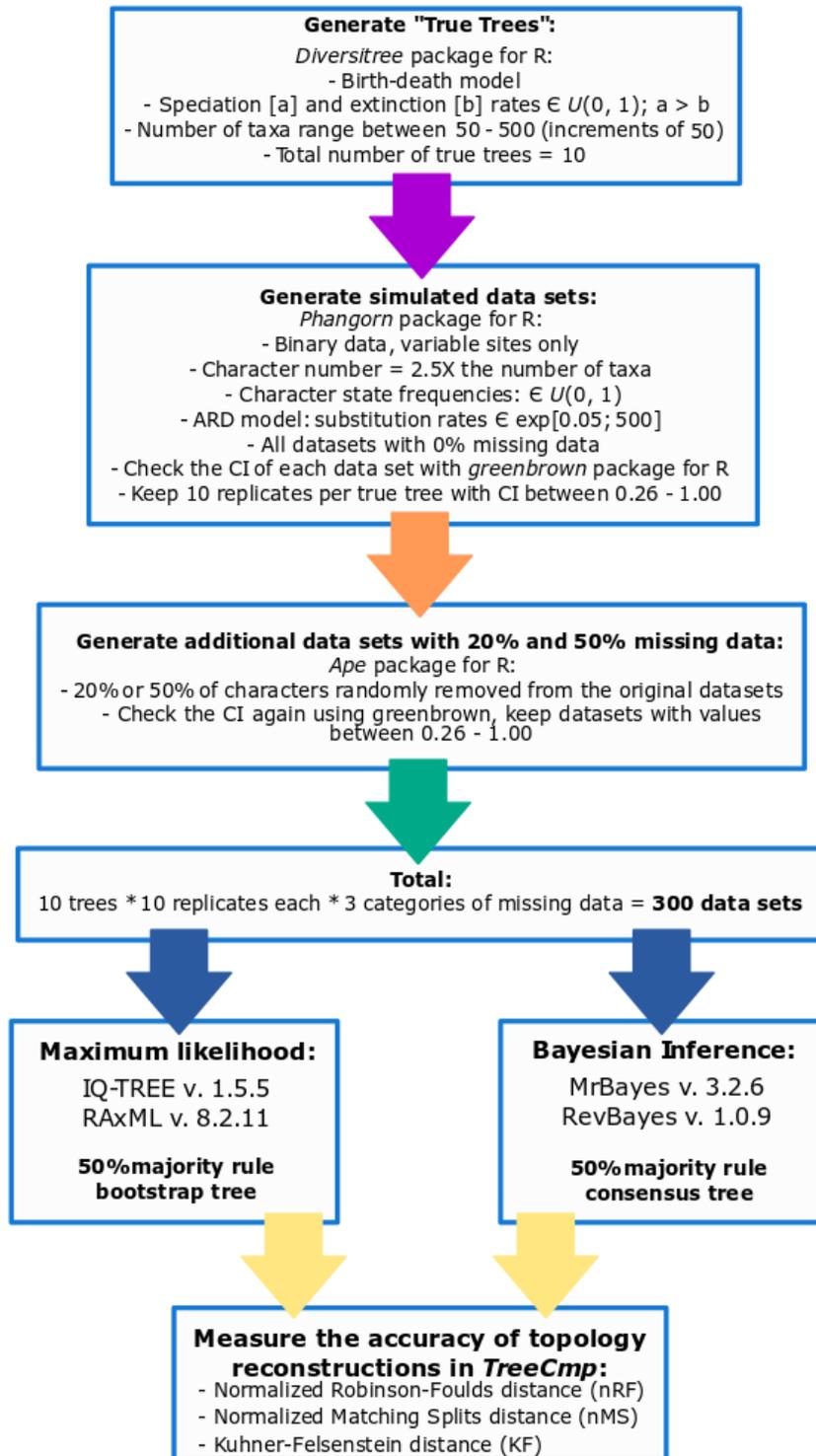


FIGURE 2-1. Flow chart of study design outlining details of each step of dataset simulations, phylogenetic analyses, and performance assessment.

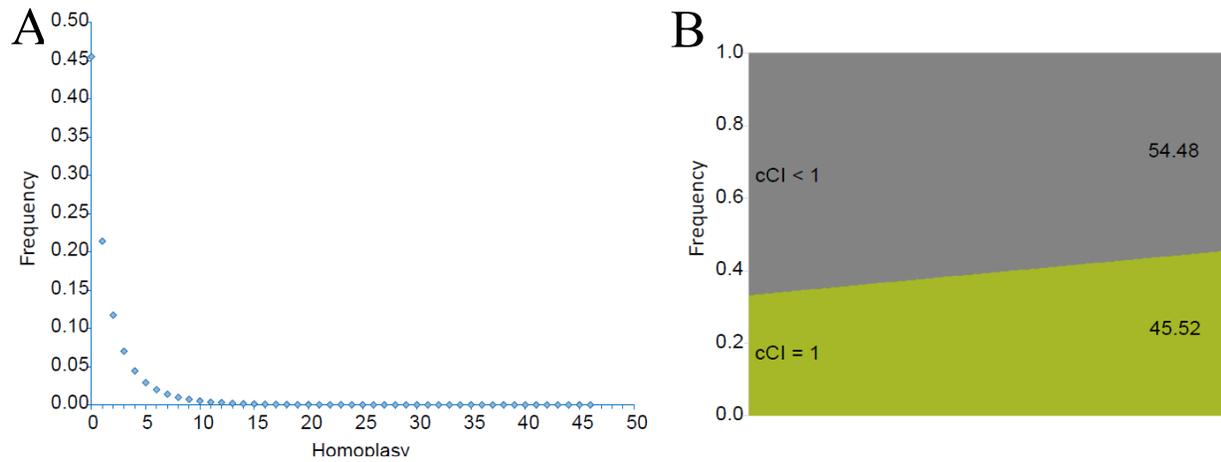


FIGURE 2-2. Homoplasy levels in simulated data sets. A – Distribution of homoplasy across individual characters in all simulated data sets used in the study (homoplasy calculated as $(1/cCI - 1)$ and represents extra number of steps for each individual character); B – Area plot showing percentage of individual characters across all simulated data sets with $cCI = 1$ (green field) and $cCI < 1$ (grey field).

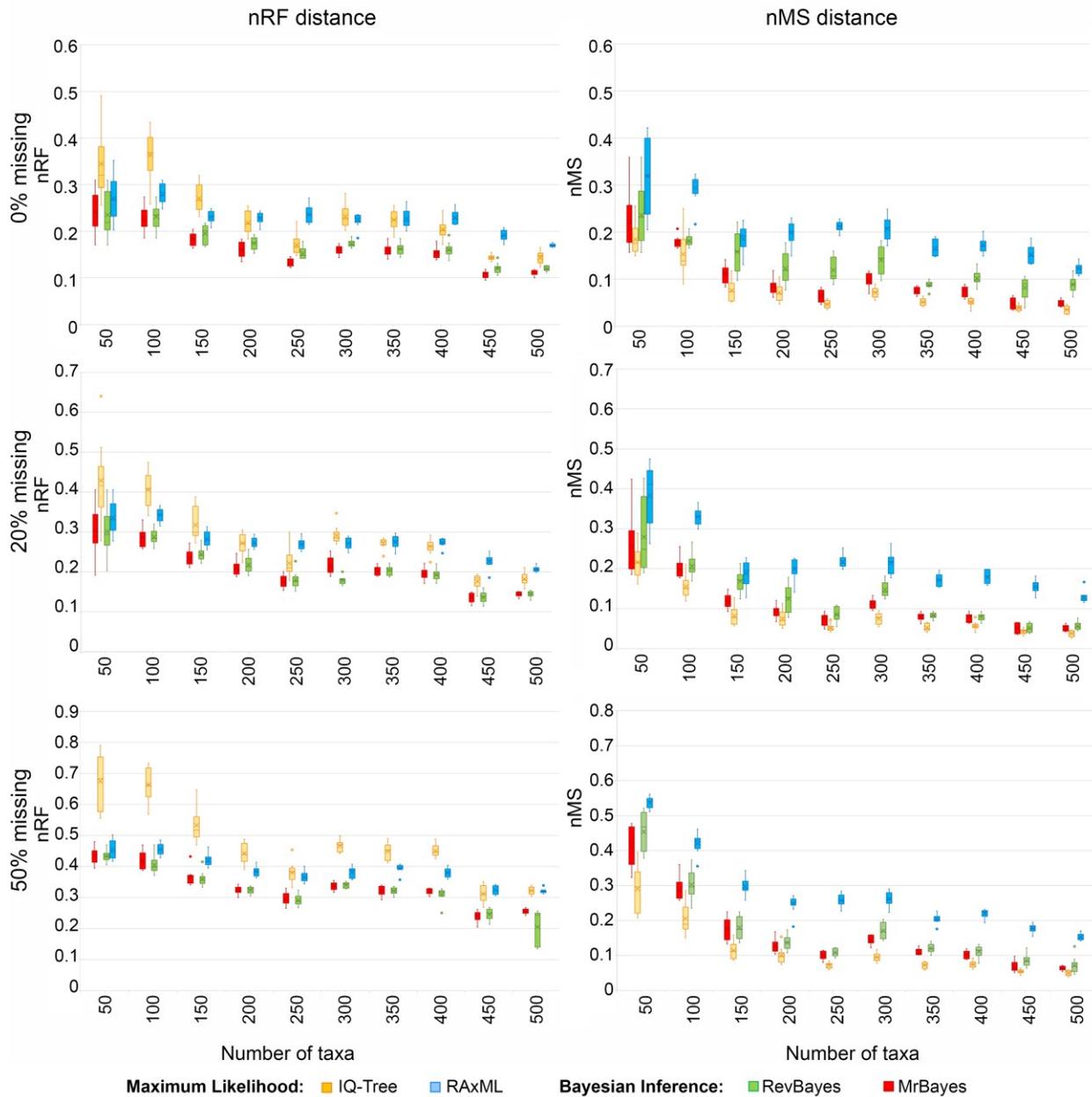


FIGURE 2-3. Normalized Robinson-Foulds (nRF, left column) and Matching Splits (nMS, right column) values for all methods across three categories of missing data: top—0% missing data; middle—20% missing data; bottom—50% missing data. Mean value for each series of replicates is indicated by ×; median values are indicated by horizontal dashes.

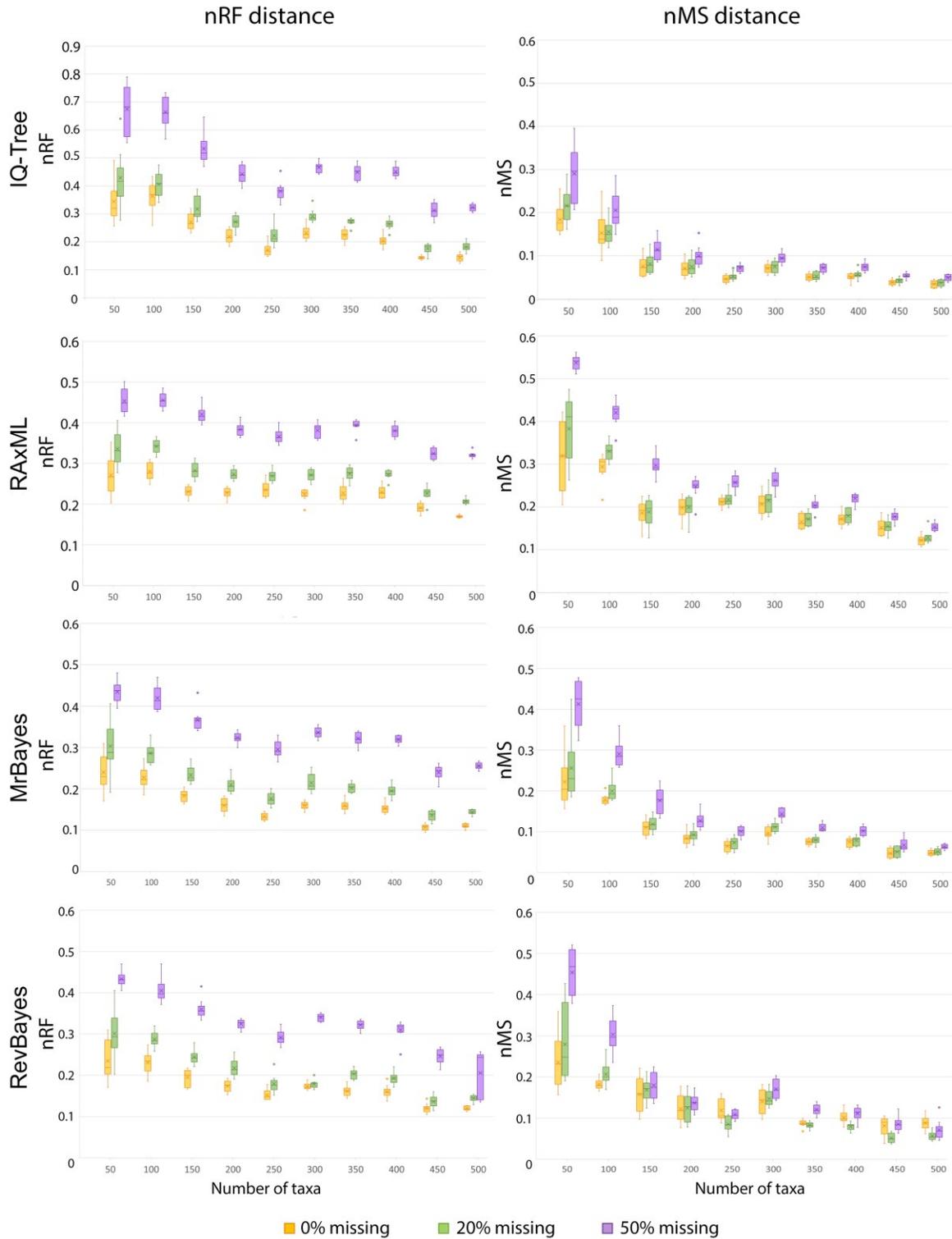


FIGURE 2-4. Normalized RF (left column) and MS (right column) metric values by individual phylogenetic software (indicated on the left) and three categories of missing data.

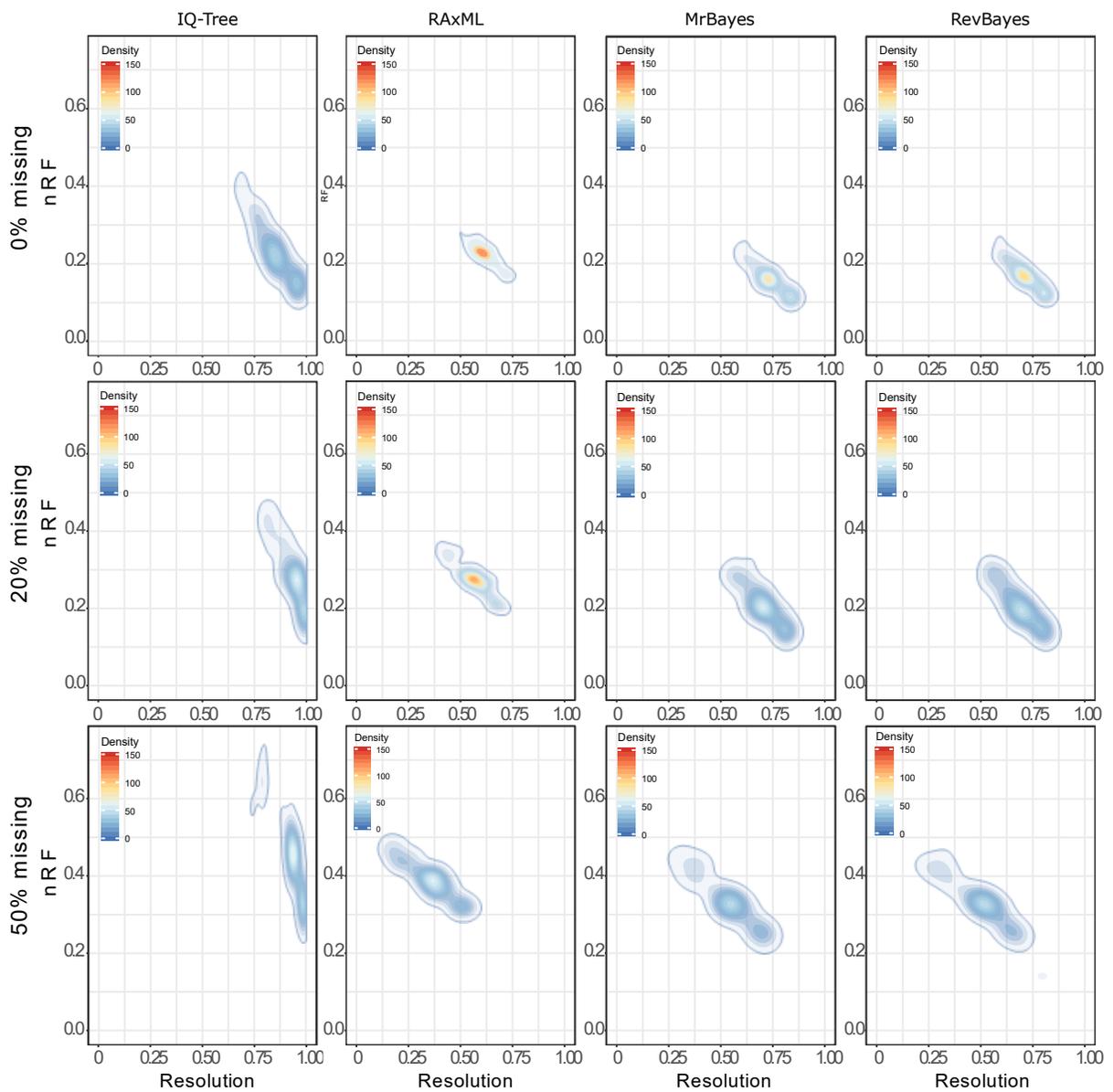


FIGURE 2-5. Contour plots showing density distribution of the results of phylogenetic analyses in the space of normalized Robinson-Foulds (nRF) distances against the relative resolution of the reconstructed trees (number of nodes on estimated topology as a fraction of the maximum possible number of nodes in a fully resolved topology with the same number of terminal taxa).

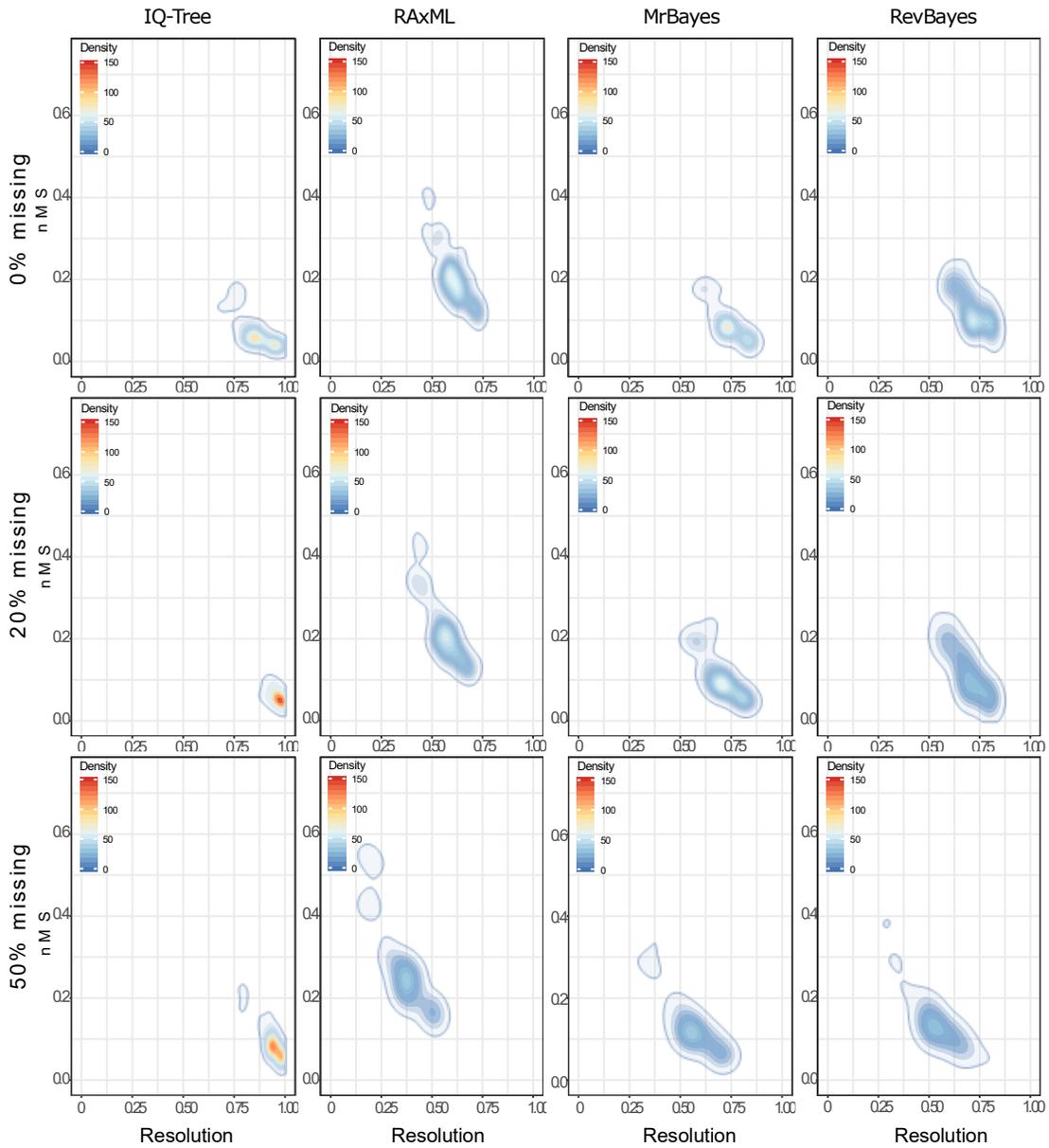


FIGURE 2-6. Contour plots showing density distribution of the results of phylogenetic analyses in the space of normalized Matching Splits (nMS) distances against the relative resolution of the reconstructed trees (number of nodes on estimated topology as a fraction of the maximum possible number of nodes in a fully resolved topology with the same number of terminal taxa).

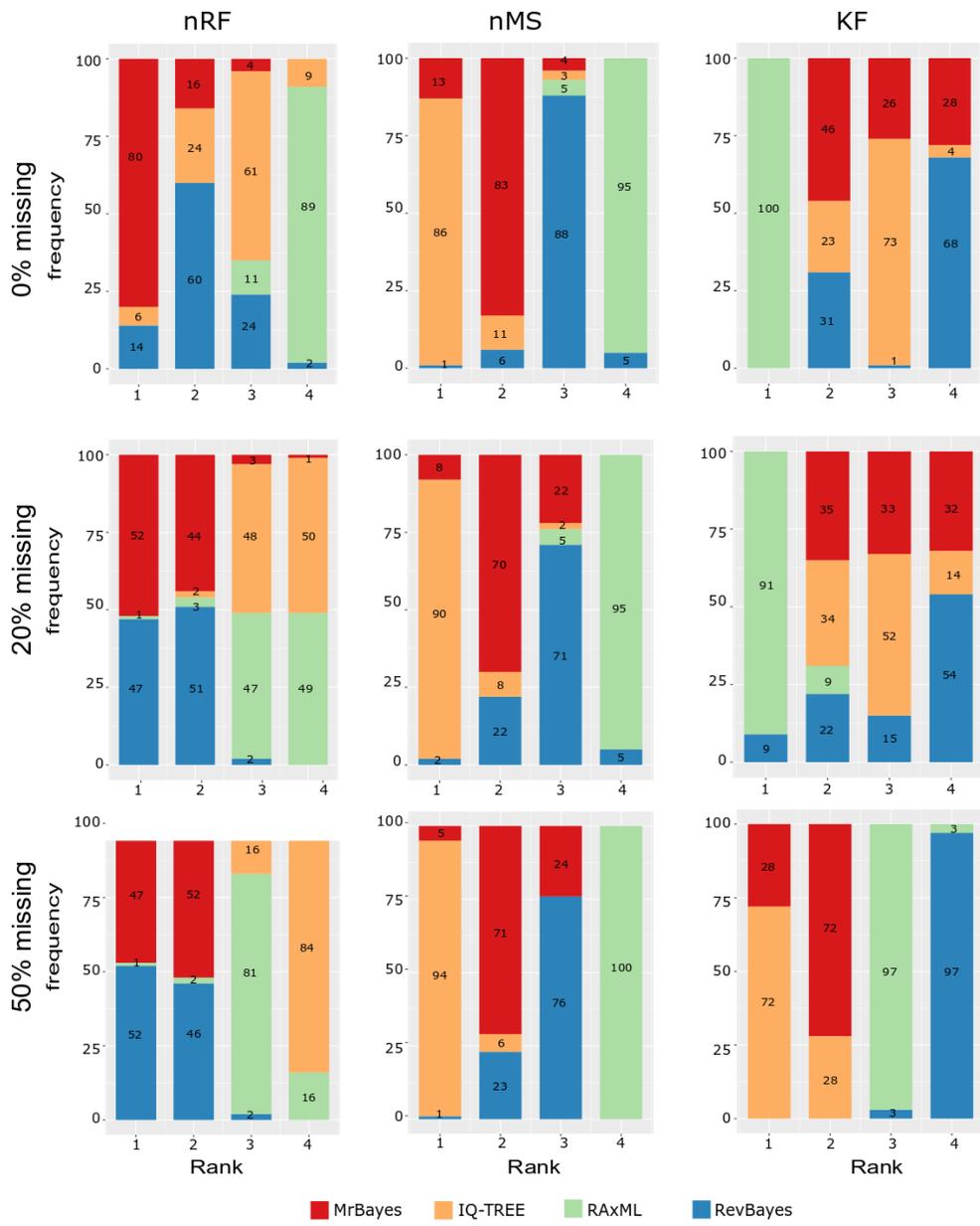
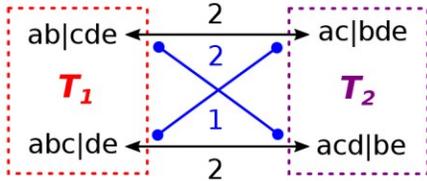


FIGURE 2-7. Results of rank tests for nRF, nMS, and KF tree comparison metrics.

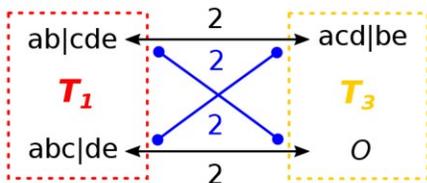
Stacked bars show how frequently each program was scored at each rank.

Calculation of MS on resolved trees is straightforward:

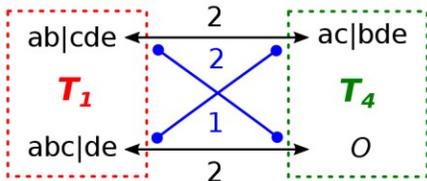


$$d_{MS}(T_1, T_2) = 2 + 1 = 3$$

Location of splits on unresolved trees affects the value of MS:



$$d_{MS}(T_1, T_3) = 2 + 2 = 4$$



$$d_{MS}(T_1, T_4) = 2 + 1 = 3$$

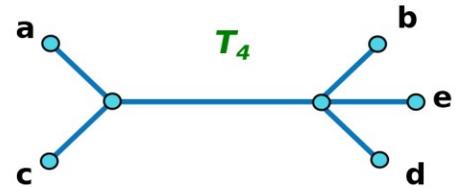
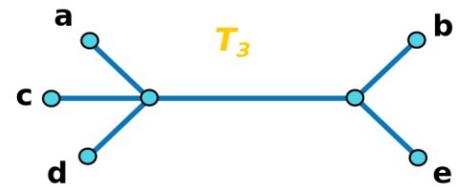
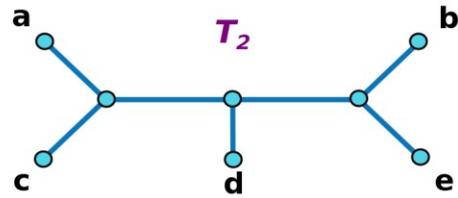
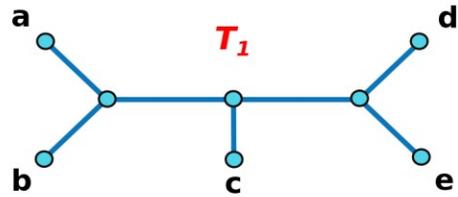


FIGURE 2-8. Calculation of the MS metric for fully resolved topologies and topologies containing polytomies.

CHAPTER 3

Using empirical character state frequencies in morphological phylogenetics under Bayesian inference

3.1 Introduction

The utility of probabilistic approaches (maximum likelihood and Bayesian inference) for reconstructing phylogenies using morphological data has long been controversial due to the apparent unrealistic specifications of available models of evolution. Although several recent studies indicate that Bayesian inference outperforms traditional parsimony when applied to discrete morphological data (Wright and Hillis, 2014; O'Reilly et al., 2016; Puttick et al., 2017a, 2017b; O'Reilly et al., 2018; Puttick et al., 2019), scepticism remains about its use. The major objections are whether constraints on the parameters used in a likelihood model describing an evolutionary scenario are applicable to morphological data (Goloboff et al., 2018). The explicit Markov model is essentially a generalized Jukes-Cantor (JC69) model (Jukes and Cantor, 1969) with k number of states (Mk model) that specifies an oversimplified evolutionary scenario with a single rate for all character state changes and equal character state frequencies – assumptions that have been criticized as being unrealistic for empirical data sets (e.g., Lewis, 2001; Wright et al., 2016; Pyron, 2017).

Development and implementation of new models of evolution are hindered by both an infinitely complex nature of morphological trait evolution and computational power constraints. Despite these limitations, optimization of model-based approaches for morphological data analysis is necessary for more accurate estimation of evolutionary relationships. A few suggestions have been proposed for overcoming some of the assumptions of the Mk model.

Wright et al. (2016) used the relationship between equilibrium character state frequencies and instantaneous substitution rates to model unequal state frequencies; they tested modifications of the beta distribution to model asymmetrical rates of character state transitions and showed that unequal transition rates can improve phylogenetic estimations in some cases. Pyron (2017) used an existing F81 model (Felsenstein, 1981) implementation for restriction site molecular data to analyse binary morphological data sets. Although intended to be more biologically plausible than the traditional Mk model, this approach has a number of limitations including restriction of the number of character states and *a priori* polarization of character states (hypothetical ancestral state scored as “0” and derived conditions scored as “1”). Direct application of the restriction site binary model as well as other models of molecular evolution to the morphological data entails a serious problem regarding the identity of each character scoring, since it is implicit in the model that character state labels are comparable across characters (Ronquist et al., 2019). This means that a morphological trait scored as “0” for one character is assumed to be comparable to features scored the same way for all other characters in a data set. Even if all characters in a data set are consistently scored as absence [0] and presence [1], presence of various morphological features (all traits scored as 1) is not comparable across characters, e. g. presence of posterodorsal spine on the opercle cannot be meaningfully compared to the presence of abdominal scutes (characters 34 and 55, both character state “1” in Wilson and Murray (2008) data set of osteoglossomorph fishes). This assumption is not true for most morphological data sets, while in molecular data sets, nucleobases (adenine, guanine, cytosine, thymine, and uracil) have the same meaning and are directly comparable across all sequence sites. Therefore, model optimization for discrete morphological data analysis under Bayesian inference remains problematic.

Herein, I present a model of discrete morphological trait evolution that uses empirical character state frequencies for each individual character in a data set. The new model is hereinafter referred to as FreqMorph (Frequency model for Morphological data) and is written and tested as a supplement code to be run with BEAST2 software (Bouckaert et al., 2014).

3.2 Materials and methods

3.2.1 Description and rationale of the FreqMorph model

The general likelihood function for a given tree topology and a data set is defined as the proportion of the conditional probability of the observed data given a model of evolution:

$$L(Q, T) \propto P(D|Q, T), \quad (1)$$

where Q is a rate matrix defining model of evolution and T is a tree topology for which likelihood of the observed data set, D , is calculated. In a phylogenetic analysis, the overall probability is calculated as the product of conditional probabilities for each character in a data set assuming that data observed at each site evolved independently of each other:

$$P(D|Q, T) = \prod_i P(D_i|Q, T) \quad (2)$$

Following the assumption of character independence that underlies the construction of morphological data sets, the general Mk model can be extended to include individual character state frequency parameters for each character in a data matrix. Individual character state frequencies do not all have to be different from each other and there can be fewer patterns than there are characters in a data set. To achieve this goal, the FreqMorph model partitions the data set with individual frequency parameter values are specified for each character – the data set is initially split into individual character partitions so that the total number of partitions equals the number of characters in the data matrix; these partitions receive individual character state

frequency parameter values while tree topology and proportionality of branch lengths is linked across partitions. Because FreqMorph uses individual rate matrices for each character in a data set, the function is modified to integrate over multiple Q matrices to calculate likelihood for a given topology:

$$P(D|Q_1, Q_2 \dots Q_i, T) = \prod_i P(D_i|Q_1, Q_2 \dots Q_i, T) \quad (3)$$

where each individual character with k character states has a rate matrix defined as:

$$Q = \begin{array}{c} \text{Char. states} \quad [0] \qquad \qquad [1] \qquad \qquad \dots \qquad \qquad [k-1] \\ \left[\begin{array}{cccc} -\mu r_{0 \rightarrow 0}(\pi_1 + \dots + \pi_{k-1}) & \mu r_{0 \rightarrow 1} \pi_1 & \dots & \mu r_{0 \rightarrow (k-1)} \pi_{k-1} \\ \mu r_{1 \rightarrow 0} \pi_0 & -\mu r_{1 \rightarrow 1}(\pi_0 + \dots + \pi_{k-1}) & \dots & \mu r_{1 \rightarrow (k-1)} \pi_{k-1} \\ \dots & \dots & \dots & \dots \\ \mu r_{(k-1) \rightarrow 0} \pi_0 & \mu r_{(k-1) \rightarrow 1} \pi_1 & \dots & -\mu r_{k-1 \rightarrow k-1}(\pi_0 + \dots + \pi_{k-2}) \end{array} \right] \end{array}$$

where r is the transition rate between two different character states (e. g. $[r_{0 \rightarrow 1}]$ is a transition rate from character state 0 to 1), μ is the probability of character state change at any given time, π is stationary character state frequency, and k is total number of character states for a given character in a data set.

Unlike the implementation of the traditional Mk model in BEAST and MrBayes (Ronquist et al., 2012), which automatically splits an entire data set into partitions according to the number of character states and applies equal character state frequencies to each partition following a simple formula, $\pi = \frac{1}{k}$, where π is a character state frequency that is equal for all characters in a partition and k is the number of character states in that partition (Lewis, 2001), the FreqMorph model treats each individual character in a data set separately with its own character state frequency parameter. The starting value of each frequency parameter is calculated based on the observed numbers for each state of a given character in a matrix. The initial vector containing relative frequencies of the character states for a partition i with k states is therefore defined as:

$\pi_i = \left[\frac{a_1}{n_i}, \frac{a_2}{n_i}, \frac{a_3}{n_i}, \dots, \frac{a_k}{n_i} \right]$, where n_i is the total number of taxa scored for a character i and a is the number of times that a particular character state is being scored for a given character (e. g. a_1 is number of taxa scored “0” for a given character in a data set, a_2 is number of taxa scored as “1”, a_3 – number of taxa scored as “2”, and a_k – number of taxa scored as $(k - 1)$). The total number of taxa scored for a character may differ across partitions depending on the amount of missing data per character. In the best-case scenario when all taxa are scored for all characters and there is no missing data in the data set, $n_1 = n_2 = n_3 = \dots = n_i =$ number of taxa in the data set. However, if some taxa are not scored for any given character, the total number n is calculated as the number of taxa with character state scores, taxa with missing data are omitted from the total number calculation. This approach allows FreqMorph to be used with empirical data sets that often contain a fraction of missing data, but it also makes the model susceptible to the additional error in the frequency parameter calculation resulting from an incomplete scoring of taxa.

To illustrate the difference between the Mk and FreqMorph approaches, consider a simple case of a data matrix with 5 taxa and 6 characters:

A	0	0	0	1	1	1
B	0	1	1	1	0	2
C	1	1	1	1	0	0
D	1	0	1	0	0	2
E	1	0	2	0	0	2

In the traditional Mk model implementation, the above data matrix is subdivided into two partitions according to the k number of states – partition I ($k = 2$) = {characters 1, 2, 4, 5} and partition II ($k = 3$) = {characters 3, 6}. In the binary partition I, all characters are assumed to have equal character state frequencies with $\pi_0 = \pi_1 = \frac{1}{k} = \frac{1}{2} = 0.5$; in partition II containing characters with three character states, base frequencies are $\pi_0 = \pi_1 = \pi_2 = \frac{1}{k} = \frac{1}{3} = 0.333$.

In case of the FreqMorph model, each character in the data set above receives an individual vector of empirical character state frequencies, e. g. for character 1, $\pi_0 = \frac{2}{5} = 0.4$ and $\pi_1 = \frac{3}{5} = 0.6$. A comparison of character state frequencies for each character in the data set under the Mk and FreqMorph models is shown below:

	Mk	FreqMorph
char.1	$\pi_0 = 0.5; \pi_1 = 0.5$	$\pi_0 = 0.4; \pi_1 = 0.6$
char.2	$\pi_0 = 0.5; \pi_1 = 0.5$	$\pi_0 = 0.6; \pi_1 = 0.4$
char.3	$\pi_0 = 0.33; \pi_1 = 0.33; \pi_2 = 0.33$	$\pi_0 = 0.2; \pi_1 = 0.6; \pi_2 = 0.2$
char.4	$\pi_0 = 0.5; \pi_1 = 0.5$	$\pi_0 = 0.4; \pi_1 = 0.6$
char.5	$\pi_0 = 0.5; \pi_1 = 0.5$	$\pi_0 = 0.8; \pi_1 = 0.2$
char.6	$\pi_0 = 0.33; \pi_1 = 0.33; \pi_2 = 0.33$	$\pi_0 = 0.2; \pi_1 = 0.2; \pi_2 = 0.6$

This approach eliminates the problem of comparable character labels across sites by calculating frequency values for each site (character) in a matrix individually. Additionally, the FreqMorph model is not restricted to a particular coding scheme and can be applied to binary as well as multistate characters. Together with the ability to estimate frequency values through the MCMC sampling procedure, this model presents an improved way of assessing morphological evolution in a probabilistic framework.

Introducing individual character state frequency parameters into an evolutionary model, however, presents the concerns of overparameterization and susceptibility of the model to statistical inconsistency (Cunningham et al., 1998; Lewis, 2001; Lemmon and Moriarty, 2004). Overparameterization is a result of including numerous free parameters in a model so that it better fits the data. Such enforced overfitting of the model results in a higher model-fit value compared to a simpler evolutionary model with fewer parameters. To account for the possibility of artificial inflation of model-fitting, it is important to use information criteria that penalize the inclusion of increasing numbers of free parameters in the model (e.g., Akaike information criterion [AIC] (Akaike, 1974), corrected AIC, Bayes factor, etc.). In addition, consistency of a

model should be tested with a large number of characters in a data set and a large number of replicates per each set of character numbers.

To test performance of the new model, I (1) compare model-fit values for three empirical morphological data sets analyzed under traditional Mk and FreqMorph models; (2) assess the accuracy of topology reconstruction using simulated data sets; and (3) assess accuracy of the topology representation using Maximum clade credibility (MCC) and majority-rule consensus (MRC) trees.

3.2.2 Empirical data sets

I analysed three published empirical data sets which were chosen to include a variety of taxonomic groups as well as variation in numbers of characters: (1) Osteoglossomorpha with 31 teleost bony-tongue fish taxa and 87 characters (Murray et al., 2016); (2) Angiospermae with 86 flowering plant taxa and 142 characters (Doyle and Endress, 2014); and (3) Hymenoptera with 107 insect taxa (bees, sawflies, wasps and ants) and 354 characters (Klopfstein et al. 2015).

3.2.3 Simulated data sets

True tree simulation: I generated four “true tree” topologies with 25, 50, 100, and 200 taxa under the birth-death model using *diversitree* (FitzJohn, 2012) package for R (R Core Team, 2018). Speciation and extinction rates were sampled randomly from a uniform distribution based on the condition that speciation rate is higher than extinction rate.

Data set simulations: For each “true tree”, I generated several sets of multistate morphological matrices using the *phangorn* package (Schliep, 2011) for R (R Core Team, 2018) including: 25 taxa with 50 and 100 characters; 50 taxa with 100, 350, and 1000 characters; 100 taxa with 200, 350, and 1000 characters; 200 taxa with 350, 500, and 1000 characters. Each set

of data matrices contained 100 replicates. Only variable characters were included in the final data sets to better represent empirical morphological data matrices. I generated data sets under the all-rates-different (ARD) model with substitution rates randomly sampled from an exponential distribution with a rate parameter λ drawn randomly from a uniform distribution (U [0.5;50]) which allows asymmetrical substitution rates violating assumptions of a symmetrical time-reversible Mk model (Lewis, 2001). Character state frequencies were randomly sampled from a uniform distribution (U [0;1]) for each individual character in a data set.

This approach to simulating morphological data matrices does not necessarily represent realistic patterns of a discrete morphological trait evolution, which are infinitely complex and vary greatly across empirical data sets, but rather my simulation design aims to generate a variety of data sets to detect a general trend in performance of the two models under an idealized evolutionary scenario when there is no dependency across characters in a data set. While this simulation approach may favour a more complex and parameter-rich model (FreqMorph) by independent sampling of character state frequencies and substitution parameters, it will only be biased toward the FreqMorph model if the observed character state frequencies recorded in the generated data matrices accurately reflect the true underlying frequency parameters used in the simulation and if those true frequency values deviate significantly from the equal character state frequencies assumed under the Mk model.

To ensure that the data sets meet empirically determined levels of homoplasy (Sanderson and Donoghue, 1989) each generated matrix was screened so the consistency index value was between 0.26 and 1.0. In total, I included 1100 simulated data sets in the study.

3.2.4 Phylogenetic analyses

Phylogenetic analyses were performed in BEAST v. 2.5.1 (Bouckaert et al., 2014) under the Fossilized Birth-Death (empirical data sets) or Birth-Death (simulated data sets) tree prior, relaxed log-normal molecular clock model and either Mk or FreqMorph substitution model with four discrete gamma categories and an ascertainment bias correction applied. Other program parameters and priors were set to their default values. Each MCMC analysis was run for 5 - 20×10⁶ generations with trees sampled every 1000 generations. Results of each analysis were checked for stationarity in Tracer v. 1.7 (Rambaut et al., 2018) and to ensure that the ESS values for the estimated parameters were ≥ 200. The maximum clade credibility and majority-rule consensus trees were built using *phangorn* (Schliep, 2011) and *ape* (Paradis and Schliep, 2018) packages for R with the initial 25% of the trees removed as the burn-in. The model-fit analyses (Bayes factor and AIC) were performed in Path Sampler and AIC for MCMC applications available with the BEAST 2.5.1 software package (Bouckaert et al., 2014). For the Path Sampler analyses, I used the Path Sampler application in BEAST 2.5.1 that implements the steppingstone sampling method (Xie et al., 2011); I used the default settings with ten steps with 1*10⁶ generations each (total of 10 million generations). To calculate AICm values, I used 1000 replicates and 25% of burn-in for each run.

All FreqMorph analyses were performed in Eclipse Neon v. 4.6 integrated development environment (Eclipse Foundation, 2016) and on Compute Canada clusters using BEAST 2.5.1 source code supplemented with the script for the FreqMorph model. The source script of the FreqMorph model is available in Appendix 3-1.

3.2.4 Performance assessment

To test performance of the new model and assess accuracy of topological reconstructions, I used the 50% majority rule consensus trees and maximum clade credibility trees (MCC) as representative results for each method. To assess the accuracy and precision of the methods used to reconstruct the phylogenies, I used normalized Robinson-Foulds distances [nRF, (Robinson and Foulds, 1981)] as implemented in *phytools* (Revell, 2012) package for R.

Because RF distance is only well-defined for perfectly bifurcating trees and can retrieve biased results for trees with polytomies, I applied a “tree resolution correction” to scale RF scores by the factor of the relative resolution (percentage of node in the consensus topology divided by the maximum number of nodes in a fully resolved tree with the same number of taxa):

$$\text{nRF}_{\text{cor}} = (\text{nRF} * \text{node}_{\text{max}}) / \text{node}_{\text{con}},$$

where node_{max} is the maximum number of nodes in a fully resolved tree, and node_{con} is the number of nodes in a consensus topology. This correction penalizes poorly resolved consensus trees that may have seemingly good RF scores due to collapsed clades. Additionally, to compare divergence between consensus topologies reconstructed under different models, I calculated the average nRF score as percentage of the total number of resolved nodes in FreqMorph and Mk consensus topologies:

$$\overline{\text{nRF}} = \frac{(N_{\text{Mk}} + N_{\text{FreqMorph}})}{2 \times N} \times 100\%$$

Additionally, for MCC trees, I calculated tree branch length distance, BS (Kuhner and Felsenstein, 1994; Felsenstein, 2004) as implemented in the Dendropy python library (Sukumaran and Holder, 2010). This metric calculates the sum of squares of the differences between the length of each branch in the true and the estimated trees. Branches that are not

shared between compared topologies are set to length 0. The branch length distance is 0 for identical trees. The BS distance was normalized by the total number of tree edges.

Finally, I used *ape* package (Paradis and Schliep, 2018) for R to assess combined accuracy of the tree topology and branch lengths of results recovered by each model. I used the RF algorithm with branch lengths (RFL, Kuhner and Yamato, 2014) tree metric that accounts for both parameters. I then compared metric results for each model using a two-tailed pair-wise t-test to assess statistical significance of the observed differences.

3.3 Results

3.3.1 Empirical data sets

For all empirical data sets, FreqMorph had a higher model fit under both AICm and Bayes factor criteria (Table 3-1) than the Mk model. Bayes factor values indicated very strong ($2\ln(\text{BF}) > 10$) to decisive ($2\ln(\text{BF}) \gg 10^2$) support for FreqMorph compared to the Mk model. Majority rule consensus topologies recovered under the FreqMorph model had slightly higher resolution than the Mk trees for corresponding data sets. Maximum clade credibility topologies recovered under FreqMorph had more nodes with higher posterior support values than the corresponding Mk trees (Table 3-2).

Average nRF difference between consensus topologies recovered under the FreqMorph and Mk models was 60% (Osteoglossomorpha), 77% (Angiospermae), and 13% (Hymenoptera); for the MCC trees, topological difference was 31% (Osteoglossomorpha), 50% (Angiospermae), and 37% (Hymenoptera).

3.3.2 Simulated data set

Results of the simulated data set analyses were consistent with the results of the empirical trials – the model fit was always higher for the FreqMorph model than the Mk model, with the Bayes Factor of FreqMorph indicating strong to decisive support. Results of analyses with fixed numbers of taxa and variable numbers of characters indicated that FreqMorph produces more accurate results overall as the number of characters increases. Notably, when the nRF metric of consensus topologies was corrected for relative resolution (nRF_{cor}), the difference in accuracy estimation between MCC and consensus trees decreased considerably (Fig. 3-1).

Overall, there was a significant difference in the accuracy of topological reconstructions between the two models for both maximum clade credibility and majority-rule consensus trees (Table 3-3). In the pooled sample of 1100 simulated data sets, FreqMorph gave higher topological accuracy in 32% (consensus trees) and 35% (MCC trees) of analyses, and both methods performed with equal topological accuracy in an additional 36% and 30% percent of all simulated data set analyses (Fig. 3-2). However, when branch lengths were taken into consideration (RFL metric), accuracy of phylogenetic reconstructions was higher under the FreqMorph model in 60.3% of the analysed data sets, and accuracy of the branch lengths estimation alone was higher for the FreqMorph MCC trees in 62.5% of all analyses (Fig. 3-2).

Performance of each method, however, varied across categories of data sets. For small data sets (25 taxa with 50 and 100 characters), both models performed with equal accuracy in more than 50% of the analyses. Disparity in the accuracy between the two models was the greatest for large data sets with 100 and 200 taxa; in these cases, the FreqMorph model generally outperformed Mk in the accuracy of topological reconstruction and branch length estimation (Fig. 3-3).

3.4 Discussion

For nearly two decades, probabilistic frameworks for morphological phylogenetics have been constrained to essentially a single model of evolution; this has left morphological phylogeneticists with no means to test alternative hypotheses of trait evolution models although this is common practice in molecular systematics (e.g., Posada and Crandall, 2001; Minin et al., 2003; Posada and Buckley, 2004; Sullivan and Joyce, 2005; Kelchner and Thomas, 2007; Luo et al., 2010; Hoff et al., 2016; Abadi et al., 2019). In the present study, the major goals are to expand the toolkit of substitution models available for morphological phylogenetics and investigate performance of the new parameter-rich model. Despite concerns deriving from adding more parameters and potentially overfitting the model, FreqMorph showed a consistent result across data sets with varying combinations of number of taxa and number of characters. The fact that FreqMorph preformed best for large data sets (100 – 200 taxa) indicates that as more taxa are involved, more model complexity is required. In empirical studies, this, of course, will depend on the sampling strategy and how heterogenic evolutionary rates are among the taxa. For closely related lineages with similar evolutionary rates, it may be possible for a simple model to capture the general evolutionary pattern for a successful phylogenetic reconstruction. However, for more distantly related lineages with differing evolutionary rates, a simple model is not satisfactory. If we accept that none of the models can accurately capture the complexity of evolution of morphological traits, we also should admit that one simple model cannot fit all data sets. To accommodate the intricacy of empirical data sets, it is necessary to expand the currently existing set of models to test alternative hypotheses and apply them in a model averaging

framework for parameter value estimation (e. g. Posada and Buckley, 2004; Posada, 2008; Li and Drummond, 2011; Darriba et al., 2012; Bouckaert and Drummond, 2017).

Model fit estimation is a common way to compare candidate models for data analysis. In phylogenetic studies, Bayes factor (BF) and Akaike information criterion (AIC) are two commonly used metrics that either implicitly penalize an increasing number of free parameters in a model (AIC) or account for them in a process of marginal likelihood estimation (BF) (Akaike, 1974; Kass and Raftery, 1995). In the analyses of empirical data sets, the more complex model (FreqMorph) had consistently higher model fit values, indicating strong to decisive support for a more complex model over the Mk model. In the absence of a known true tree, model-fit and node-support values are often used to choose between alternative models, and in all cases tested herein, these metrics indicate support in favour of selecting the FreqMorph model. However, these results of the empirical data set analyses must be taken with caution, considering previous findings by Heled and Bouckaert (2013) as well as the simulation results in the present study which show that higher model support value does not necessarily directly translate into higher accuracy of a phylogenetic reconstruction inferred under the selected model.

Model selection and topology inference are often treated as separate processes in investigative studies that examine only one of these factors; however these processes are integrated under the common goal of producing an accurate phylogenetic estimation. Recently, Abadi et al. (2019), in their in-depth comparison of different model selection criteria and models of evolution, showed that different selection criteria can produce similar phylogenetic end-results and that the model selection process could be omitted in favour of the most complex model. In their simulation study, opting for a parameter-rich GTR+ Γ +I model did not have a crucial effect on the final result when the primary focus was topology or ancestral sequences. Results of the

simulation studies here suggest that this might also be the case for morphological data. Topological differences between MCC trees inferred under Mk and FreqMorph differed significantly in favour of a simpler model in only one category of data sets (25 taxa with 100 characters). Similarly, consensus topologies summarized under a simpler Mk model were significantly better for two sets of replicates (25 taxa with 100 characters and 50 taxa with 100 characters), but after correcting for resolution, this difference was reduced to the same set of matrices as for the MCC results (25 taxa with 100 characters). In the remaining cases, topologies reconstructed under both models either were not significantly different between models or were better under the parameter-rich FreqMorph model. Selecting a more complex model, however, also entails a number of issues such as increased computational time and resources.

The results of simulated data analyses indicate that the FreqMorph model is sensitive to sampling. While the underlying simulation procedure used in the present study favours random sampling of unequal character state frequencies, and could potentially create bias in favour of the FreqMorph model, the current implementation of the FreqMorph model relies on a representative sampling of character states; this, however, can be problematic, especially in case of small number of taxa (25 – 50 taxa) which can produce empirical frequencies very different from the underlying character state frequencies used to generate data sets. This model sensitivity can be corrected by thorough taxonomic sampling and estimation of the character state frequencies during the MCMC sampling procedure.

For the empirical data sets, it is impossible to assess phylogenetic accuracy of the reconstructed topologies; however, the observed discrepancies in topological resolution and support values between the Mk and FreqMorph results can be attributed to individual data set properties such as strength of phylogenetic signal, amount of missing data, and presence of

autapomorphies in a data set. Tree-length distributions for the three empirical data sets show similar degrees of right skewness, indicating a considerable degree of consistency of signal in the data sets (Fig. 3-4). However, the Hymenoptera data set is characterized by a high proportion of missing data per character - twice the amount in the Osteoglossomorpha and Angiospermae data sets (Fig. 3-5). Large amounts of missing data will result in high heterogeneity in the posterior sample of the model parameters and topologies, and consequently poor resolution of consensus topologies and low nodal support.

It has been suggested that topology estimation should be less affected by model selection than by other estimated parameters, such as branch lengths (Posada, 2001; Minin et al., 2003; Abdo et al., 2005; Kelchner and Thomas, 2007; Dornburg et al., 2018). This was observed in the simulation experiments. When the branch length parameter was considered, performance disparity between the two models increased almost twofold, with FreqMorph having higher accuracy than Mk in 62% of the cases. Significant correlation between branch length estimates and substitution model parameters including base frequencies follows from the fact that base frequencies are directly incorporated into the transition rate matrix and, together with the substitution rates, are used to calculate branch lengths. Simple models assuming equal base frequencies and substitution rates tend to underestimate true branch lengths (Yang et al., 1994; Adachi and Hasegawa, 1995; Yang et al., 1995; Posada, 2001; Schwartz and Mueller, 2010). This issue becomes especially important when estimating divergence times and evolutionary rates that rely on accurate estimation of the branch length parameter.

It is important to note limitations to assessing accuracy of topological reconstructions using strictly quantitative metrics such as RF. Equal RF scores between compared topologies inferred under different models would indicate no statistical significance in the results; however, different

placement of even a single taxon on a tree may have drastic effects on the subsequent interpretation of the inferred phylogenies, but strictly quantitative statistical assessment is not sensitive to such nuances. In the simulation study, about one third of the results had equal RF scores under FreqMorph and Mk; however, this does not necessarily mean that both models inferred the same topology, but rather that the recovered trees had the same number of incorrectly inferred clades compared to the true tree. In empirical studies, alternative placements of taxa can be a key factor in choosing between topologies, since researchers can interpret those options using their conception of the ingroup. It is, therefore, important to keep in mind, when interpreting results of statistical analyses based on the RF or similar tree distance metrics, that they only account for the quantitative side of the comparison and that distinct topologies may have the same RF distance from the true tree.

Overall, the present work lays the basis for future explorations and improvements in assessing morphological evolution in a probabilistic framework by relaxing assumptions of the current Mk model. The current implementation of the FreqMorph model in BEAST 2 requires users to run an R script for data parsing into multiple partitions. In future, this step can be integrated into the BEAUTi workflow with a proper template for parsing morphological data matrices. Additionally, FreqMorph which is essentially an extension of the Mk model that uses empirical character state frequencies is not restricted to a time-calibrated framework and can be applied to non-clock analyses as implemented in other popular Bayesian inference software (MrBayes, RevBayes, PhyloBayes, etc.) and even beyond the MCMC framework, in other likelihood-based methods (e. g. maximum likelihood) that implement data set partitioning.

3.5 Conclusions

Modelling morphological evolution is a daunting task; however, phylogenetic reconstructions do not require an exact realistic model, but rather a sensible approximation of the evolutionary processes. In the present study, I show that relaxing the assumption of equal character state frequencies produces consistent results across a wide range of analysed data sets, and can provide more accurate estimation of topology and branch lengths compared to the traditional Mk model. Although results show that FreqMorph always has higher model-fit than Mk, this should be taken with caution since model fit does not directly correlate with the accuracy of phylogenetic reconstructions. Based on the combination of data sets analysed in the present study, the FreqMorph model should be used for data sets with a relatively high number of taxa (> 50) and in cases when branch lengths and evolutionary rates are of particular interest.

References

- Abadi, S., D. Azouri, T. Pupko, and I. Mayrose. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications* 10: 10.1038/s41467-019-08822-w.
- Abdo, Z., V. N. Minin, P. Joyce, and J. Sullivan. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. *Molecular Biology and Evolution* 22:691–703.
- Adachi, J. and M. Hasegawa. 1995. Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. *Journal of Molecular Evolution* 40: 622–628.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–723.
- Bouckaert, R. R., and A. J. Drummond. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology* 17: 10.1186/s12862-017-0890-6.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10:e1003537.
- Cunningham, C. W., H. Zhu, and D. M. Hillis. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. *Evolution* 52:978–987.
- Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:10.1038/nmeth.2109.

- Dornburg, A., Z. Su, and J. P. Townsend. 2018. Optimal rates for phylogenetic inference and experimental design in the era of genome-scale data sets. *Systematic Biology* 68(1):145–156.
- Doyle, J. A. and P. K. Endress. 2014. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: ANITA lines and relatives of Chloranthaceae. *International Journal of Plant Sciences* 175:555–600.
- Eclipse Foundation. 2016. Eclipse Neon v. 4.6. Available at <https://eclipse.org/downloads/index-developer.php>. Accessed on September 28, 2018.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368–376.
- Felsenstein, J. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer associates, 580 pp.
- FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3:1084–1092.
- Goloboff, P. A., A. Torres, and J. S. Arias. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* 34:407–437.
- Heled, J. and R. R. Bouckaert. 2013. Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology* 13:10.1186/1471-2148-13-221.
- Hoff, M., S. Orf, B. Riehm, D. Darriba, and A. Stamatakis. 2016. Does the choice of nucleotide substitution models matter topologically? *BMC Bioinformatics* 17:10.1186/s12859-016-0985-x.
- Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules; pp. 21–132 in H. N. Munro(ed.), *Mammalian protein metabolism*. Academic Press, New York, 590 pp.

- Kass, R. E. and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–795.
- Kelchner, S. A. and M. A. Thomas. 2007. Model use in phylogenetics: nine key questions. *Trends in Ecology and Evolution* 22:87–94.
- Klopfstein, S., L. Vilhelmsen, and F. Ronquist. 2015. A nonstationary Markov model detects directional evolution in hymenopteran morphology. *Systematic Biology* 64:1089–1103.
- Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* 11:459–468.
- Kuhner, M. K. and J. Yamato. 2014. Practical performance of tree comparison metrics. *Systematic Biology* 64:205–214.
- Lemmon, A. R. and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology* 53:265–277
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Li, W. L. S. and A. J. Drummond. 2011. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* 29:751–761.
- Luo, A., H. Qiao, Y. Zhang, W. Shi, S. Y. Ho, W. Xu, A. Zhang, and C. Zhu. 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evolutionary Biology* 10:10.1186/1471-2148-10-242.
- Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology* 52:674–683.

- Murray, A. M., M. G. Newbrey, A. G. Neuman, and D. B. Brinkman. 2016. New articulated osteoglossomorph from Late Cretaceous freshwater deposits (Maastrichtian, Scollard Formation) of Alberta, Canada. *Journal of Vertebrate Paleontology* 36:e1120737.
- O'Reilly, J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* 61:105–118.
- O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biological Letters* 12:20160081.
- Paradis, E. and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.
- Posada, D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25:1253–1256.
- Posada, D. and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50:580–601.
- Posada, D. and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53:793–808.
- Puttick, M. N., J. E. O'Reilly, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, and P. C. Donoghue. 2017b. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B* 284:p.20162290.

- Puttick, M. N., J. E. O'Reilly, D. Oakley, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, and P. C. J. Donoghue. 2017a. Parsimony and maximum-likelihood phylogenetic analyses of morphology do not generally integrate uncertainty in inferring evolutionary history: a response to Brown *et al.* *Proceedings of the Royal Society B* 284: 20171636.
- Puttick, M. N., J. E. O'Reilly, D. Pisani, and P. C. Donoghue. 2019. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology* 62:1–17.
- Pyron, R. A. 2017. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Systematic Biology* 66:38–56.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67:901–904.
- Revell, L. J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223.
- Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53:131–147.
- Ronquist, F., J. P. Huelsenbeck, and M. Teslenko. 2019. MrBayes version 3.2 manual: tutorials and model summaries. Available online:

https://github.com/NBISweden/MrBayes/blob/develop/doc/manual/Manual_MrBayes_v3.2.pdf

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: 539–542.

Sanderson, M. J. and M. J. Donoghue. 1989. Patterns of variation in levels of homoplasy.

Evolution 43:1781–1795.

Schliep, K. P. 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* 27: 592e593.

Schwartz, R. S. and R. L. Mueller. 2010. Branch length estimation and divergence dating:

estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evolutionary Biology* 10:10.1186/1471-2148-10-5.

Sukumaran, J. and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing.

Bioinformatics 26:1569–1571.

Sullivan, J. and P. Joyce. 2005. Model selection in phylogenetics. *Annual Review of Ecology,*

Evolution, and Systematics 36:445–466.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods),

version 4.0a. Sinauer, Sunderland, Massachusetts. Available at

<https://paup.phylosolutions.com>.

Wilson, M. V. H. and A. M. Murray. 2008. Osteoglossomorpha: phylogeny, biogeography, and

fossil record and the significance of key African and Chinese fossil taxa. *Geological*

Society, London, Special Publications 295:185–219.

- Wright, A. M. and D. M. Hillis. 2014. Bayesian Analysis Using a Simple Likelihood Model Outperforms Parsimony for Estimation of Phylogeny from Discrete Morphological Data. *PLoS ONE*:9:e109210.
- Wright, A. M., G. T. Lloyd, and D. M. Hillis. 2015. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. *Systematic Biology* 65:602–611.
- Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 60:150–160.
- Yang, Z. I. H. E. N. G., N. Goldman, and A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Molecular Biology and Evolution* 11:316–324.
- Yang, Z., N. Goldman, and A. Friday. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Systematic Biology* 44:384–399.

TABLE 3-1. Results of the model fit analyses for empirical data sets. Akaike information criterion for MCMC (AIC_m) values are shown as $-\ln L$. Bayes Factor (BF) values are calculated as based on the marginal likelihood estimations from the stepping stone analyses ($BF = 2^{*(\ln L_{Mk} - \ln L_{FreqMorph})}$).

	Osteoglossomorpha			Hymenoptera			Angiospermae		
	AIC_m	$-\ln L$	BF	AIC_m	$-\ln L$	BF	AIC_m	$-\ln L$	BF
Mk	2021.21	-1089.57		11956.96	-6436.04		9296.28	-4949.21	
			36.64*			357.28*			1941.04*
FreqMorph	2007.21	-1071.25		11715.05	-6257.40		7455.59	-3978.69	

TABLE 3-2. Relative resolution and posterior support values for the summary topologies of the empirical data set runs recovered under the Mk and FreqMorph models.

	Osteoglossomorpha		Hymenoptera		Angiospermae	
	Relative resolution (%) of consensus	Nodes with pp>50% in MCC	Relative resolution (%) of consensus	Nodes with pp>50% in MCC	Relative resolution (%) of consensus	Nodes with pp>50% in MCC
Mk	59	13	57	54	64	49
FreqMorph	79	21	58	54	69	51

TABLE 3-3. Results of two-tailed pairwise t-tests comparing performance of Mk and FreqMorph models. Statistically significant values ($p < 0.05$) are indicated in bold font.

	25 taxa		50 taxa			100 taxa			200 taxa			Total
	50ch	100ch	100ch	350ch	1000ch	200ch	350ch	1000ch	350ch	500ch	1000ch	
nRF (MCC)	0.1775	0.0214	0.1612	0.9686	1.0000	0.2701	0.7181	0.0028	0.3194	0.8725	0.6517	0.0332
nRF (Con)	0.0279	0.0022	<<0.001	0.7237	0.1272	0.2372	0.3008	0.0004	0.4571	0.2501	0.0062	<<0.001
nRF _{cor}	0.0569	0.0002	0.6894	0.7934	0.1578	0.2882	0.4528	0.0003	0.5336	0.3511	0.0111	0.0060
RFL	0.4962	<<0.001	0.0029	0.0954	0.0731	0.0289	<<0.001	<<0.001	0.6043	0.3319	0.0277	0.9882
BS	0.4589	<<0.001	0.0018	0.1578	0.1123	0.0148	<<0.001	<<0.001	0.8590	0.1622	0.0094	0.0078

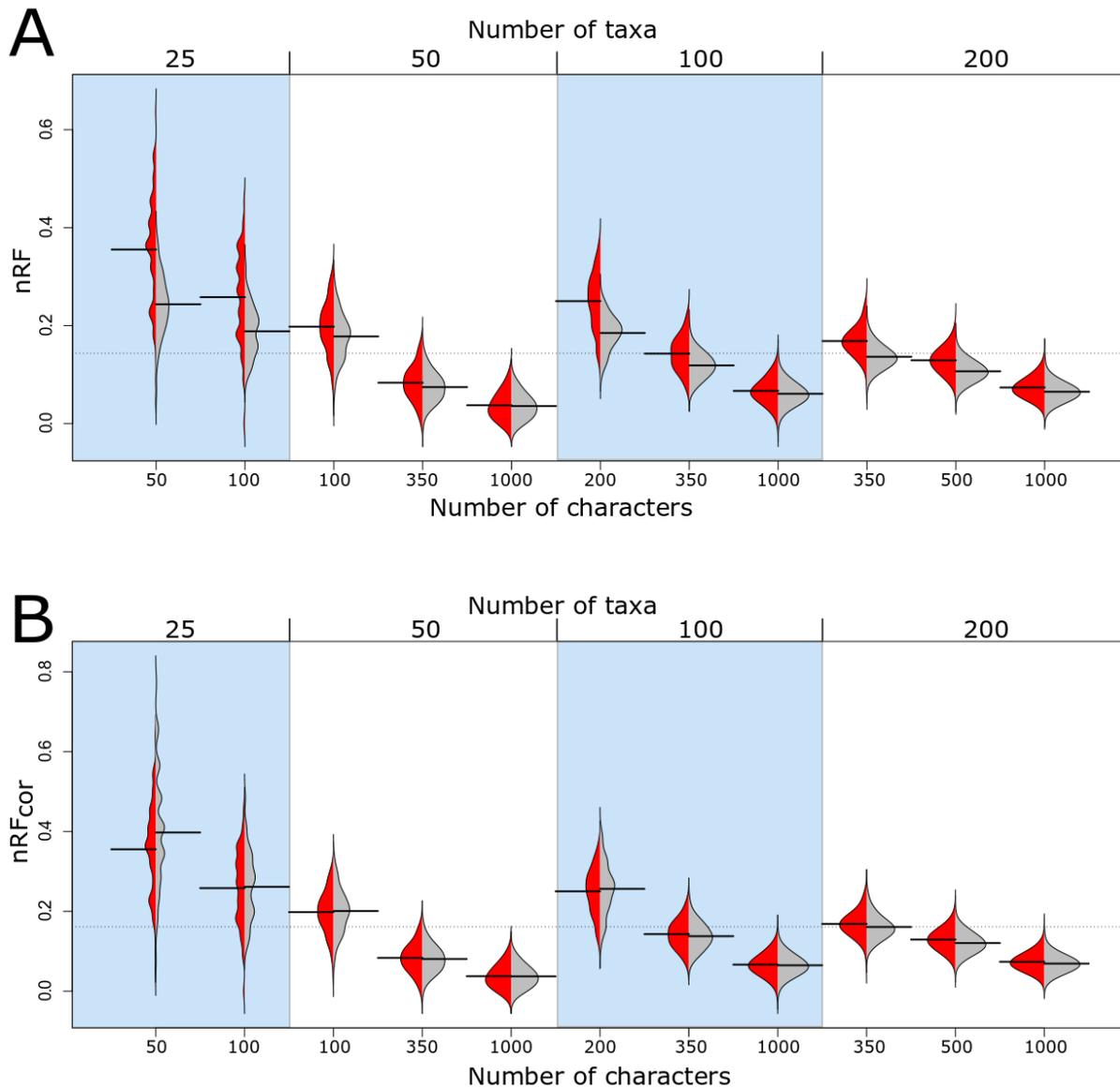


FIGURE 3-1. Assymmetric beanplots showing distribution of normalized Robinson-Foulds distances (nRF) for the maximum clade credibility trees (MCC) and consensus topologies reconstructed under FreqMorph for individual categories of data sets. In the top panel (A), each ‘bean’ shows distribution of nRF values for MCC topologies on the left side (red color) versus distribution of uncorrected nRF values for consensus trees on the right side (grey color). In the bottom panel (B), distribution of nRF values for MCC topologies (left side – red color) is plotted against distribution of the corrected nRF (nRF_{cor}) for consensus topologies; overall, distribution

of nRF_{cor} matches distribution of nRF values for fully resolved MCC trees better than uncorrected nRF values for consensus topologies indicating that the “resolution correction” reduces artificial inflation of the RF metric for trees with polytomies. Number of taxa is indicated above the plot and number of characters is shown below the plot for each set of simulated data matrices. Horizontal black lines indicate mean values for each distribution.

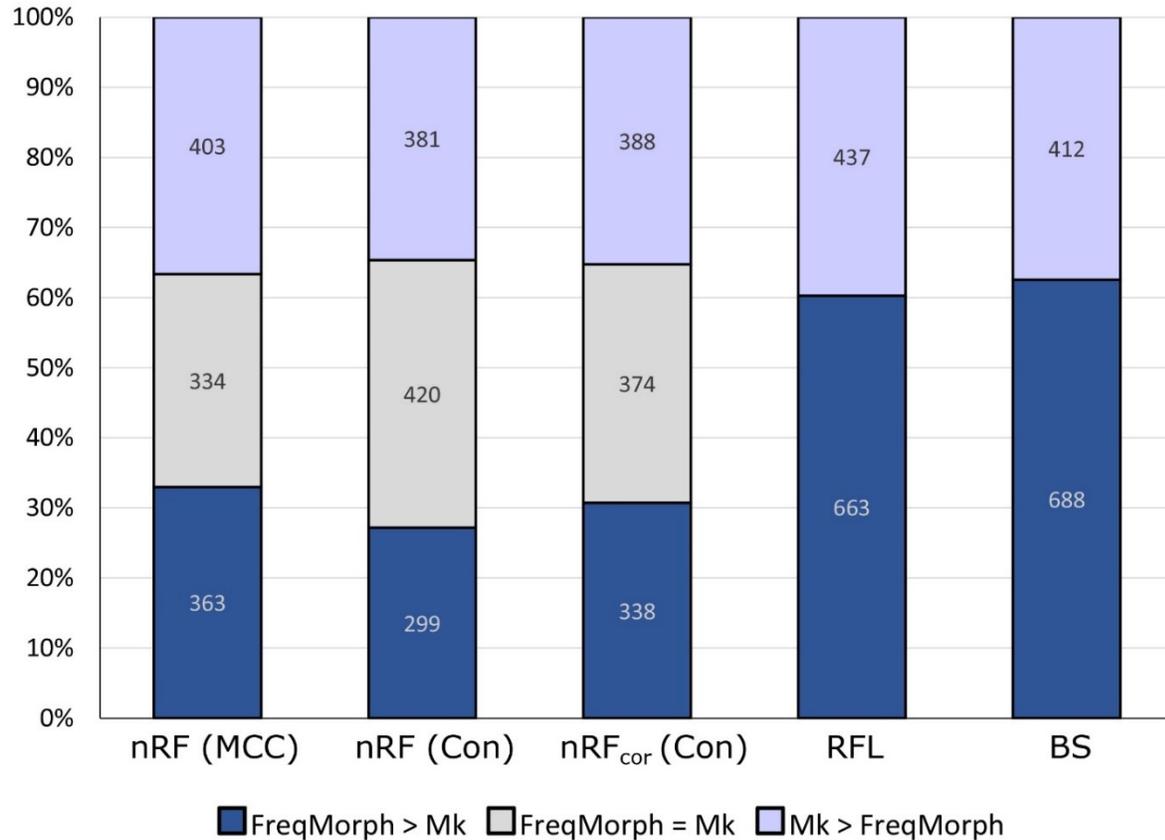


FIGURE 3-2. Stacked column charts showing results in a pooled sampled of 1100 simulated data sets using different metrics: normalized Robinson-Foulds distance for MCC trees [nRF(MCC)]; normalized Robinson-Foulds distance for consensus trees [nRF(Con)]; normalized Robinson-Foulds distance corrected for resolution of consensus topologies [nRF_{cor}], Robinson-Foulds distance with branch lengths [RFL], and branch score [BS]. Each stacked column represents 100% of analysed data sets (n=1100) with percentage values plotted on the y-axis. Numbers within each section of a column are absolute numbers of data sets that fall within each category of results. FreqMorph > Mk and Mk > FreqMorph signs indicate whether the FreqMorph or the Mk model showed more accurate results; FreqMorph = Mk indicate that both models performed with equal accuracy.

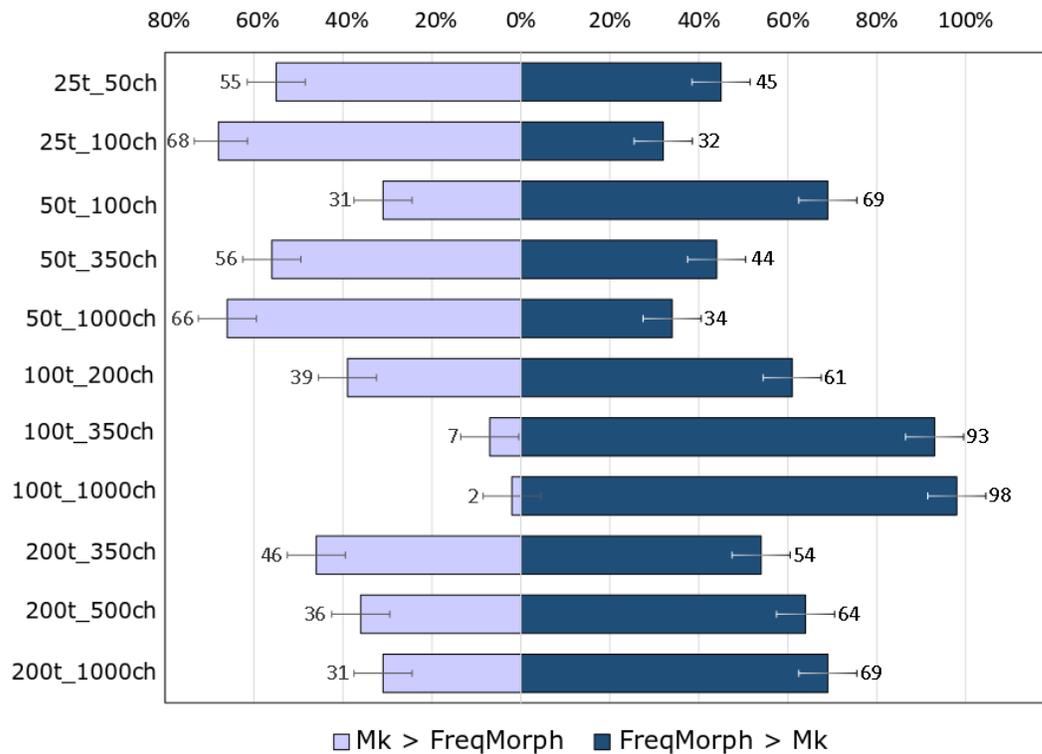


FIGURE 3-3. Comparison of the two models' phylogenetic accuracy under the RFL metric per individual categories of data sets. Data set specifications are abbreviated as follows taxa = t and characters = ch (25t_50ch specifies data set with 25 taxa and 50 characters).

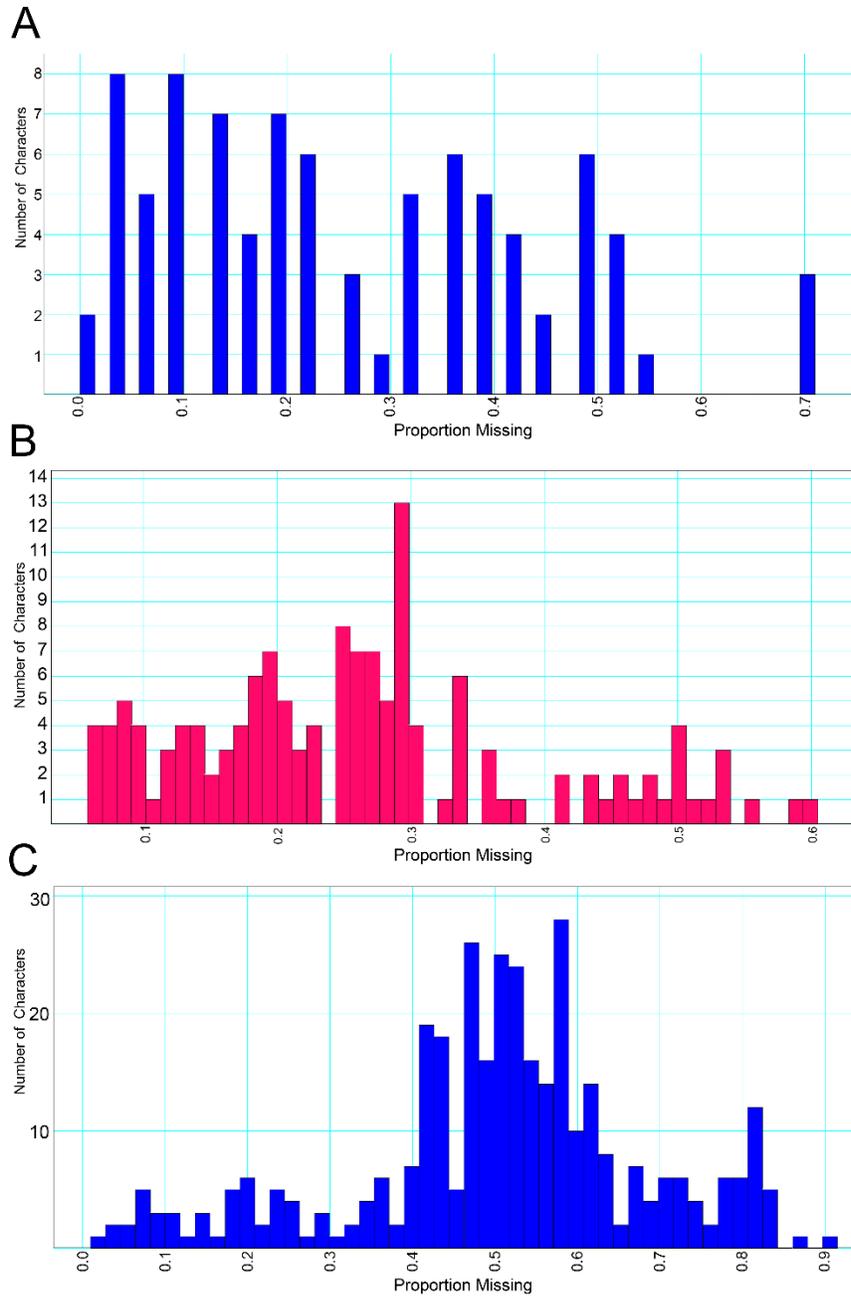


FIGURE 3-5. Proportion of missing data per character in the three empirical data sets analysed in the study: (A) – Osteoglossomorpha (n=87), mean=0.259, median=0.226; (B) – Angiospermae (n=142), mean=0.260, median=0.244; (C) – Hymenoptera (n=354), mean=0.505, median=0.509. Charts were generated in Mesquite v. 3.61 (Maddison and Maddison, 2015) using the Chart Wizard menu option.

APPENDIX 3-1. Source code of the FreqMorph model for the BEAST2 software.

```
package beast.evolution.substitutionmodel;
import beast.core.Citation;
import beast.core.Description;
import beast.core.Input;
import beast.core.Input.Validate;
import beast.core.util.Log;
import beast.core.parameter.RealParameter;
import beast.evolution.datatype.DataType;
import beast.evolution.alignment.Alignment;
import beast.evolution.alignment.FilteredAlignment;
import beast.evolution.datatype.Binary;
import beast.evolution.datatype.StandardData;
import beast.evolution.tree.Node;
import beast.evolution.alignment.TaxonSet;
import beast.evolution.substitutionmodel.Frequencies;
import java.util.Arrays;
import java.util.List;
import java.util.Collections;

/**
 * @author OksanaV
 *
 */
@Description("Morph-Model as implemented by A.Gavryushkina & J.Heled")
@Citation("Lewis, Paul O. A likelihood approach to estimating phylogeny from discrete
morphological character data. Systematic biology 50.6(2001): 913 - 925.")

public class FM extends SubstitutionModel.Base {

    public Input<Integer> nrOfStatesInput = new Input<Integer>("stateNumber", "the
number of character states");
    public Input<DataType> dataTypeInput = new Input<DataType>("datatype", "datatype,
used to determine the number of states", Validate.XOR, nrOfStatesInput);
    public Input<Boolean> estimateInput = new Input<>("estimate", "whether to estimate the
frequencies from data or assume uniform distribution over characters", true);
    public Input<TaxonSet> taxonSetInput = new Input<>("taxa", "An optional taxon-set
used only to sort the sequences into the same order as they appear in the taxon-set.", new
TaxonSet(), Validate.OPTIONAL);
    public Input<Alignment> dataInput = new Input<>("data", "Sequence data for which
frequencies are calculated");

    boolean hasFreqs;
    private boolean updateFreqs;
```

```

public FM () {frequenciesInput.setRule(Validate.OPTIONAL);
    try {
        } catch (Exception e) {e.printStackTrace();
            throw new RuntimeException("initAndValidate() call failed when
constructing FM()");
        }
    }
    double totalSubRate;
    double[] frequencies;
    int pattern;
    int counts;
    EigenDecomposition eigenDecomposition;
    private void setFrequencies() {
        final Frequencies frequencies1 = frequenciesInput.get();
        Frequencies frequencies2 = new Frequencies();
        if (frequencies1 != null) {
            if (frequencies1.getFreqs().length != nrOfStates) {
                throw new RuntimeException ("number of stationary frequencies
does not match number of states.");
            }
            System.arraycopy(frequencies1.getFreqs(), 0, frequencies, 0, nrOfStates);
            totalSubRate = 1;
            for (int k = 0; k < nrOfStates; ++k) {
                totalSubRate -= frequencies[k]*frequencies[k];
            }
            hasFreqs = true;
        } else {frequencies = frequencies2.getFreqs();
        }
        Log.info.println("Starting frequencies: " + Arrays.toString(frequencies));
        hasFreqs = false;
    }
    @Override
    public void initAndValidate() {
        if (nrOfStatesInput.get() != null) {
            nrOfStates = nrOfStatesInput.get();
        } else {nrOfStates = dataTypeInput.get().getStateCount();
        }
        frequencies = new double[nrOfStates];
        setFrequencies();
    }
    @Override
    public double[] getFrequencies() {
        return frequencies;
    }
    @Override

```

```

    public void getTransitionProbabilities(Node node, double fStartTime, double
fEndTime, double fRate, double[] matrix) {
        if( updateFreqs ) {
            setFrequencies();
        }
        if( hasFreqs ) {
            final double e1 = Math.exp(-(fStartTime - fEndTime) * fRate/totalSubRate);
            final double e2 = 1 - e1;
            for( int i = 0; i < nrOfStates; ++i ) {
                final int r = i * nrOfStates;
                for( int j = 0; j < nrOfStates; ++j ) {
                    matrix[r + j] = frequencies[j] * e2;
                }
                matrix[r + i] += e1;
            }
        } else {
            double fDelta = (nrOfStates / (nrOfStates - 1)) * (fStartTime - fEndTime);
            double fPStay = (1.0 + (nrOfStates - 1) * Math.exp(-fDelta * fRate)) / nrOfStates;
            double fPMove = (1.0 - Math.exp(-fDelta * fRate)) / nrOfStates;
            Arrays.fill(matrix, fPMove);
            for (int i = 0; i < nrOfStates; i++) {
                matrix[i * (nrOfStates + 1)] = fPStay;
            }
        }
    }

    @Override
    public EigenDecomposition getEigenDecomposition(Node node) {
        return eigenDecomposition;
    }

    @Override
    public boolean canHandleDataType(DataType dataType) {
        if (dataType instanceof StandardData || dataType instanceof Binary) {
            return true;
        }
        return false;
    }

    protected boolean requiresRecalculation() {
        if( ! hasFreqs ) {
            return false;
        }

        updateFreqs = true;
        return true;
    }
}

```

CHAPTER 4

Delimitation of *Alosa* species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches

A version of this chapter has been published as Vernygora, O.V., C. S. Davis, A. M. Murray, and F. A. H. Sperling. 2018. Delimitation of *Alosa* species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches. *Journal of Fish Biology* 93(6):1216–1228.

4.1 Introduction

Shad species of the genus *Alosa* Linck, 1790 are commercially important fishes in Europe and North America. The group includes approximately 16 species of anadromous, freshwater, and marine fishes (Whitehead, 1985; Nelson et al., 2016). The native range of the genus is split into three main regions – Western Atlantic, Northeast Atlantic and Mediterranean Sea, and the Ponto-Caspian Basin (Whitehead, 1985; Alexandrino et al., 2006). Each region is characterized by endemic species and subspecies of *Alosa* that constitute important components of regional fisheries (FAO, 2016).

Despite the economic significance of the genus *Alosa*, species delimitation within the group is challenging due to high interspecific morphological similarity, complex population structures and intraspecific phenotypic plasticity (McDowal, 2001, 2003; Alexandrino et al., 2006), and interspecific hybridization (Alexandrino et al., 2006; Faria et al., 2011; Jolly, 2011; McBride et al., 2014). Within the Mediterranean and Northeast Atlantic group the species status of *A. fallax* (Lacepède, 1803) and *A. alosa* (Linnaeus, 1758) have historically been a subject of

contention due to their close genetic relationships and extensive hybridization (Boisneau et al., 1992; Alexandrino et al., 2006; Faria et al., 2006; Coscia et al., 2010). Delimitation of some North American species have also been problematic especially for the morphologically similar species *A. pseudoharengus* (Wilson, 1811) and *A. aestivalis* (Mitchill, 1814), *A. sapidissima* (Wilson, 1811) and *A. alabamae* (Jordan and Evermann, 1896), as well as for delimitating ecomorphs of a single species (Chapman et al., 1994; Nolan et al., 2003; Czesny et al., 2012). Current understanding of the evolutionary relationships and taxonomy of North American, Mediterranean and Northern European species of *Alosa* has been advanced with extensive molecular studies (Alexandrino et al., 2006; Bowen et al., 2008; Czesny et al., 2012; Faria et al., 2004, 2006, 2011, 2012; Jolly et al., 2011, 2012; Hasselman et al., 2013); however, the Ponto-Caspian assemblage of *Alosa* species remains understudied. Relatively few molecular studies have investigated genetic interrelationships of shads from the Black Sea (Boyadzhieva-Doychinova et al., 2012; Keskin and Atar, 2013; Turan et al., 2015) and no studies have included samples from the Sea of Azov, an area characterized by distinct spawning stocks and the presence of resident forms of shads that likely do not enter the Black Sea (Pavlov, 1959; Svetovidov, 1964; Mezhzherin and Fedorenko, 2009).

In the Sea of Azov, three species of *Alosa* have been reported: two large, anadromous shad species, *A. immaculata* Bennett, 1835 and *A. maeotica* (Grimm, 1901), that spawn in rivers or brackish waters respectively, and a small semi-anadromous species, *A. caspia* (Eichwald, 1838) (Fig. 4-1) that only migrates short distances and spawns in estuaries. The first two species are endemic to the Black Sea and Sea of Azov, while the third was originally described from the Caspian Sea. A subspecies of *A. caspia*, *A. caspia tanaica* (Grimm, 1901), has been described from the Sea of Azov; however, the taxonomic status of *A. c. tanaica* is still debated

(Svetovidov, 1952, 1964; Whitehead, 1985; Hoestlandt, 1991; Kottelat, 1997; Kottelat and Freyhof, 2007; Mezhzherin et al., 2009). In the original descriptions, distinctions between species were based on the number of fin rays, colouration, and body proportions, i.e., eye diameter, body height, head length and head height. The two larger species, *A. immaculata* and *A. maeotica*, were described as having an elongate ‘herring-like’ body while *A. caspia* was characterized by a deep ‘shad-like’ body, large eyes, and a relatively short head (Bennet, 1835; Eichwald, 1838; Grimm, 1901; Svetovidov, 1952, 1964). The number of gill rakers was not indicated in the original descriptions of *A. caspia* and *A. immaculata*, and neither of the species initially had type specimens designated.

Extensive biological and fisheries research of the Ponto-Caspian ichthyofaunas in the mid to late 20th century revealed great similarity of the shad species from the Azov – Black Sea and Caspian Sea basins, prompting comprehensive comparative studies (e.g., Bulgakov, 1926; Svetovidov, 1945; Berg, 1949; Pavlov, 1959). Svetovidov (1952) amassed a knowledge of morphology, biology, and biogeography of the Ponto-Caspian species of *Alosa*; he designated neotype and lectotype specimens for *A. caspia* and *A. immaculata* and provided detailed descriptions including the number of gill rakers as a key diagnostic feature of each species. According to his classification, *A. caspia* can be distinguished from other congeneric species occurring in the Sea of Azov by large eyes (eye diameter is 24 – 28.5% of the head length, compared to 20 – 23.8% in *A. maeotica* and 18.9 – 23.6% in *A. immaculata*) and high number of gill rakers (59 – 85, compared to 33 – 47 in *A. maeotica* and 50 – 60 in *A. immaculata*). However, these morphological characteristics can be particularly prone to variation, for example in response to environmental factors, such as availability and type of food resources, type of habitat (marine vs. freshwater), migratory strategies, and predatory risks. Similar variability of

these morphological traits is observed in other anadromous species (e.g., Reist et al., 1995; Kahilainen et al., 2011; Czesny et al., 2012; Chavarie et al., 2013; Dodson et al., 2013; Roesch et al., 2013). Because of the potential uncertainty of species delimitation using these variable characters, I here test the congruence of the morphological species delimitation with molecular data for the species of *Alosa* in the Sea of Azov.

In the present study, I present novel mitochondrial cytochrome *b* (cyt *b*) and cytochrome oxidase subunit I (COI) sequence data, and report on SNP variation discovered using ddRAD sequencing (Peterson et al., 2012) in *Alosa* species from the Sea of Azov. The new genetic data are used to: (i) investigate phylogenetic relationships among *Alosa* species from the Black Sea and Sea of Azov basins; (ii) compare species delimitation based on morphological characters to genetic structure determined from molecular data; and (iii) evaluate the phylogenetic position of *Alosa* species from the Black Sea and Sea of Azov relative to the Mediterranean and North American members of the genus.

4.2 Material and methods

Individuals of *Alosa* spp. were opportunistically sampled as bycatch in commercial catches of other species, *Belone belone* (Linnaeus, 1761) and *Engraulis encrasicolus* (Linnaeus, 1758), in the north-western and southern areas of the Sea of Azov between May and June, 2016 (Fig. 4-2). Individuals were identified in the field to species level using the morphological characters previously proposed to distinguish species, in particular, the number of gill rakers on the first gill arch: *A. maeutica* with fewer than 40 gill rakers, *A. immaculata* with 40 – 59 gill rakers, and *A. caspia* with more than 59 gill rakers following Svetovidov (1952), Pavlov (1959), and Kottelat and Freyhof (2007). None of the collected individuals had a gill raker count lower than 40;

therefore, *A. maeotica* was not identified in the sample. Twenty specimens were selected for mitochondrial sequence analysis: four putative *A. caspia* and 16 putative *A. immaculata*. Meristic and morphometric characters traditionally used for species delimitation (Svetovidov, 1952; Pavlov, 1959; Mezhzherin and Fedorenko, 2005) were recorded for each sampled individual (Table 4-1, Fig. 4-3). Principal component analysis (PCA) of nine morphological characters (Table 4-2) was performed in the *factoextra* package for R 3.4.1 (R Core Team, 2017). To determine which morphological traits are most congruent with the genotypic grouping of the sampled individuals, I conducted a linear discriminant analysis using the MASS package (Venables and Ripley, 2002) for R.

4.2.1 DNA extraction and mitochondrial gene sequencing

Fin clips were collected in the field and stored in 96% ethanol until DNA extraction. Total genomic DNA was extracted from fin tissue using the DNeasy blood and tissue kit (QIAGEN) following the manufacturer's protocol. Partial coding regions of two mitochondrial genes, cytochrome oxidase subunit I (COI) and cytochrome *b* (cyt *b*) were amplified by PCR in 20 μ l reactions containing: 2 μ l of 10x buffer, 2 μ l of 25 mM MgCl₂, 0.4 μ l of 10 mM dNTPs, 0.4 μ l of 10 μ M forward and reverse primers, 0.04 μ l of TopTaq DNA polymerase (QIAGEN), 5 μ l of extracted DNA, and 9.76 μ l of ddH₂O. PCR amplification was conducted using the following conditions: (1) initial denaturation at 94°C for 2 min; (2) 35 cycles of denaturation at 94°C for 30 s; annealing at 54°C for 30 s (COI)/ 62°C for 30 s (cyt *b*); extension at 72°C for 1 min (COI)/ 72°C for 30 s (cyt *b*), and (3) final extension at 72°C for 5 min. Amplification and sequencing were performed using previously published primers: Alocytbfl and Alocytr1 for cyt *b* (Alexandrino et al., 2006) and Fish-F1 and Fish-R1 for COI (Keskin and Atar, 2013). PCR

products were purified using ExoSAP-IT and sequenced using BigDye Terminator v 3.1 cycle sequencing kit (Applied Biosystems). Sequencing reactions were purified by ethanol precipitation and resolved on an ABI 3730 DNA Analyzer (Applied Biosystems) in the Molecular Biology Sequencing Unit (MBSU), University of Alberta.

Consensus sequences from forward and reverse reads were generated in BioEdit 7.2.5 (Hall, 1999). Multiple sequence alignment of consensus sequences was performed in MAFFT 7.245 (Kato et al., 2002; Kato and Standley, 2013) under default settings.

4.2.2 Next generation ddRAD sequencing

A subsample of ten specimens was used for genome-wide SNP genotyping using a ddRAD sequencing approach. The subsample included four putative *A. caspia* and six putative *A. immaculata* specimens from the southern Sea of Azov, and two putative *A. immaculata* from the north-western Sea of Azov (Table 4-1).

Library construction and ddRAD sequencing conceptually followed the protocol of Peterson et al. (2012). Genomic DNA was digested with PstI and MspI. Adapter ligation used PstI/MspI specific linkers in place of the Peterson et al. (2012) flex adapters where PstI linkers contained an 8 bp in line Illumina Nextera XT i5 index (one of 16 S5XX index sequences) and the common MspI linker was forked. Ligation reactions were pooled across individuals possessing different i5 indexes. Illumina flowcell sequences, sequencing primer binding sites were introduced in a PCR reaction using a common forward primer and one of 24 Illumina i7 (N7XX) indexed reverse primers. PCR products were pooled over i7 indexes and the final library was purified size selected to 400-500 bp on a 1.5% agarose gel. The final ddRAD library was sequenced (single-end, 75 bp reads) on a single high output flowcell of an Illumina NextSeq 500.

4.2.3 ddRAD data processing

Raw reads were quality checked, demultiplexed, and sorted by individual samples using the *Stacks* v. 1.35 *process_radtags* program (Catchen et al., 2011). Reads with Phred quality score below 20 and those that failed Illumina's chastity filter were discarded. Individual inline index sequences (8 bp) and the PstI sites (5 bp) were removed, resulting in reads of 62 bp in length. PstI sites were trimmed using the *Cutadapt* v. 1.9.1 program (Martin, 2011).

Since no reference genome for any species of *Alosa* or a phylogenetically close taxon is available, I used a *de novo* approach to build loci. Trimmed sequence reads were aligned and stacked into putative loci using the *ustacks* program in *Stacks* with a minimum coverage depth of 3 reads per locus and a maximum number of 2 mismatches within the stacks. Loci were identified and catalogued in the *cstacks* program. Each individual was then genotyped at each locus to determine SNP-containing loci (performed in *sstacks*). To build the final catalog I allowed 3 mismatches between primary stacks and secondary reads and set a 0.05 chi-square significance level to call heterozygote or homozygote loci.

The final data file with information on each ddRAD locus containing SNPs for each individual was compiled in the *populations* program in *Stacks* (Catchen et al., 2011) and output in a .vcf format. Since few individuals per species and locality were sampled, all ten specimens were assigned to a single population for final processing with *populations*. I tested values of minimum coverage per individual between 100% and 10% (in increments of 10%) to evaluate the effect on the number of SNPs called. Minimum minor allele frequency was set to 0.2 to account for small sample size and minimize the probability of erroneous locus calls resulting from sequencing errors.

4.2.4 Phylogenetic analyses and divergence time estimation

To assess phylogenetic relationships among species of *Alosa* from the Sea of Azov and evaluate their phylogenetic position relative to the Mediterranean and North American members of the genus I performed a series of phylogenetic analyses using the following data sets: (i) individual and concatenated mitochondrial COI/cyt b data sets, including 20 samples sequenced in the present study and an outgroup species, *Alosa alosa* (GenBank accession NC009575), that is the closest, yet distinct, congeneric species to the ingroup taxa; (ii) SNP data set with no outgroup available; therefore, the resulting phylogeny was rooted at the mid-point; and (iii) individual and concatenated data sets for the mitochondrial markers (COI and cyt b) supplemented with GenBank sequences available for the Mediterranean and North American species of *Alosa*. I used the most closely related non-congeneric member of subfamily Alosinae, *Brevoortia tyrannus* (Latrobe, 1802), as an outgroup to the *Alosa* ingroup taxa (Li and Orti, 2007; Lavoué et al., 2013). The SNP data set included invariant flanking sequences for each SNP. Detailed information for each data set is presented in Table 4-3.

Phylogenetic analyses were performed using Maximum Likelihood (ML) implemented in IQ-Tree web server v. 1.3.10 (Minh et al., 2013; Nguyen et al., 2015). The best model of sequence evolution was selected based on AIC scores (Akaike, 1974) calculated by jModelTest v. 2.1.7 (Guindon and Gascuel, 2003; Posada, 2008; Dariba et al., 2012), which recovered TIM as the best-fit model for the first codon position and F81 as the best-fit model for the second and third codon positions in the molecular dataset. All ML analyses were performed under default settings with all model parameters estimated by the program. Node support was estimated by

ultrafast bootstrap (Minh et al., 2013) and the SH-aLRT test (Guidon et al., 2010) with 1000 replicates each.

A time-calibrated analysis of the concatenated COI/cyt b data set (58 ingroup taxa, 1099 bp) was performed in BEAST v2.4.6 (Bouckaert et al., 2014) under the Birth-Death model. Because there is uncertainty in the accurate identification of fossils as Alosinae or *Alosa*, a root node age was sampled from a uniform distribution over the range from 15 to 40 mya as estimated by the recent time-calibrated analyses of the Clupeoidei (Lavoué et al., 2013; Bloom and Lovejoy, 2014; Egan et al., 2018). All tip ages represented by extant taxa were set to 0 mya.

The partitioning scheme and the best-fit model for each partition were determined with PartitionFinder v 2.1.1 (Lanfear et al., 2016). I used the relaxed log-normal clock model with default settings. The MCMC analysis was run for 50,000,000 generations with trees sampled every 1000 generations. Results of the analysis were checked for convergence in Tracer v1.6 (Rambaut et al., 2014). The maximum clade credibility tree with the median node heights was built in TreeAnnotator v.2.4.6 (available with the BEAST 2 software) after 50% of the trees were removed as a burn-in.

4.3 Results

Two mitochondrial markers, COI and cyt b, were successfully sequenced for all 20 specimens, with fragment lengths of 651 and 448 bp respectively. The newly generated sequence data were deposited to GenBank under accession numbers MG200032 – 200051, MG490169-490188.

The ddRAD sequencing generated a total of 16,459,864 reads. After the initial quality check and filtering, 13,324,616 reads were retained (~81%). The initial *de novo* catalog

contained 155,275 putative loci with a mean read depth of 20. SNP data sets with the highest coverage values per individual (100 – 70%) all resulted in the same phylogenetic tree topology; therefore, I only present results for the maximum coverage (no missing data). The final SNP data set contained 806 markers (49972 bp). The raw reads data files were deposited to Sequence Read Archive (SRA) under accession numbers SAMN08057894–08057903.

4.3.1 PCA results

Results of the PCA indicate that the first three principal components accounted for 78.5% of the variance in the morphometric data (Fig. 4-4). The other six axes had eigen values less than 1.00 and were not considered in the further analyses following Kaiser's stopping rule (Kaiser, 1960).

The PCA scatterplot shows that the two putative species identified in my sample, *A. caspia* and *A. immaculata*, occupy the same morphospace and do not form separate morpho-clusters (Fig. 4-4). The first principal component was most strongly negatively correlated with the preanal length (loading = -0.85) as well as with the eye diameter (loading = -0.65) and positively correlated with head length, predorsal, and prepelvic lengths (loadings > 0.65). The second principal component was positively correlated with the maximum body depth (loading = 0.71) and negatively correlated with the predorsal length (loading = - 0.68). Variance along the third axis was attributed to the interorbital width (loading = -0.62) and number of gill rakers (loading = - 0.58). Gill raker counts showed a strong positive correlation with PC 4 (loading = 0.72); that principal component, however, accounted for only 9% of the variance in the data set.

4.3.2 Phylogenetic results

Maximum likelihood analyses of individual COI and cyt b data sets recovered overall identical topologies (Appendices 4-1 and 4-2); therefore, I concatenated COI and cyt b sequences into a single data set (Fig. 4-5A) which recovered two well-supported clades (bootstrap values >80%) within my sample of 20 *Alosa* specimens. One clade comprised seven individuals sampled from the north-western part of the Sea of Azov and a sister clade comprised all ten individuals sampled from the southern part of the Sea of Azov along with three north-western specimens (AZ-VI, AZ-VII, AZ-VIII). Notably, individuals identified as putative *A. caspia* based on the number of gill rakers on the first gill arch do not form a monophyletic group, but instead each have a sister-group relationship with a putative *A. immaculata* individual.

Similar patterns are recognized in the topology recovered by the maximum likelihood analysis of the SNP data set (Fig. 4-5B), with two well defined sister groups – a north-western *Alosa* group (AZ-I and AZ-II) and a clade comprising southern individuals with the four specimens of putative *A. caspia* not forming a monophyletic group.

In the ML topologies retrieved by the analyses of the COI, cyt b, and concatenated (COI + cyt b) data sets supplemented with GenBank sequences for the North American and European species of *Alosa* (Appendix 4-3), individuals sampled from the Sea of Azov form a well-supported clade (bootstrap >75%) similar in structure and composition to the topology of the COI/cyt b tree in Figure 4-5. It is important to note that there is no sequence data available for *A. maeotica*; therefore, its phylogenetic position was not assessed in my analyses. Sequences of *A. caspia* and *A. immaculata* from individuals sampled in the Black Sea retrieved from GenBank are deeply nested within a clade with my samples from the Sea of Azov, therefore confirming close genetic relationships between the Black Sea and the Sea of Azov shad populations.

Notably, *A. macedonica* and *A. vistonica*, species endemic to freshwater lakes in Greece, are also recovered deeply nested in a clade with the specimens from the Black Sea and Sea of Azov. The concatenated mitochondrial tree topology shows a moderate degree of geographical structure with the European (*A. alosa*, *A. agone* (Scopoli, 1786), and *A. fallax*) and North American (*A. sapidissima*, *A. pseudoharengus*, *A. mediocris* (Mitchill, 1814), *A. alabamae*, and *A. aestivalis*) forming mostly paraphyletic species. This geographic clustering is more evident in the ML topology of the COI only data set, whereas the ML topology for the cyt b data set shows less defined geographic clusters of *Alosa* species (Appendices 4-1 and 4-2).

4.3.3 Linear discriminant analysis results

We used the composition of the clades recovered in the phylogenetic analyses of the concatenated mtDNA data set to assign individuals to two groups, the north-western Sea of Azov or the southern Sea of Azov. Results of this analysis indicate that the two groups show a fair amount of morphological differentiation with almost a clear separation between the two clusters of individuals (Fig. 4-6). A single outlier (AZ-IV) was recovered in the sample. Linear discriminant coefficients indicate that the morphological disparity between the two groups is mainly attributed to the caudal peduncle depth, head length, and preanal length (LD coefficients - 1.41, -0.66, -0.57, respectively). Individuals of the north-western group (excluding the AZ-IV specimen) are characterized by a narrower caudal peduncle (\bar{x} = 8.43% of the standard length, SL), shorter head (\bar{x} = 25% of SL) and greater preanal length (\bar{x} = 72.65% of SL) while individuals of the southern group had a thicker caudal peduncle (\bar{x} = 9.03% of SL), more elongate head (\bar{x} = 26.89% of SL), and shorter preanal distance (\bar{x} = 71.98% of SL).

4.4 Discussion

Analyses of all data sets tested in my study retrieved topologies that indicate two common patterns: (i) morphological species assignments based on the number of gill rakers are not congruent with the genetic clustering as indicated by both mitochondrial markers and genome-wide SNPs; and (ii) there is evidence of geographical genetic structure within the *Alosa* species complex in the Sea of Azov that is evident from the consistent recovery of two well-supported clades, one consisting of individuals sampled only in the north-western part of the Sea of Azov and another group comprising specimens from the southern and north-western areas of the sea.

The observed discrepancy between morphological identification based on characters of Svetovidov (1952) and my molecular analyses challenges the species status of traditionally recognized groups of *Alosa* in the region. In my sample, two groups, representing individuals identified as *A. immaculata* and *A. caspia* based on the number of gill rakers, showed extensive overlap in other features of their morphology and in their genetics (Fig. 4-4). The eye diameter, another diagnostic feature used to delimit species of *Alosa*, significantly contributed to the first principal component (Fig. 4-4) but it was lower than values assigned to either of the species (Svetovidov 1952), closer to the values described for *A. maeotica*. The four putative *A. caspia* specimens do not form a distinct phenotypic cluster and can only be distinguished from the other specimens based on the single morphological criterion of gill raker count.

Large amounts of intraspecific morphological trait variation is common in diadromous fish with complex population structures (e. g., McCart and Andersen, 1967; Walker, 1997; Narum et al., 2004; Østbye et al., 2005; Michel et al., 2008; Palkovacs et al., 2008; Dodson et al., 2012; Wund et al., 2012; Bakhshalizadeh and Bani, 2017). This variation represents an adaptive response to environmental conditions when a species is fragmented into spatially and/or

temporally separated stocks. Morphometric traits are normally more prone to change than meristic characters (Lindsey, 1981; Beacham, 1990; Swain and Foote, 1999; Robinson and Parsons, 2002); however, gill raker number has been shown to vary among different populations of a single species (Lindsey, 1981; Swain and Foote, 1999). Coscia et al. (2010) found that gill raker count was a largely inadequate criterion for identifying *A. alosa*, *A. fallax*, and their hybrids; using these counts resulted in a 16% misidentification rate in their study sample relative to the results of the genetic analyses based on the microsatellite and mtDNA markers.

In other species of *Alosa*, phenotypic variation has been used to differentiate among populations, stocks, and ecomorphs of *A. sapidissima*, *A. fallax*, and *A. pseudoharengus* (Melvin et al., 1992; Turan and Basusta, 2001; Czesny et al., 2012; Jones et al., 2013). These species show intraspecific morphological variation comparable to that observed among the putative species in the Sea of Azov. The uncertain taxonomic status of *Alosa* shads from the Black Sea and Sea of Azov has been previously reported in a few studies that examined morphological and allozyme diversity within the *Alosa* species complex and it was suggested that the observed variation can be explained by intraspecific plasticity, and that multiple species described from the Black Sea and Sea of Azov basin might better be considered morphotypes of a single species (Pravotorov, 2000; Mezhzherin et al., 2009).

In cases when traditional morphology-based taxonomy fails to identify clear species boundaries, an integrative taxonomic approach provides a more holistic picture of biological systems by combining molecular and morphological information (Borsa, 2002; Dayrat, 2005; Schlick-Steiner et al., 2010; Victor, 2010; Fujita et al., 2012; Thomas et al., 2014; Stern et al., 2016). Phylogenetic analyses of the COI and cyt b sequences generated in my study recovered two distinct clades that do not correspond to the morphological groupings based on numbers of

gill raker. Instead, the two clades predominantly show an association with the sampling locality and have a fair degree of morphological divergence based on morphometric traits including the caudal peduncle depth, preanal length, and head length. These body proportions provide a good overall description of body shape and, based on the congruence with the genetic data, are better identifiers of distinct morphotypes than the number of gill rakers.

The average genetic divergence among individuals within those two major clades was relatively low for both COI and cyt b markers, being 0.47% ($\bar{x} = 3$ bp) and 0.34% ($\bar{x} = 1.5$ bp) in the north-western and southern clades, respectively. These values fall within the range of intraspecific nucleotide divergence reported for *A. alosa*, *A. caspia*, *A. fallax*, *A. immaculata*, and *A. sapidissima* based on mtDNA sequences (Faria et al., 2006; Bowen et al., 2008; Turan et al., 2015), and indicate that each population comprises a single species. The previously reported interspecific divergence between species ranged from 1.10% (*A. fallax* – *A. immaculata*) to 2.99% (*A. alosa* – *A. sapidissima*), which is similar to the divergence between my two major clades (0.82 – 1.77%). Therefore, the southern and north-western populations in the Sea of Azov would correspond to separate species.

Fast evolving mtDNA can, however, mislead determination of species boundaries, especially given a limited number of markers and a spatial and temporal population structure such as is common for *Alosa* species (Meyer, 1994; Moritz and Cicero, 2004; Alexandrino et al., 2006; Dupuis et al., 2012; Faria et al., 2012). When considering the across-genome SNP divergence within the subsampled group, I found overall an extremely low degree of genetic divergence (0.08 – 0.10% [40 – 49.9 bp] across both clades), which would not be expected if the two populations represent two different species. It should be noted, however, that the recent genome-wide study of a related clupeid species, *Clupea harengus*, showed that this species has

the lowest mutation rates among all vertebrates examined to date (Feng et al., 2017). It is possible that the low mutation rates are also characteristic of other clupeid taxa, which could explain the small nucleotide divergence of *Alosa* species, and can support the conclusion that the two Sea of Azov populations do indeed represent separate species.

The relatively large divergence of mtDNA sequences compared to the genome-wide set of SNPs from the Sea of Azov specimens could have resulted from an introgressive hybridization event such as has been reported for the closely related Mediterranean species *A. fallax* and *A. alosa* (Alexandrino et al., 2006; Coscia et al., 2010; Jolly et al., 2011; Faria et al., 2012). It is conceivable that introgressive contact may have occurred between different species due to their limited spawning area in the Sea of Azov together with the inaccessibility and disturbance of their preferred spawning grounds by dams that block upstream migrations of the anadromous shads, forcing them to share spawning sites with the semi-anadromous congeners. Analogously, the current construction of a bridge across the Kerch Strait, which connects the Black Sea and Sea of Azov, blocks migration routes for the anadromous Black Sea shads that enter the Sea of Azov to access their spawning, feeding and nursing grounds. Consequences of such habitat disturbances have yet to be evaluated. It is likely that such anthropogenic barriers on the major migration route may result in unsuccessful spawning seasons for diadromous *A. immaculata*, causing stock depletion.

4.4.1 Divergence time estimation

Monophyly of the genus *Alosa* has been repeatedly supported based on mitochondrial, nuclear, and combined molecular data (Faria et al., 2006; Li and Orti, 2007; Bowden et al., 2008; Lavoué et al., 2013); however, divergence times and rates of evolution have been a subject of

debate (Bentzen et al., 1993; Faria et al., 2006; Lavoué et al., 2013; Bloom and Lovejoy, 2014). The time-calibrated analyses (Fig. 4-7) retrieved a topology identical to the composition of the major clades. The estimated median age of the root is 20.1 million years and is consistent with the recent hypotheses on the divergence between *Alosa* and *Brevoortia* lineages in the early Miocene (23.8 – 16.4 mya). Previous biogeographic studies that focused on large-scale analysis of the clupeiforms suggested that the common ancestor of *Alosa* and *Brevoortia* might have originated in the Western Atlantic, followed by subsequent eastward dispersal that established European groups of shads (Lavoué et al., 2013; Bloom and Lovejoy, 2014). This hypothesis offers a parsimonious biogeographic scenario corroborated by the restricted present-day distribution of *Brevoortia* on the Atlantic coast of the Americas. An alternative scenario would be the presence of a widespread amphi-Atlantic common ancestor that gave rise to the North American and European lineages through a series of variance and dispersal events. Although beyond the scope of the present study, testing these hypotheses of the origin and divergence within the *Brevoortia* – *Alosa* clade would require careful consideration and reassessment of the fossil record of each lineage.

An early Miocene (23.8 – 16.4 mya) marine connection between the Eastern Paratethys and the Atlantic Ocean could have allowed the amphi-Atlantic *Alosa* to give rise to two major lineages, North American shads excluding *A. sapidissima* and European shads including the ancestor of *A. sapidissima* and its relatives (Fig. 4-7, node C) coinciding with a gradual closing of the Eastern Tethys during the late middle and middle Miocene (12 – 5.7 mya). The subsequent colonization of the Atlantic coast of North America by the common ancestor of *A. sapidissima* and *A. alabamae* likely happened in the late Miocene or early Pliocene (median estimate 5.2 mya; Fig. 4-7, node B). During that time (ca 5.96 – 5.33 mya), the Mediterranean region

experienced the Messinian salinity crisis which was characterized by progressive isolation of the Mediterranean basin from the Atlantic Ocean, dramatic sea level drop, and salt deposition (Popov et al., 2004; Krijnsman et al., 2010; Bianchi et al., 2012). The lineage of *Alosa* that remained in the landlocked Mediterranean basin had to exhibit remarkable euryhalinity to be able to survive in the hypersaline environment at the beginning of the salinity crisis followed by the hyposaline conditions during the ‘Lago Mare’ (=Lake Sea) event (Bianchi et al., 2012). Such a transition seems feasible for *Alosa* considering its high tolerance along a broad salinity gradient, and the ability to rapidly adapt to freshwater environments in landlocked populations observed in extant species (Palkovacs et al., 2008; Lavoué et al., 2013; Bloom and Lovejoy, 2014; Velotta et al., 2015). Subsequent reestablishment of the Mediterranean – Atlantic Ocean connection, together with fragmentation of the greater Mediterranean basin into a series of landlocked seas (Black Sea, Sea of Azov, Caspian and Aral Seas), would have subdivided the European *Alosa* complex into areas in which populations would have evolved in isolation, becoming endemic to their regions (Fig. 4-7, node A). Recent divergence times among the endemic European species, *A. vistonica*, *A. macedonica*, *A. caspia*, and *A. immaculata*, are indicated by their extremely low amount of genetic divergence from one another. Close phylogenetic relationships between the endangered shads that are endemic to Greece (*A. vistonica* and *A. macedonica*) and those from the Black Sea and Sea of Azov basin have significant implications for conservation measures aimed at preserving phylogenetic diversity of shads from the Black Sea, Sea of Azov, and Greek lakes as part of the same species complex.

The results of my time calibrated phylogenetic analysis suggest higher rates of the mtDNA evolution (~0.70% per Myr) than it has been estimated previously using fossil calibrations (Bentzen et al., 1993). Bentzen et al. (1993) estimated mtDNA substitution rates at 0.22% per

Myr for *Alosa* species, but they questioned their estimate and suggested that the fossil calibration of the clade ages might not be accurate. Although, such a low evolutionary rate is indeed realistic, given that the mtDNA divergence rates reported in sharks and some groups of teleost fishes can be as low as 0.15×10^{-3} substitution per site per Myr (Martin et al., 1992; Cantatore et al., 1994), it is conceivable that the fossils used to calibrate age estimations of *Alosa* belong to a different evolutionary lineage.

4.5 Conclusions

Our study shows that the traditional taxonomic delimitation of species of *Alosa* in the Sea of Azov based on the single meristic character of gill raker count does not reflect the underlying genetic structure of the group. Considering the high degree of morphological homogeneity and low genetic divergence between shads identified as *A. caspia* and *A. immaculata* in my sample, I conclude that individuals collected from the same geographical region and characterized by similar values of the preanal length, caudal peduncle depth, and head length belong to a single lineage despite the observed disparity in the gill raker counts. The taxonomic status of this lineage should be further investigated in order to establish whether it represents a separate species. Higher levels of genetic divergence between the two clades recovered in my study, one representing the north-western region and the other representing the southern region, indicate that the *Alosa* species complex in the Sea of Azov is, in fact, structured. The underlying natural structure is, however, not reflected by the designated ranges of the single meristic feature of gill raker count. Failure to discover and recognize natural structure may lead to inefficient stock management. Fisheries regulations in the region are primarily concerned with identifying stocks

based on a particular meristic-based morphotype designation of species of *Alosa*; they should instead focus on preserving phylogenetic and ecological diversity of the species complex.

Such taxonomic uncertainty resulting in frequent misidentifications of species leads to poor conservation and stock management. An integrative approach is required to establish reliable diagnoses for each species of *Alosa* that can be used in the field by fishers and ensure proper stock protection and exploitation measures. Although the sample sizes and geographical sampling of my study is somewhat limited and does not allow an immediate conclusion on the taxonomic status of the recovered genetic groups, I present here the first genetic resources for the endemic Ponto-Caspian shads from the Sea of Azov and show the importance of gaining a true understanding of population structures. Based on the results of this investigation, I recommend further research, including: (i) re-examination of the type specimens and revision of the species diagnoses for the taxa in question, *A. immaculata*, *A. caspia*, *A. maeotica*, as well as other species and subspecies of *Alosa* reported from the Ponto-Caspian basin; (ii) in-depth investigation of genetic structure of the *Alosa* species complex in the Black Sea and Sea of Azov employing greater temporal and geographical sampling and targeting a rare species not sampled in the present study – *A. maeotica*; (iii) developing a framework for sustainable fisheries in the region that would promote conservation of shad species endemic to the Black Sea and Sea of Azov.

References

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on automatic control* 19:716–723.
- Alexandrino, P., R. Faria, D. Linhares, F. Castro, M. Le Corre, R. Sabatié, J. L. Bagliniere, and Weiss, S. 2006. Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. *Journal of Fish Biology* 69:242–259.
- Bakhshalizadeh, S. and A. Bani. 2017. Use of geometric morphometrics to identify ecophenotypic variation of juvenile Persian sturgeon *Acipenser persicus*. *Scientia Marina* 81:187–193.
- Beacham, T. D. 1990. A genetic analysis of meristic and morphometric variation in chum salmon (*Onchorynchus keta*) at three different temperatures. *Canadian Journal of Zoology* 68:225–229.
- Bennett, E. T. 1835. Characters of several previously undescribed fishes from Trebizond, collected by Keith Abbott. *Proceedings of the Zoological Society of London B* 3:91–92.
- Bentzen, P., W. C. Leggett, and G. G. Brown. 1993. Genetic relationships among the shads (*Alosa*) revealed by mitochondrial DNA analysis. *Journal of Fish Biology* 43:909–917.
- Berg, L. S. 1948. *Freshwater fishes of the USSR and adjacent countries*. vol. 1. Izdatelstvo Akademii Nauk SSSR, Moskva and Leningrad, 504 pp.
- Bianchi, C. N., C. Morri, M. Chiantore, M. Montefalcone, V. Parravicini, and A. Rovere. 2012. Mediterranean Sea biodiversity between the legacy from the past and a future of change; pp1–55 N. Stambler (ed.), *Life in the Mediterranean Sea: a look at habitat changes*. New York, NY: Nova Science Publishers, 749 pp.

- Bloom, D. D. and N. R. Lovejoy. 2014. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proceedings of the Royal Society B* 281:10.1098/rspb.2013.2081.
- Boisneau, P., C. Mennesson-Boisneau and R. Guyomard. 1992. Electrophoretic identity between allis shad *Alosa alosa* L. and twaite shad, *Alosa fallax* (Lacepede). *Journal of Fish Biology* 40:731–738.
- Borsa, P. 2002. Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of the Linnean Society* 75:261–269.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* 10:e1003537.
- Bowen, B. R., B. R. Kreiser, P. F. Mickle, J. F. Schaefer, and S. B. Adams. 2008. Phylogenetic relationships among North American *Alosa* species (Clupeidae). *Journal of Fish Biology* 72:1188–1201.
- Boyadzhieva-Doychinova, D., F. Tserkova, M. Gevezova, D. Klisarova, and I. Denev. 2012. Application of variability in its sequences for studying biodiversity of Black Sea clupeid fishes (Pisces: Clupeidae). *Acta Zoologica Bulgarica* 4:95–102.
- Bulgakov, G. P. 1926. To the knowledge of Caspian herrings. Description of a new species of the genus *Caspialosa*. *Byulleten' Sredneaziatskii Gosudarstvennyi Universitet, Taschkent* 13:27–39.

- Cantatore, P., M. Roberti, G. Pesole, A. Ludovico, F. Milella, M. N. Gadaleta, and C. Saccone. 1994. Evolutionary analysis of cytochrome b sequences in some Perciformes: evidence for a slower rate of evolution than in mammals. *Journal of Molecular Evolution* 39:589–597.
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* 1:171–182.
- Chapman, R. W., J. C. Patton, and B. Eleby. 1994. Comparisons of mitochondrial DNA variation in four alosid species as revealed by the total genome, the NADH dehydrogenase I and cytochrome b regions; pp. 249–262 in A. Beaumont (ed.), *Genetics and Evolution of Aquatic Organisms*. Chapman and Hall, London. 539 pp.
- Chavarie, L., K. L. Howland, and W. M. Tonn. 2013. Sympatric polymorphism in Lake Trout: the coexistence of multiple shallow-water morphotypes in Great Bear Lake. *Transactions of the American Fisheries Society* 142:814–823.
- Coscia, I., V. Rountree, J. J. King, W. K. Roche, and S. Mariani. 2010. A highly permeable species boundary between two anadromous fishes. *Journal of fish biology* 77:1137–1149.
- Czesny, S., J. Epifanio, and P. Michalak. 2012. Genetic divergence between freshwater and marine morphs of alewife (*Alosa pseudoharengus*): a ‘next-generation’ sequencing analysis. *PloS ONE* 7:e31803.
- Dariba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:772.
- Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85:407–415.

- Dodson, J. J., N. Aubin-Horth, V. Thériault, and D. J. Páez. 2013. The evolutionary ecology of alternative migratory tactics in salmonid fishes. *Biological Reviews* 88:602–625.
- Dupuis, J. R., A. D. Roe, and F. A. H. Sperling. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology* 21:4422–4436.
- Egan, J. P., D. D. Bloom, C.-H. Kuo, M. P. Hammer, P. Tongnunu, S. P. Iglesias, M. Sheaves, C. Grudpan, and A. M. Simons. 2018. Phylogenetic analysis of trophic niche evolution reveals a latitudinal herbivory gradient in Clupeoidei (herrings, anchovies, and allies). *Molecular Phylogenetics and Evolution* 124:151–161.
- Eichwald, C. E. von. 1838. Faunae Caspii maris primitiae. *Bulletin de la Société Impériale des Naturalistes de Moscou* 11:125–174.
- FAO. 2016. FAO Yearbook. Fishery and Aquaculture Statistics. 2014. Rome/Roma, FAO. 105 pp.
- Faria, R., A. Pinheiro, T. Gabaldón, S. Weiss and P. Alexandrino. 2011. Molecular tools for species discrimination and detection of hybridization between two closely related Clupeid fishes *Alosa alosa* and *A. fallax*. *Journal of Applied Ichthyology* 27:16–20.
- Faria, R., B., Wallner, S. Weiss, and P. Alexandrino. 2004. Isolation and characterization of eight dinucleotide microsatellite loci from two closely related clupeid species (*Alosa alosa* and *A. fallax*). *Molecular Ecology Notes* 4:586–588.
- Faria, R., S. Weiss, and P. Alexandrino. 2006. A molecular phylogenetic perspective on the evolutionary history of *Alosa* spp. (Clupeidae). *Molecular Phylogenetics and Evolution* 40:298–304.

- Faria, R., S. Weiss, and P. Alexandrino. 2012. Comparative phylogeography and demographic history of European shads (*Alosa alosa* and *A. fallax*) inferred from mitochondrial DNA. *BMC Evolutionary Biology* 12:10.1186/1471-2148-12-194.
- Feng, C., M. Pettersson, S. Lamichhaney, C.-J. Rubin, N. Rafati, M. Casini, A. Folkvord, and L. Andersson. 2017. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *eLife*:6:e23907.
- Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology and evolution* 27:480–488.
- Grimm, O. von. 1901. The herrings of the Sea of Azov. *Vestnik Rybopromyshlennosti* 16:57–70.
- Guindon, S. and O. Gascuel. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52:696–704.
- Guindon, S., J. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59:307–321.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nuclear Acids Symposium Series* 41:95–98.
- Hasselman, D. J., D. Ricard, and P. Bentzen. 2013. Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular ecology* 22:1558–1573.
- Hoestlandt, H. 1991. The freshwater fishes of Europe. Vol. 2. Clupeidae, Anguillidae. Aula, Wiesbaden. 447 pp.

- Jolly, M. T., M. W. Aprahamian, S. J. Hawkins, P. A. Henderson, R. Hillman, N. O'Maoiléidigh, and M. J. Genner. 2012. Population genetic structure of protected allis shad (*Alosa alosa*) and twaite shad (*Alosa fallax*). *Marine biology* 159:675–687.
- Jolly, M. T., P. S. Maitland, and M. J. Genner. 2011. Genetic monitoring of two decades of hybridization between allis shad (*Alosa alosa*) and twaite shad (*Alosa fallax*). *Conservation Genetics* 12:1087–1100.
- Jones, A. W., E. P. Palkovacs, and D. M. Post. 2013. Recent parallel divergence in body shape and diet source of alewife life history forms. *Evolutionary Ecology* 27:1175–1187.
- Jordan, D. S. and B. W. Evermann. 1896. Description of a new species of shad (*Alosa alabamae*) from Alabama; pp. 203–205 in B. W. Evermann, Report to the United States Commission on Fisheries. G.P.O., Washington, 610 pp.
- Kahilainen, K. K., A. Siwertsson, K. Ø. Gjelland, R. Knudsen, T. Bøhn, and P. A. Amundsen. 2011. The role of gill raker number variability in adaptive radiation of coregonid fish. *Evolutionary Ecology* 25:573–588.
- Kaiser, H. F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20:141–151.
- Katoh, K. and D. M. Standley. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.
- Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066.

- Keskin, E. and H. H. Atar. 2013. DNA barcoding commercially important fish species of Turkey. *Molecular Ecology Resources* 13:788–797.
- Kottelat, M. 1997. European freshwater fishes. An heuristic checklist of the freshwater fishes of Europe (exclusive of former USSR), with an introduction for non-systematists and comments on nomenclature and conservation. *Biologia, Bratislava, Section Zoology* 52:1–271.
- Kottelat, M. and J. Freyhof. 2007. Handbook of European freshwater fishes. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany, 646 pp.
- Krijgsman, W., M. Stoica, I. Vasiliev, and V. V. Popov. 2010. Rise and fall of the Paratethys Sea during the Messinian Salinity Crisis. *Earth and Planetary Science Letters* 290:183–191.
- Lacepède, B. G. E. 1803. Histoire naturelle des poissons. Tome cinquième. Chez Plassan, Paris, 803 pp.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular biology and evolution* 34:772–773.
- Latrobe, B. H. 1802. A drawing and description of the *Clupea tyrannus* and *Oniscus praegustator*. *Transactions of the American Philosophical Society* 5:77–81.
- Lavoué, S., M. Miya, P. Musikasinthorn, W. J. Chen, and M. Nishida. 2013. Mitogenomic evidence for an Indo-West Pacific origin of the clupeioidi (Teleostei: Clupeiformes). *PLoS ONE* 8:e56485.
- Li, C. and G. Orti. 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 44:386–398.

- Linck, H. F. 1790. Versuch einer Eintheilung der Fische nach den Zähnen. Magazin für das Neueste aus der Physik und Naturgeschichte, Gotha 6:28–38.
- Lindsey, C. C. 1981. Stocks are chameleons: plasticity in gill rakers of coregonid fishes. Canadian Journal of Fisheries and Aquatic Sciences 38:1497–1506.
- Linnaeus, C. 1758. Systema Naturae per Regna Tria Naturae, secundum Classes, Ordines, Genera, Species, cum Characteribus, Differentiis, Synonymis, Locis. Tonus I. Editio Decima, Reformata. Laurentii Salvii, Holmiae, 824 pp.
- Linnaeus, C. 1761. Fauna Suecicasitens animalia Suecicaeregni: quadrupedia, aves, amphibian, pisces, insect, vermes. Impensis Direct. Laurentii Salvii, Stockholm, 579 pp.
- Martin, A. P., G. J. P. Naylor, and S. R. Palumbi. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357:153–155.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17:10–12.
- McBride, M. C., T. V. Willis, R. G. Bradford, and P. Bentzen. 2014. Genetic diversity and structure of two hybridizing anadromous fishes (*Alosa pseudoharengus*, *Alosa aestivalis*) across the northern portion of their ranges. Conservation Genetics 15:1281–1298.
- McCart, P. and B. Andersen. 1967. Plasticity of gillraker number and length in *Oncorhynchus nerka*. Journal of the Fisheries Research Board of Canada 24:1999–2002.
- McDowall, R. M. 2001. Diadromy, diversity and divergence: implications for speciation process. Fish and Fisheries 2:278–286.
- McDowall, R. M. 2003. Shads and anadromy: implications for ecology, evolution, and biogeography; pp. 11–23 in K. E. Limburg and J. R. Waldman (eds.), Biodiversity, Status,

and Conservation of the World's Shads. Baltimore, MD: American Fisheries Society, 370 pp.

- Melvin, G. D., M. I. Dadswell, and A. McKennie. 1992. Usefulness of meristic and morphometric characters in discriminating populations of American shad (*Alosa sapidissima*) (Ostreichthyes:Clupeidae) inhabiting a marine environment. *Canadian Journal of Fisheries and Aquaculture Sciences* 49:266–280.
- Mezhzherin, S. V. and L. V. Fedorenko. 2005. Genetic structure of the spawning population of Danube shad, *Alosa pontica*, Eichwaldt 1838 (Clupeiformes, Alosiinae). *Tsitologiya i genetika* 39:35–40.
- Mezhzherin, S. V., L. V. Fedorenko, and D. B. Verlatyi. 2009. Differentiation and allozyme variability of shads genus *Alosa* (Clupeiformes, Alosiinae) from Azov-Black Sea basin. *Cytology and genetics* 43:118–122.
- Michel, C., B. J. Hicks, K. N. Stolting, A. C. Clarke, M. I. Stevens, R. Tana, A. Meyer, and M. R. van den Heuvel. 2008. Distinct migratory and non-migratory ecotypes of an endemic New Zealand eleotrid (*Gobiomorphus cotidianus*)—implications for incipient speciation in island freshwater fish species. *BMC Evolutionary Biology* 8:49–63.
- Minh, B. Q., M. A. T. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30:1188–1195.
- Mitchill, S. L. 1814. Report, in part, of Samuel L. Mitchill, M. D., Professor of Natural History, on the fishes of New-York. D. Carlisle, New York, 28 pp.
- Moritz, C. and C. Cicero. 2004. DNA barcoding: promises and pitfalls. *Public Library of Science Biology* 2:1529–1531.

- Narum, S. R., C. Contor, A. Talbot, and M. S. Powell. 2004. Genetic divergence of sympatric resident and anadromous forms of *Oncorhynchus mykiss* in the Walla Walla River, USA. *Journal of Fish Biology* 65:471–488.
- Nelson, J. S., T. C. Grande, and M. V. H. Wilson. 2016. *Fishes of the World*. Fifth edition. John Wiley and Sons, Hoboken, New Jersey, 707 pp.
- Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274.
- Nolan, K. A., J. R. Waldman, and I. Wirgin. 2003. Intraspecific and interspecific molecular differentiation of American shad and Alabama shad: a synthesis; pp. 295–302 in K. E. Limburg and J. R. Waldman (eds.), *Biodiversity, Status, and Conservation of the World's Shads*. American Fisheries Society, Symposium 35, Maryland, 370 pp.
- Østbye, K., T. Bernatchez, T. F. Næsje, K. – J. M. Himberg, and K. Hindar. 2005. Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology* 14:4371–4387.
- Palkovacs, E. P., K. B. Dion, D. M. Post, and A. Caccone. 2008. Independent evolutionary origins of landlocked alewife populations and rapid parallel evolution of phenotypic traits. *Molecular Ecology* 17:582–597.
- Pavlov, P. I. 1959. *Shads of genus Alosa in the North-western Part of Black Sea*. Kyiv: Akademia Nauk. UkrSSR, 252 pp.
- Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135.

- Popov, S. V., F. Rögl, A. Y. Rozanov, F. F. Steininger, I. G. Shcherba, and M. Kovac. 2004. Lithological-Paleogeographic maps of Paratethys-10 maps Late Eocene to Pliocene, 46 pp.
- Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* 25:1253–1256.
- Pravotorov, B. I. 2000. On population structure of the Black Sea – Azov Sea fish *Alosa caspia tanaica* (Grimm), (Clupeidae). *Hydrobiological Journal* 36:101–105.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <https://www.R-project.org/>
- Rambaut, A., Suchard, M. A., Xie, D. and Drummond, A. J. 2014. Tracer v1.6. Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Reist, J., E. Gyselman, J. Babaluk, J. Johnson, and R. Wissink. 1995. Evidence for two morphotypes of Arctic char (*Salvelinus alpinus* (L.)) from Lake Hazen, Ellesmere Island, Northwest Territories, Canada. *Nordic Journal of Freshwater Research* 71:396–410.
- Robinson, B. W. and K. J. Parsons. 2002. Changing times, spaces, and faces: tests and implications of adaptive morphological plasticity in the fishes of northern postglacial lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 59:1819–1833.
- Roesch, C., B. Lundsgaard-Hansen, P. Vonlanthen, A. Taverna, and O. Seehausen. 2013. Experimental evidence for trait utility of gill raker number in adaptive radiation of a north temperate fish. *Journal of Evolutionary Biology* 26:1578–1587.
- Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of entomology* 55:421–438.

- Stern, N., B. Rinkevich, and M. Goren. 2016. Integrative approach revises the frequently misidentified species of *Sardinella* (Clupeidae) of the Indo-West Pacific Ocean. *Journal of Fish Biology* 89:2282–2305.
- Svetovidov, A. N. 1945. On *Clupeonella delicatula* (Nordmann) of the Caspian and that of the Black and Azov seas. *Comptes Rendus de l'Academie des Sciences de l'URSS* 46:207–208.
- Svetovidov, A. N. 1952. *Fauna of USSR. Fishes. Vol. II (1). Clupeidae*. Moscow: USSR Academy of Sciences publishing, 331 pp.
- Svetovidov, A. N. 1964. Systematics of North American anadromous Clupeoid fishes of the genera *Alosa*, *Caspialosa* and *Pomolobus*. *Copeia* 196:118–130.
- Swain, D. P. and C. J. Foote. 1999. Stocks and chameleons: the use of phenotypic variation in stock identification. *Fisheries Research* 43:113–128.
- Turan, C. and N. Basusta. 2001. Comparison of morphometric characters of twaite shad (*Alosa fallax nilotica*, Geoffroy Saint-Hilaire, 1808) among three areas in Turkish seas. *Bulletin Français de la Pêche et de la Pisciculture* 362/363:1027–1035.
- Turan, C., D. Ergüden, M. Gürlek, C. Çevik, and F. Turan. 2015. Molecular systematic analysis of shad Species (*Alosa* spp.) from Turkish marine waters using mtDNA Genes. *Turkish Journal of Fisheries and Aquatic Sciences* 15:149–155.
- Velotta, J. P., S. D. McCormick, and E. T. Schultz. 2015. Trade-offs in osmoregulation and parallel shifts in molecular function follow ecological transitions to freshwater in the Alewife. *Evolution* 69:2676–2688.
- Venables, W. N. and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York, 498 pp.

- Walker, J. A. 1997. Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (Gasterosteidae) body shape. *Biological Journal of the Linnaean Society* 61:3–50.
- Whitehead, P. J. P. 1985. *FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part I. Chirocentridae, Clupeidae and Pristigasteridae.* *FAO Fisheries Synopsis* 125:1–303.
- Wund, M. A., S. Valena, S. Wood, and J. Baker. 2012. Ancestral plasticity and allometry in threespine stickleback fish reveal phenotypes associated with derived, freshwater ecotypes. *Biological Journal of the Linnaean Society* 105:573–583.

TABLE 4-1. Measurements and counts of *Alosa* spp. specimens from the Sea of Azov.

Specimen numbers follow Fig. 2. Measurements and abbreviations follow Fig.3. Measurements and abbreviations: BD_{max} – maximum body depth; ED – eye diameter; HL – head length; Io – interorbital width; pD – number of dorsal fin rays; PA – preanal length; pA – number of anal fin rays; PD – predorsal length; Pd – caudal peduncle depth; pP – number of pectoral fin rays; PV – preventral length; SL – standard length; TL – total length. All linear measurements are in millimeters. * indicates specimens used for the ddRAD sequencing.

Specimen Number	Species assignment	TL	SL	BD _{max}	Pd	HL	PD	PV	PA	ED	Io	pD	pA	pP	Gill rakers
AZ-I*	<i>A. immaculata</i>	162	139	37	12	37	68	71	106	7	7	16	18	14	48
AZ-II*	<i>A. immaculata</i>	205	175	39	13	45	84	86	125	8	9	17	19	16	45
AZ-III	<i>A. immaculata</i>	170	146	35	13	38	71	73	108	7	7	16	18	16	50
AZ-IV	<i>A. immaculata</i>	129	104	33	10	30	47	51	77	6	5	16	18	16	54
AZ-V	<i>A. immaculata</i>	182	156	39	13	40	73	78	114	7	8	17	20	16	48
AZ-VI	<i>A. immaculata</i>	169	144	38	13	36	67	71	103	7	6	18	18	16	54
AZ-VII	<i>A. immaculata</i>	197	166	45	15	44	71	79	114	8	7	16	20	16	50
AZ-VIII	<i>A. immaculata</i>	152	128	38	12	38	57	64	93	6	5	17	16	16	45
AZ-IX	<i>A. immaculata</i>	160	137	31	11	36	63	67	94	7	6	17	18	16	54
AZ-X	<i>A. immaculata</i>	185	156	36	13	40	75	82	113	7	7	17	20	16	54
AZ-02*	<i>A. caspia</i>	167	140	40	13	36	64	68	102	8	8	17	20	16	61
AZ-04*	<i>A. immaculata</i>	124	104	31	10	30	55	56	79	7	7	15	20	14	41
AZ-05*	<i>A. caspia</i>	184	154	37	13	40	73	75	109	7	7	16	17	16	62
AZ-06*	<i>A. immaculata</i>	143	117	31	12	32	55	58	90	7	5	17	20	16	42
AZ-13*	<i>A. caspia</i>	135	113	33	11	31	50	56	80	6	6	17	20	16	61
AZ-14	<i>A. immaculata</i>	118	96	28	9	28	47	50	72	6	5	17	21	16	46
AZ-16	<i>A. immaculata</i>	179	154	38	12	41	73	79	111	7	6	17	18	16	52
AZ-21*	<i>A. caspia</i>	185	155	38	13	41	72	80	111	9	7	16	20	16	60
AZ-27*	<i>A. immaculata</i>	197	172	46	15	43	72	78	117	8	7	17	20	16	44
AZ-29*	<i>A. immaculata</i>	195	167	43	14	43	78	82	115	8	7	17	20	16	46

TABLE 4-2. Morphological data recorded for the sampled individuals of *Alosa* spp. Body proportions are indicated as percentages (%). Abbreviations: Measurements and abbreviations follow Fig.3. Measurements and abbreviations: BDmax – maximum body depth; ED – eye diameter; HL – head length; Io – interorbital width; pD – number of dorsal fin rays; PA – preanal length; pA – number of anal fin rays; PD – predorsal length; Pd – caudal peduncle depth; pP – number of pectoral fin rays; PV – preventral length; SL – standard length; TL – total length.

	BD/SL	Pd/SL	HL/SL	PD/SL	PV/SL	PA/SL	ED/HL	Io/HL	Gill rakers
AZ-I	26.62	8.63	26.62	48.92	51.08	76.26	18.92	18.92	48.00
AZ-II	23.61	8.33	25.69	47.22	50.00	71.53	18.92	18.92	45.00
AZ-III	23.97	8.90	26.03	48.63	50.00	73.97	18.42	18.42	50.00
AZ-IV	31.73	9.62	28.85	45.19	49.04	74.04	20.00	16.67	54.00
AZ-V	25.00	8.33	25.64	46.79	50.00	73.08	17.50	20.00	48.00
AZ-IX	22.63	8.03	26.28	45.99	48.91	68.61	19.44	16.67	54.00
AZ-X	23.08	8.33	25.64	48.08	52.56	72.44	17.50	17.50	54.00
AZ-VI	26.39	9.03	25.00	46.53	49.31	71.53	19.44	16.67	54.00
AZ-VII	27.11	9.04	26.51	42.77	47.59	68.67	18.18	15.91	50.00
AZ-VIII	29.69	9.38	29.69	44.53	50.00	72.66	15.79	13.16	45.00
AZ-02	28.57	9.29	25.71	45.71	48.57	72.86	22.22	22.22	61.00
AZ-04	29.81	9.62	28.85	52.88	53.85	75.96	23.33	23.33	41.00
AZ-05	24.03	8.44	25.97	47.40	48.70	70.78	17.50	17.50	62.00
AZ-06	26.50	10.26	27.35	47.01	49.57	76.92	21.88	15.63	42.00
AZ-13	29.20	9.73	27.43	44.25	49.56	70.80	19.35	19.35	61.00
AZ-14	29.17	9.38	29.17	48.96	52.08	75.00	21.43	17.86	46.00
AZ-16	24.68	7.79	26.62	47.40	51.30	72.08	17.07	14.63	52.00
AZ-21	24.52	8.39	26.45	46.45	51.61	71.61	21.95	17.07	60.00
AZ-27	26.74	8.72	25.00	41.86	45.35	68.02	18.60	16.28	44.00
AZ-29	25.75	8.38	25.75	46.71	49.10	68.86	18.60	16.28	46.00

TABLE 4-3. Data sets used for the phylogenetic analyses.

Data set	Size			Accession numbers (ingroup)
	Base pairs	Taxa (ingroup)	Outgroup (Accession number)	
COI+ <i>cyt b</i>	1099	20	<i>Alosa alosa</i> (NC_009575)	This study
SNP	49972	20	n/a	This study
mtDNA supplemented:				<i>A. alosa</i> : KC500190, KC500192, AP009131, JX080177, JX080175; <i>A. fallax</i> : KJ768202, KJ204651, KJ204650, EU492080, EU492310; <i>A. sapidissima</i> : GU440215, KC015152, KC015144, EU552616, EF653234; <i>A. pseudoharengus</i> : KU564521, JN024722, KC015143, DQ419776, NC_009576; <i>A. aestivalis</i> : KC015129, KC015128, EU523898, EU552615, EF653229; <i>A. agone</i> : KJ552682, KJ552649; <i>A. macedonica</i> : KJ552493, KJ552563, KF631310, KF631309; <i>A. vistonica</i> : KJ552463; <i>A. immaculata</i> : KJ552592; <i>A. alabamae</i> : NC_028275, KJ158091, EF653230; <i>A. mediocris</i> : EF653233; <i>A. caspia</i> : DQ419770.
- COI	651	42	<i>Brevoortia</i>	
- <i>cyt b</i>	448	37	<i>tyrannus</i>	
- COI/ <i>cyt b</i>	1099	58	(NC_014266)	

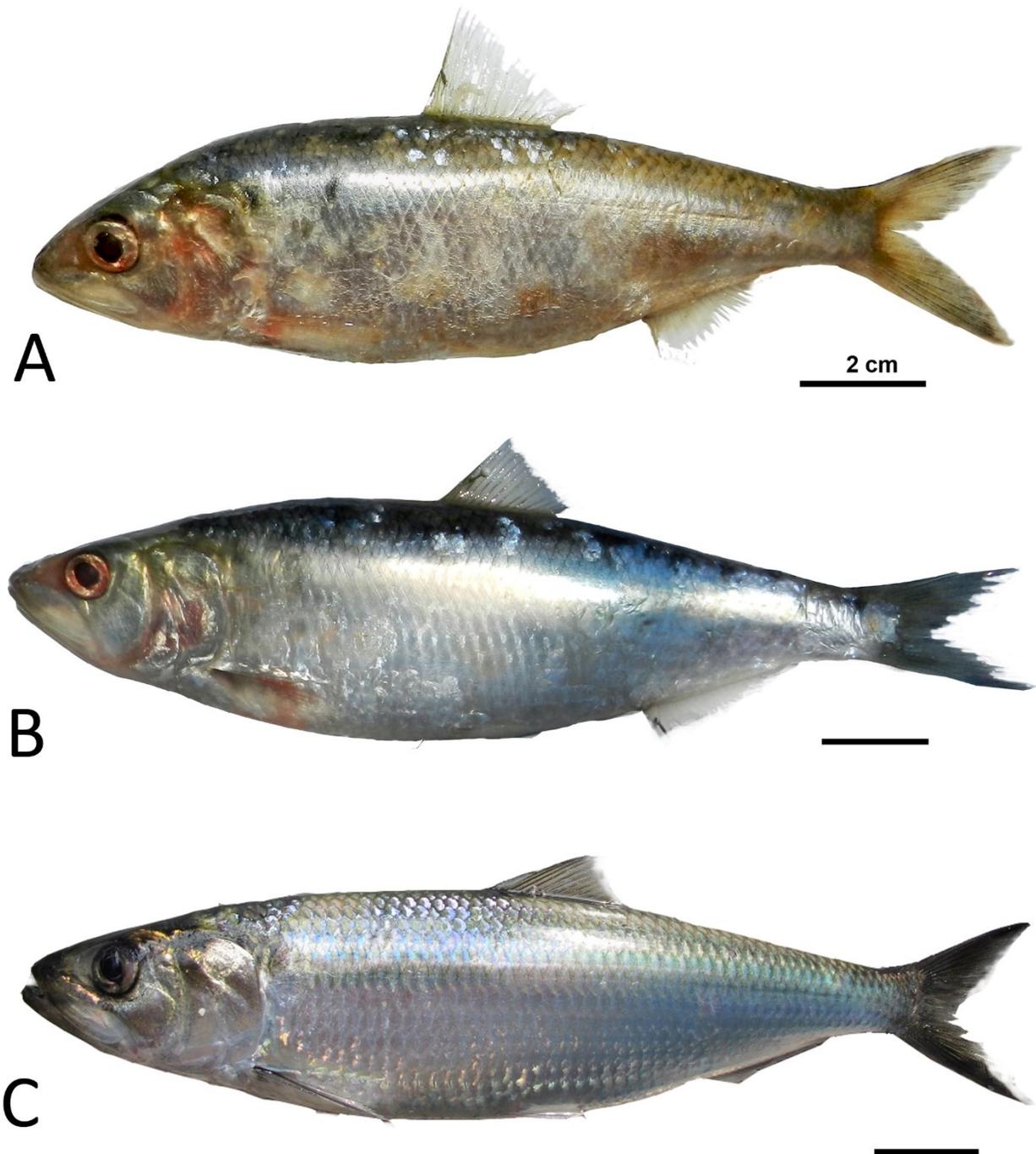


FIGURE 4-1. Species of *Alosa* in the Sea of Azov: A – *A. caspia* (AZ-02, this study); B – *A. immaculata* (AZ-IV, this study); C – *A. maeotica* (image from www.wikimedia.org). Scale bars = 2 cm.

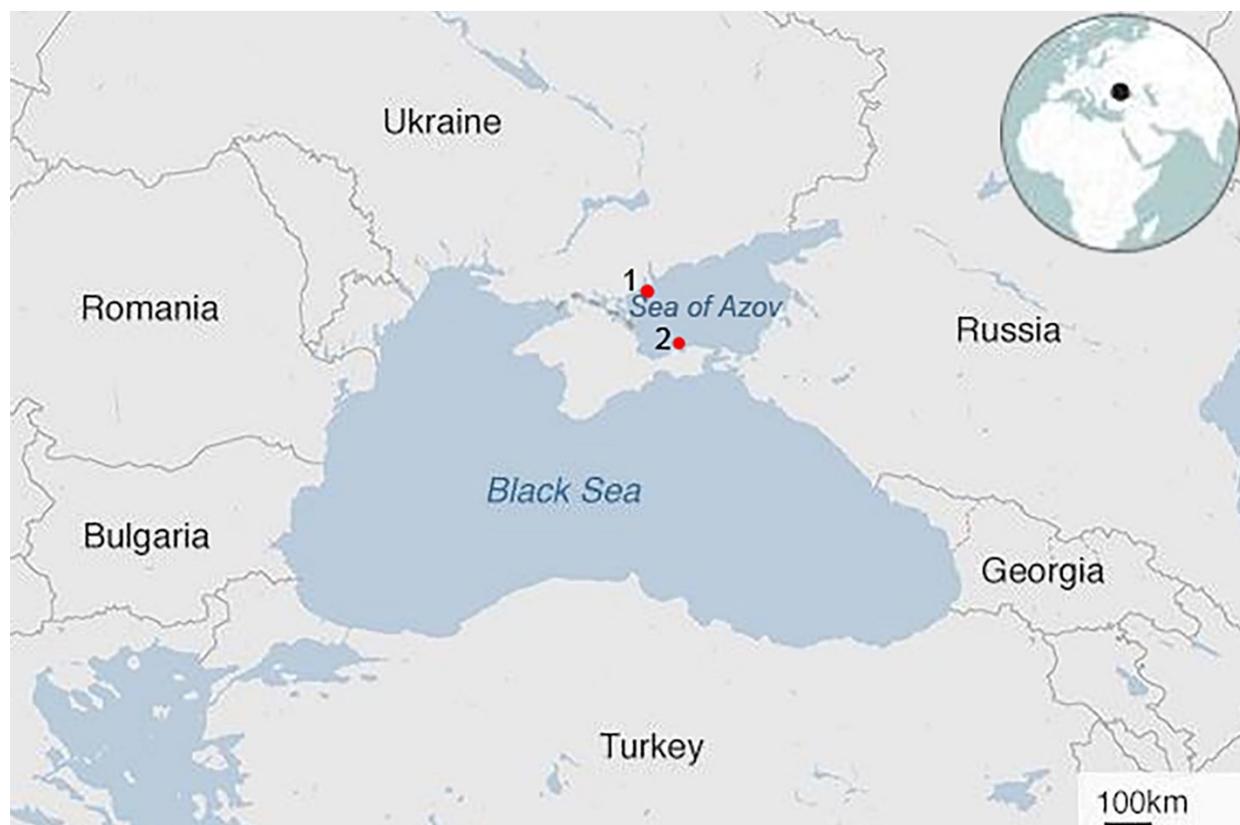


FIGURE 4-2. Map showing sampling localities in the Sea of Azov. 1 – North-western locality – specimens from this area are indicated in the study with the Roman numerals AZ_I – X; 2 – Southern locality – specimens from this area are indicated in the study with the Arabic numerals AZ_02 – 29.

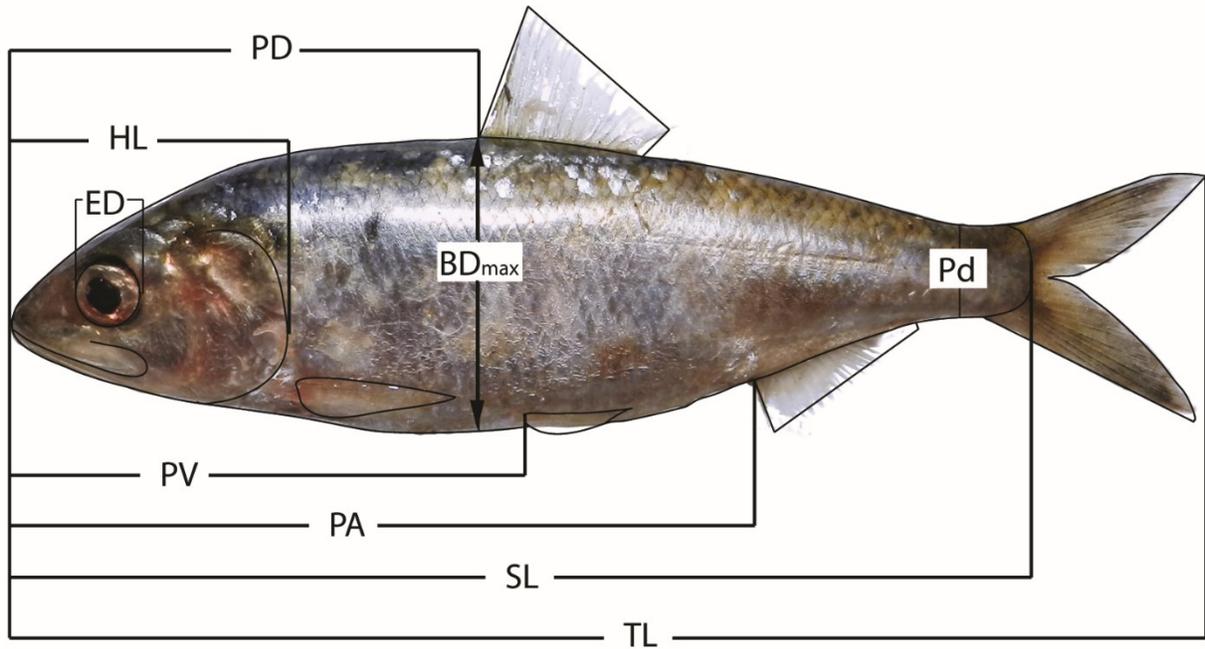


FIGURE 4-3. Morphometric measurements. Abbreviations: TL – total length; SL – standard length; BDmax – maximum body depth; Pd – caudal peduncle depth; HL – head length; PD – predorsal length; PV – preventral length; PA – preanal length; ED – eye diameter.

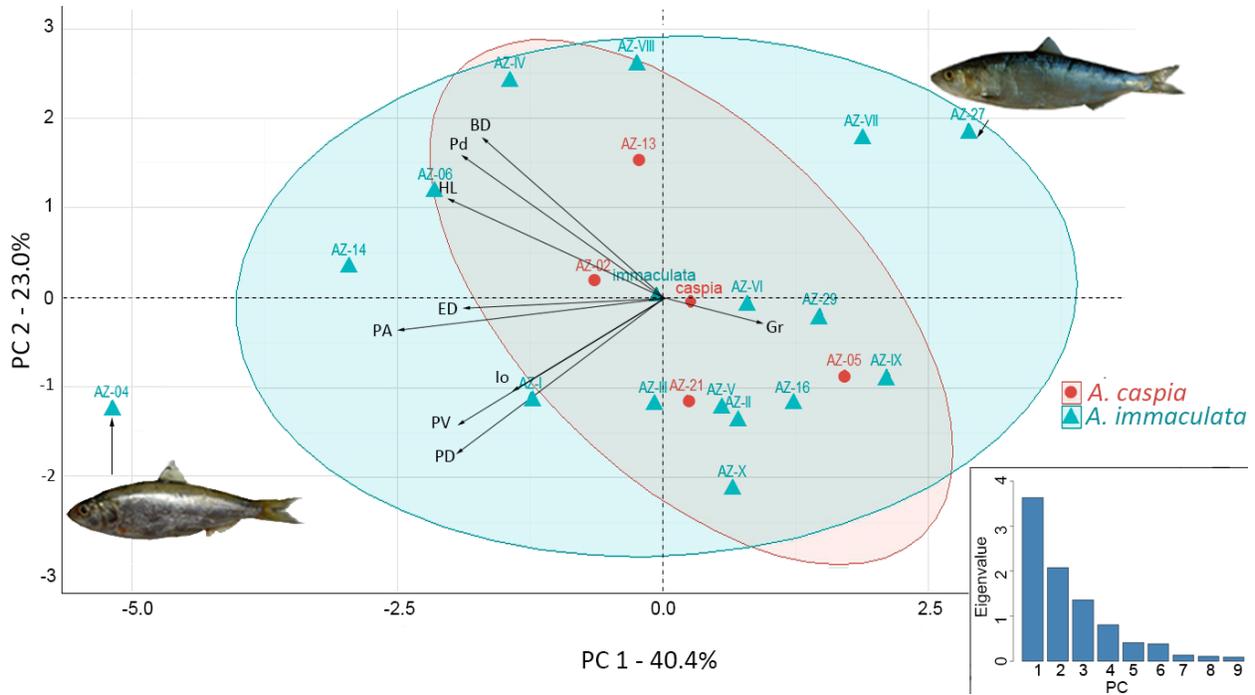


FIGURE 4-4. Principal component analysis biplot showing distribution of the 20 sampled individuals in the PC1 – PC2 morphospace along with the contributions of the morphometric ratios used in the analysis. Individuals are grouped by species according to their identification based on gill raker counts. Insert in the bottom right shows eigenvalues for each principal component.



FIGURE 4-5. Maximum likelihood topologies: A – concatenated COI/cyt b data set and B – SNPs data set. Bootstrap values are indicated above branches; SH-aLRT values are indicated below the branches. Asterisks indicate specimens with the gill raker count >59. Individuals (AZ_VI – VIII) sampled from the north-western part of the Sea of Azov but nested within the southern clade are highlighted in yellow.

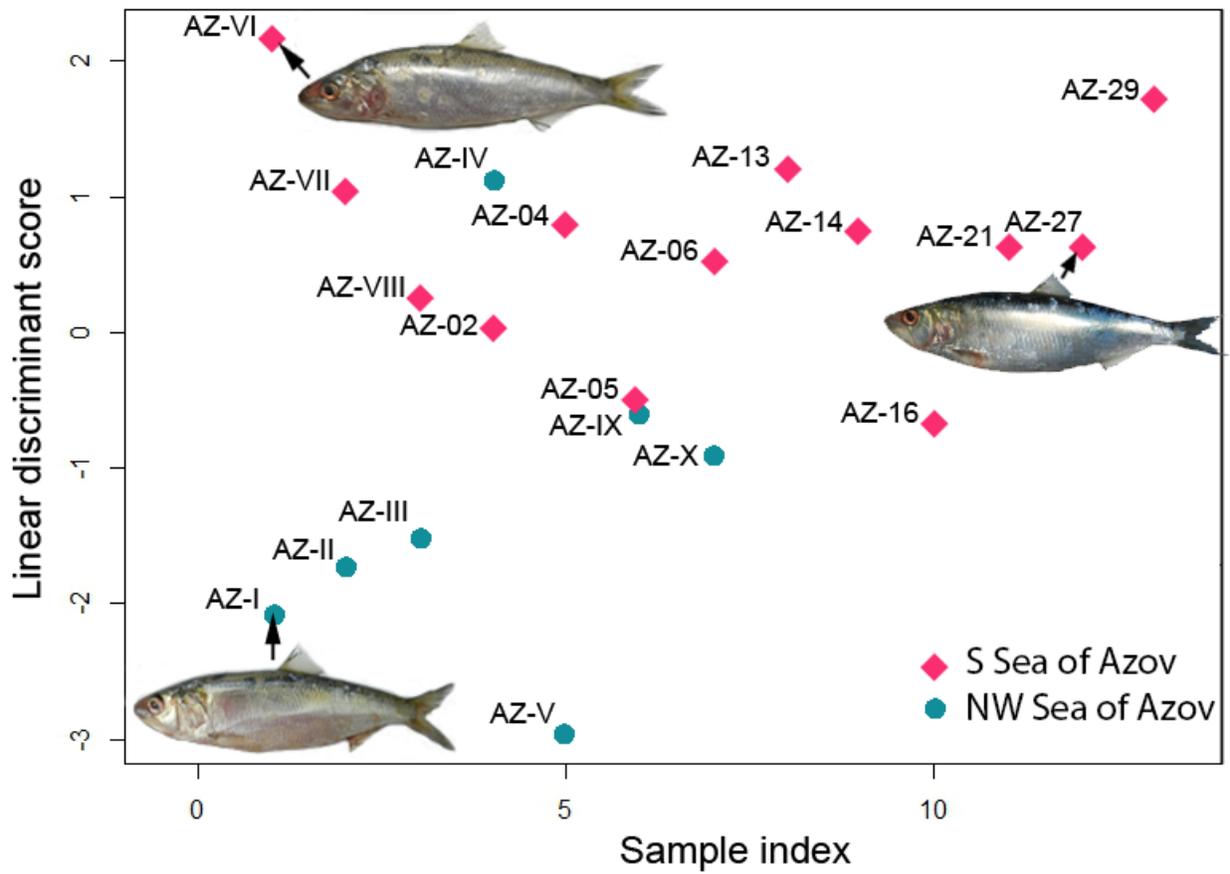


FIGURE 4-6. Linear discriminant analysis scatter plot showing separation of the two genetic clusters in the morphospace. Individuals were assigned to the two groups, south (S) and north-west (NW), based on the results of the phylogenetic analysis of the concatenated mtDNA data set.

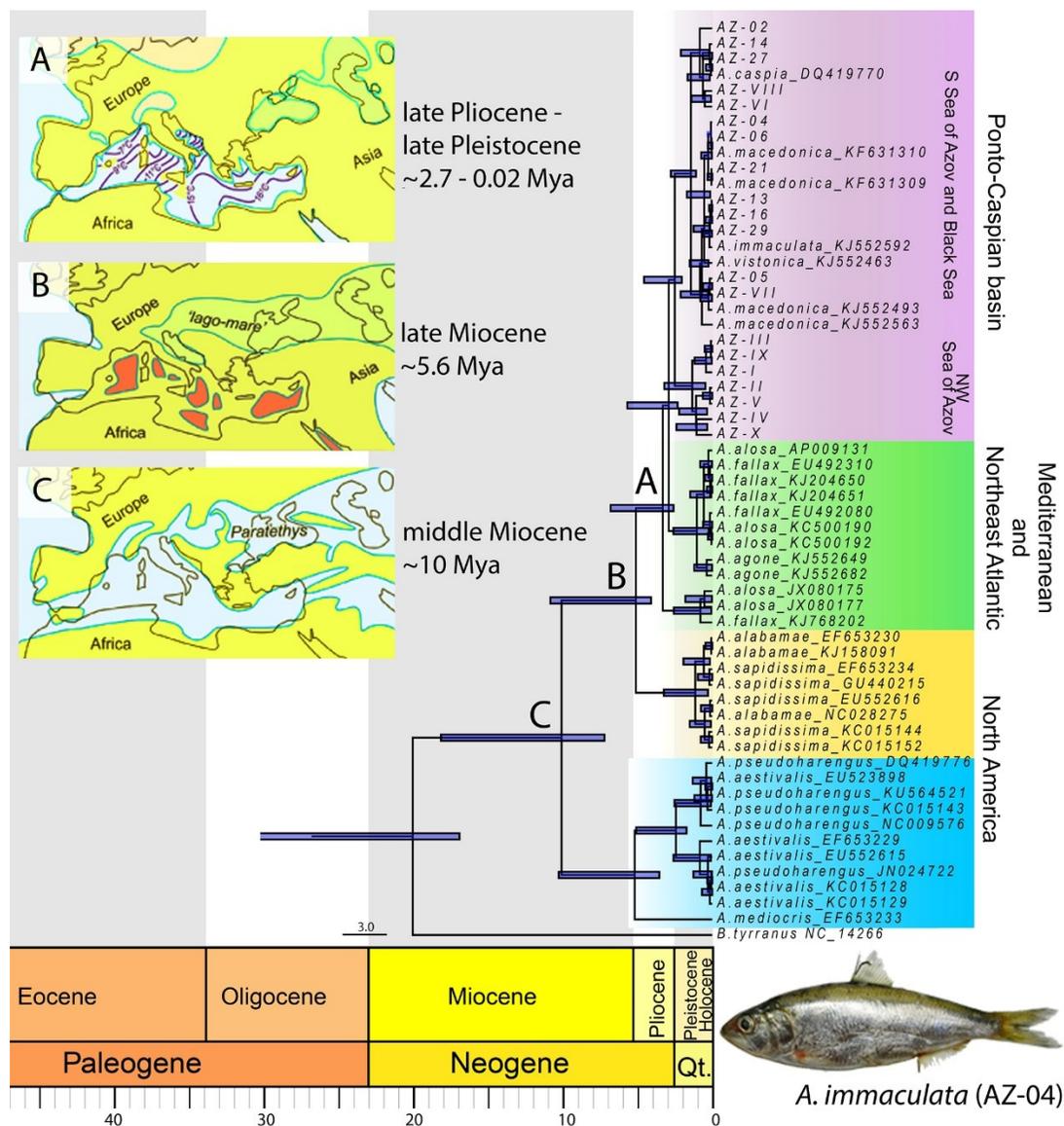


FIGURE 4-7. Phylogenetic chronogram based on the maximum clade credibility tree from a time-calibrated Bayesian analysis (BEAST v. 2.4.6) of the concatenated COI/cyt b data set for the genus *Alosa*. Timescale is in million years before present (mya). Bars at the nodes indicate 95% age credibility intervals. Figure inserts A – D show paleo-geographical reconstructions of the Mediterranean basin during the key divergence events labeled on the chronogram with the corresponding letters (figures adapted from Bianchi et al., 2014).



APPENDIX 4-1. Maximum likelihood topology of the COI data set for the genus *Alosa*. Support values are indicated at the nodes: SH-aLRT/Bootstrap.

CHAPTER 5

Morphological variation among the species of †*Armigatus* (Teleostei: Clupeomorpha: †Ellimmichthyiformes) and new material of †*Armigatus alticorpus* from the Late Cretaceous (Cenomanian) of Hakel, Lebanon

A version of this chapter has been accepted for publication as Vernygora, O.V. and A. M. Murray. 2020. Morphological variation among the species of †*Armigatus* (Teleostei: Clupeomorpha: †Ellimmichthyiformes) and new material of †*Armigatus alticorpus* from the Late Cretaceous (Cenomanian) of Hakel, Lebanon. *Cretaceous Research*. In Press.

5.1 Introduction

Members of the extinct clupeomorph genus †*Armigatus* were small marine fishes abundant in the Late Cretaceous Tethys Sea (e.g., Forey et al., 2003; Murray and Wilson, 2013; Vernygora and Murray, 2015; Murray et al., 2016). The oldest fossil record of the group is from the Cenomanian deposits of Levant that have yielded three of the five known species of the genus, the type species †*A. brevissimus* (Blainville, 1818) from Hakel and Hajula, and two species, †*A. namourensis* Forey et al., 2003 and †*A. alticorpus* Forey et al., 2003, described from Namoura. The other two species are known from Cenomanian – Turonian deposits of Morocco, †*A. oligodontatus* Vernygora and Murray, 2015, and Campanian deposits of Croatia, †*A. dalmaticus* Murray et al., 2016. Members of the genus are most readily distinguished from other †ellimmichthyiform clupeomorphs by the presence of an incomplete predorsal series of heart-shaped or oval scutes. Within the genus, however, species delimitation is primarily based on morphometric and meristic traits which in some cases have overlapping ranges making species

identification a problematic task. This issue is particularly pertinent to distinguishing between spatially and temporally co-occurring species, †*A. brevissimus*, †*A. alticorpus*, and †*A. namourensis*. In a recent review of relationships of the species of †*Armigatus*, Murray et al. (2016) summarized key distinguishing characteristics for each of the five species which for the most part included number of the dorsal and anal fin rays, numbers of vertebrae, ribs, and predorsal scutes, as well as relative body depth. Because of the partial preservation of some specimens, not all of these meristic and morphometric traits are known for all of the species making it impossible to directly compare all species in the genus.

Of the five species, †*A. alticorpus* is the least studied taxon. This species was originally described from the Cenomanian deposits of Namoura, Lebanon (Forey et al., 2003). The very brief original description was limited to a general comparison with †*A. namourensis* and †*A. brevissimus*, and did not provide detailed information on the morphology of this species. As a result, the most recent cladistic analyses of the group (Murray et al., 2016) excluded †*A. alticorpus* because of the lack of sufficient information for the taxon.

Here, I report new material of †*Armigatus* from Hakel, Lebanon collected in 1998 and housed in the Royal Tyrrell Museum of Palaeontology (TMP, Drumheller, Alberta, Canada). Hakel is one of the most famous Late Cretaceous (Cenomanian) fossiliferous localities in Lebanon along with Namoura and Hajula. These localities are characterized by presence of ‘fish-beds’ preserving multiple individuals in a single layer. These mass mortality beds presumably represent anoxic events caused by ‘blooms’ of planktonic dinoflagellates triggered by coastal upwellings (Hemleben, 1977; Schram et al., 1999). Geology and conditions of deposition at the three localities have been described in detail elsewhere (e. g., Dubertret and Vautrin, 1937; Dubertret, 1959, 1966; Hückel, 1970, 1974; Saint Marc, 1974; Forey et al. 2003). As was noted

by Forey et al. (2003), ichthyofaunal composition of the three Lebanese localities is much more similar between Hakel and Hajula than between either of those and Namoura. According to their estimates, less than a quarter of species described from Namoura are shared with the Hakel ichthyofauna (Forey et al., 2003). Among the species that were considered unique to Namoura are several clupeomorph taxa including †*Scombroclupea diminuta*, †*Sorbinichthys elusivo*, †*Triplomystus noorae*, †*Triplomystus oligoscutatus*, †*Armigatus namourensis*, and †*Armigatus alticorpus*.

Specimens referred to in this study are preserved on a large limestone slab, TMP 1998.65.11, which preserves multiple clupeomorph fishes including specimens of two species of †*Armigatus*, †*A. namourensis* and †*A. alticorpus*. Although in many cases individual fish are obscured by overlying specimens, it is possible to identify anatomical details of several fish. Overall, preservation of the †*A. alticorpus* specimens provides necessary information for addition of this species to a phylogenetic analysis. Results of phylogenetic reassessment of †*A. alticorpus* provide an updated and more complete hypothesis of interrelationships within the genus, as well as shed light on historical biogeography of the group and ichthyofaunal connections between the Late Cretaceous Lebanese localities.

5.2 Materials and methods

The material described here is housed in the Royal Tyrrell Museum of Palaeontology (Alberta, Canada) and is catalogued under TMP 1998.65.11. This material consists of a large limestone slab (TMP 1998.65.11) which preserves multiple clupeomorph fishes including at least four identifiable specimens of †*A. alticorpus*. Photographs of the specimens were taken using a

Canon EOS Rebel and an Olympus OM-D E-M10II camera as well as a Dino-Lite Pro Digital Microscope AM-413ZTAS.

Phylogenetic analysis – The phylogenetic analysis is based on an existing character list used to analyse relationships within †Ellimmichthyiformes in earlier studies by Zaragüeta-Bagils (2004), Alvarado-Ortega et al. (2008), Murray and Wilson (2013), with modifications discussed in Vernygora and Murray (2015), Murray et al. (2016), and Marramà and Carnevale (2017) (Appendix 5-1). The final character matrix included 37 ingroup taxa and five outgroup taxa, the ostariophysan *Chanos chanos* and four clupeiform species *Denticeps clupeoides*, †*Palaeodenticeps tanganyikae*, *Chirocentrus dorab*, and *Odaxothrissa vittata*. I performed a cladistic analysis in TNT v 1.5 (Goloboff et al., 2008) using an heuristic search method with 1000 starting replicates and 100 trees saved per replication, random taxon addition, and tree bisection and reconnection (TBR) swapping algorithm. Statistical branch support values, bootstrap and Bremer support, were also calculated in TNT v 1.5. Mesquite version 3.61 (Maddison and Maddison, 2015) was used to calculate consistency (CI) and retention (RI) indices for the 50% majority-rule consensus of the most parsimonious trees (MPTs).

Comparative Material – *Alosa sapidissima* CMN Z 366; †*Armigatus alticarpus* NHMUK P.63134 (holotype), TMP 1998.65.11; †*Armigatus brevissimus* UALVP 5087, 17620, 47258; †*Armigatus dalmaticus* CNHM 9423 (holotype), 9259, 9277, 9287; †*Armigatus oligodentatus* UALVP 51679 (holotype), 47146, 47155, 51602, 51622, 51623, 51680, 51681; *Chanos chanos* UAMZ F8463, F8550; *Chirocentrus dorab* CMN 77-0025; *Chirocentrus nudus* ROM R8500; *Denticeps clupeoides* (syntype) MNHN IC-1960-0391; †*Diplomystus birdi* CMN FV 30564; †*Diplomystus dentatus* UALVP 17830, TMP 86.224.81, 86.224.85, 86.224.89; †*Diplomystus dubertreti* MNHN F SHA2053 (holotype); †*Foreyclupea loonensis* UALVP 17535 (holotype);

†*Horseshoeichthys armaserratus* TMP 2001.045.0093 (holotype); †*Palaeodenticeps tanganikae* WM 352/92, 301/96, 432/96, 128/96, 100/96; †*Sorbinichthys africanus* UALVP 51640 (holotype), 47186, 51641; †*Thorectichthys marocensis* UALVP 47178 (holotype), 47134, 51647, 51649; †*Thorectichthys rhadinus* UALVP 51653 (holotype), 51664, 51715; †*Tycheoichthys dunveganensis* CMN 52730 (holotype). Information for other species was taken from the literature.

Anatomical abbreviations –**ach**, anterior ceratohyal; **ber.f**, beryciform foramen; **br**, branchiostegal ray; **cl**, cleithrum; **cor**, coracoid; **d**, dentary; **pcd.s**, precaudal scute; **ect**, ectopterygoid; **ep**, epural (numbered); **f.r**, fin rays; **fr**, frontal; **hy**, hypural; **io**, infraorbital; **io. c**, infraorbital sensory canal; **let**, lateral ethmoid; **mes**, mesethmoid; **mx**, maxilla; **n.s**, neural spine; **op**, opercle; **ors**, orbitosphenoid; **pa**, parietal; **pas**, parasphenoid; **ph**, parhypural; **pmx**, premaxilla; **pop**, preopercle; **r**, rib; **pr**, procurrent rays; **ptt**, posttemporal; **pu**, preural centrum; **sca**, scapula; **scl**, supracleithrum; **smx**, supramaxilla; **sn**, supraneural (numbered); **sph**, sphenotic; **u**, ural centrum; **un**, uroneural; **vc**, vertebral centrum.

Measurement abbreviations - **BD**, body depth (measured as the greatest vertical distance of the body); **HD**, head depth (vertical distance taken at the greatest height of the supraoccipital to the ventral border of head); **HL**, head length (measured as the horizontal distance between the tip of the premaxilla and the posterior limit of the opercle); **PA**, preanal length (taken from the anteriormost point of the head to the origin of anal fin); **PD**, predorsal length (distance between anterior end of the head to the origin of dorsal fin); **PV**, prepelvic length (taken from the anterior end of head and the origin of pelvic fin); **SL**, standard length (distance between anterior end of head to the posteriormost end of hypural); **TL**, total length (distance between anterior end of head to the posteriormost fin ray of caudal fin).

5.3 Systematic paleontology

Subdivision Teleostei MÜLLER, 1845

Cohort Clupeocephala PATTERSON & ROSEN, 1977

Superorder Clupeomorpha GREENWOOD, ROSEN, WEITZMAN & MYERS, 1966

Order †Ellimmichthyiformes GRANDE, 1985

Family †Armigatidae MURRAY AND WILSON, 2013

†*Armigatus* GRANDE, 1982

Included species: †*A. alticorpus* Forey et al., 2003, †*A. brevissimus* (de Blainville, 1818), †*A. dalmaticus* Murray et al., 2016, †*A. namourensis* Forey et al., 2003, †*A. oligodentatus* Vernygora and Murray, 2015.

†*Armigatus alticorpus* Forey et al., 2003

Emended diagnosis: member of the genus †*Armigatus* based on the presence of incomplete series of heart-shaped predorsal scutes. The species is distinguished from other members of the genus by having a comparably greater body depth, 46 – 52% of standard length, 15 – 17 dorsal fin rays (16 or more in †*A. brevissimus* and †*A. namourensis*, 11 – 14 in †*A. oligodentatus* and †*A. dalmaticus*); unornamented bones of the skull roof (compared to the ornamented skull roof of †*A. brevissimus* and †*A. namourensis*), 27 anal pterygiophores (16 or fewer in †*A. oligodentatus* and †*A. dalmaticus* but 24 – 27 in †*A. namourensis* and 27 or more in †*A. brevissimus*), 33 – 35 preural centra (31 – 32 in †*A. brevissimus*, 36 – 38 in †*A. namourensis*,

35 – 37 in †*A. oligodentatus*, but 32 – 35 in †*A. dalmaticus*), and 7 – 9 predorsal scutes (17 – 19 in †*A. namourensis*, 10 – 11 in †*A. dalmaticus*, but 9 in †*A. brevissimus* and †*A. oligodentatus*).

Holotype: NHMUK P.63134 (Fig. 5-1; Forey et al., 2003:fig. 55).

Referred material: TMP 1998.65.11 (4 specimens; Figs. 5-2 and 5-3)

Type locality and age: Namoura, Lebanon; Middle Cenomanian.

5.3.1 Description

General Body Form – This is a small sized clupeomorph fish reaching a standard length (SL) of 60 mm (Table 5-1). Body is deep and relatively short with a smooth gently arched dorsal outline not forming a steep incline between the occiput to the dorsal fin origin. The greatest body depth varies between 46% and 52% of SL. All referred specimens are preserved in lateral view, indicating that the fish most likely had a laterally compressed body, typical of clupeomorphs. The head is triangular in lateral view; it is as deep (or slightly deeper) as it is long. The length from the tip of the snout to the posterior edge of the opercle reaches 37.5% of SL. Dorsal fin originates close to the midpoint of the body with the predorsal length being 47.9 – 51.1% of SL. The pelvic fin origin is below the origin of the dorsal fin. The caudal fin is deeply forked.

Skull roof – The frontals constitute majority of the skull roof (Fig. 5-1). The frontal bones are long, tapering anteriorly to articulate with the mesethmoid and expanding posteriorly to meet the parietals. The suture between the two frontals is straight and continuous; the anterior fontanelle is absent. In the holotype (NHMUK P.63134), there is a partially preserved bone-enclosed supraorbital sensory canal in the posterior half of the frontal (Fig. 5-1). The parietals are relatively small and roughly square; they contact each other in the midline excluding the

supraoccipital bone from the contact with the frontals (medioparietal skull condition). Details of the supratemporal sensory canal cannot be determined in the specimens. The supraoccipital is partially preserved in the holotype and specimens A and D; it is a small bone with a low median crest. Bones of the skull roof are smooth, without ornamentation.

Orbital region - Impressions of the lateral ethmoid, at least three infraorbitals, and the dermosphenotic are preserved in specimens A, C and D (TMP 1998.65.11), but details of the individual bones cannot be determined. Remains of a laminar bone in the upper part of the orbit are interpreted as a possible orbitosphenoid. The parasphenoid crosses the orbit slightly below the midpoint; it is narrow in lateral view and lacks the basipterygoid process. The details of the infraorbital sensory canal cannot be determined.

Jaws – The upper jaw comprises the premaxilla, maxilla, and two supramaxillae; it is best preserved in specimen A (Fig. 5-4). The premaxilla has a typical clupeomorph comma shape with an expanded anterior portion. The maxilla is long and gently curved. Anteriorly, the maxilla ends in a robust articular head that contacts the premaxilla. There is a strengthening ridge that runs along the maxillary blade close to the ventral margin of the bone. The supramaxillary bones are tear drop-shaped and positioned over the posterior end of the maxilla. The posterior supramaxilla has an oval body with a long spine-like anterior process which extends over the smaller anterior supramaxilla. The dentary is narrow at the symphysis and deepens posteriorly. The anguloarticular is crushed in all specimens and nothing can be inferred about its appearance. All the jaw bones appear to be edentulous, however, this might be an artifact of preservation and it is possible that some jaw elements could have minute teeth.

Suspensorium and Branchial Region – The suspensorium is poorly preserved in all specimens. The hyomandibula, as well as the pterygoid series, are crushed or missing. Specimen

A (TMP 1998.65.11) preserves an impression and partial remains of the head of the hyomandibula which appears wide and forms a single condyle for articulation with the skull. The pterygoid series are partially preserved in the holotype and specimens A, and D. The metapterygoid is a large thin bone; its shape cannot be determined. The endopterygoid is partially preserved and appears oval in shape. Details of the quadrate, ectopterygoid, and palatine are not clear in any of the specimens. Teeth are not preserved on any of the pterygoid series of bones or the palatine.

The anterior ceratohyal is preserved in the holotype and specimen A (TMP 1998.65.11). It is an elongate hourglass shaped bone with a large 'beryciform' foramen in the centre (Fig. 4), as is typical of †ellimmichthyiforms and some clupeiforms. There are six branchiostegal rays visible in the holotype.

Opercular Series – The opercle is relatively narrow; it is approximately three times as tall as it is wide. The posterior margin of the bone curves ventrally to meet with its anterior edge at an acute angle. Anteriorly, the opercle is thickened and has a straight margin.

The preopercle is best preserved in specimen D (TMP 1998.65.11); it is L-shaped with well-developed, narrow dorsal and ventral limbs that meet almost at a right angle. The dorsal limb is approximately twice as long as the ventral limb. A thick ridge, partially preserved close to the anterior margin of the preopercle, is interpreted as the remains of the bone-enclosed sensory canal, but no canal pore openings can be identified. Details of the subopercle and interopercle are not clear in any of the specimens. The surface of all bones in the opercular series is smooth, without ridges or striations.

Paired Fins and Girdles – The posttemporal is a thin bone with a long and slender dorsal limb; the ventral limb of the bone appears to be wider and shorter than the dorsal limb. Details of

the supracleithrum are not clear in any specimen. The cleithrum is S-shaped, forming two rounded angles; it is slender and tapers ventrally. The coracoid is best seen in the holotype and specimen A (TMP 1998.65.11). It is an expanded thin bone that articulates posterodorsally with the scapula. There are 10 or 12 pectoral fin rays. The pectoral fin is short with its distal end not reaching the level of the origin of the dorsal fin.

The pelvic girdle is poorly preserved with only fragments of the fin rays visible in the holotype and specimens A, and D (TMP 1998.65.11). The pelvic fin inserts only slightly posterior to the level of the dorsal fin origin. The pelvic girdle is obscured by the abdominal scutes and details are not clear.

Median and Caudal Fins - The dorsal fin is triangular with a gradual decrease in height towards its posterior end. It originates anterior to the midpoint of the body above the 12–15th abdominal centra. The dorsal fin comprises approximately 15–17 fin rays supported by 15–17 pterygiophores. The two anterior-most dorsal fin rays are not segmented and unbranched. The anal fin is poorly preserved in all specimens with only pterygiophores visible. The fin appears to be long, extending under almost the entire caudal region of the trunk. There are 27 long anal fin pterygiophores that reach the distal ends of the haemal spines.

The caudal fin is strongly forked, similar to that of other species of *Armigatus* (Fig. 5-5). There are 19 principal and approximately seven dorsal and ventral procurrent caudal fin rays with a dorsal caudal scute in front of the procurrent fin rays. The caudal fin rays are supported by six hypurals. The first hypural is autogenous; it is broad distally and tapers proximally, reaching the first ural centrum. The second hypural is fused with the first ural centrum as is normal in clupeomorphs. There is no diastema between the second and third hypurals. The first ural centrum is roughly the same size as the first preural centrum. The parhypural is fused with the

first preural centrum. There are three slender epurals filling the space between the neural spine of the second preural centrum and the first uroneural. The neural spine of the first preural centrum is much shorter than that of the second preural centrum. There are two thin uroneurals; the first is long, reaching anteriorly to the second preural centrum. The posterior end of the second uroneural reaches the posterior end of the first uroneural. The presence of a neural arch on the first preural centrum cannot be confirmed.

Supraneurals and Scute Series – There are five or six supraneural bones preserved in the specimens (six in the holotype). The supraneurals are narrow, without the laminar expansions at the distal ends, which are present in larger specimens of †*Diplomystus*; the proximal ends of the supraneurals are gently curved anteriorly.

There are two series of scutes, the predorsal and the abdominal. The predorsal series extends from close to the origin of the dorsal fin anteriorly to about half-way to the occiput, leaving an unscuted gap behind the head, as it is characteristic of †*Armigatus*. There are seven or eight roughly heart-shaped scutes in this series, all of which have a similar size and smooth unornamented surfaces (Fig. 5-6).

The abdominal series of scutes consists of a total of 21–24 scutes and extends from the isthmus to the anal fin with 8–10 of the series located posterior to the pelvic fin. Each scute has a pair of prominent spine-like ascending lateral wings and a short posteriorly-directed median spine. The ascending lateral wings of the abdominal scutes are relatively short, not reaching half the distance between the ventral margin of the body and the vertebral column. These lateral wings do not meet one another at their dorsal extent, but expand ventrally toward the median keel and contact the neighboring scutes ventrally.

Vertebral Column - There are 35–37 vertebrae including the two ural centra. Of these, 16–18 are abdominal. The anteriormost 11–12 centra are associated with paired neural spines. Posteriorly, the right and left halves of the neural spines are fused in the midline. In the caudal region, each centrum is ornamented with two lateral ridges and three pits, a characteristic pattern observed in the †ellimmichthyiforms. There are about 13–15 pairs of ribs. Ribs articulate with well-developed parapophyses which significantly decrease in size in the anterior portion of the vertebral column. Two series of intermuscular bones are preserved in the specimens, representing the epineural and epipleural series. Both series continue into the caudal region.

5.4 Phylogenetic results

Parsimony analysis recovered 738 most parsimonious trees (MPTs) of 264 steps. A majority rule consensus tree of the MPTs has a length of 271 steps and consistency (C.I.) and retention (R.I.) indices equal 0.3358 and 0.6457, respectively. In the consensus tree (Fig.5-7), extant clupeiform outgroup species and an extinct taxon, †*Palaeodenticeps tanganikae*, form a monophyletic group which, however, does not include two extinct clupeiforms, †*Pseudoellimma gallae* and †*Santanaclupea silvasantosi*, that were recovered in a sister group relationship to †Ellimmichthyiformes including a problematic taxon, †*Ornategulum sardinoides*.

Within †Ellimmichthyiformes, four distinct families were recovered, †Scutatuspinosidae, †Sorbinichthyidae, †Armigatidae, and †Paraclupeidae (Fig.5-7). Composition of each family is overall consistent with the phylogenetic results of Murray and Wilson (2013) and Vernygora et al. (2016) who found a monophyletic family †Armigatidae (†*Armigatus* + †*Diplomystus*), a monotypic family †Sorbinichthyidae, †Scutatuspinosidae (or subfamily †Scutatuspinosinae) comprising the Early Cretaceous clupeomorphs †*Scutatuspinosus itapagipensis*, †*Ranulfoichthys*

dorsonudum, and †*Foreyclupea loonensis*, and a large family †Paraclupeidae which includes a diversity of more derived †ellimmichthyiform taxa.

Addition of new recently described or revised clupeomorph taxa revealed new patterns and interrelationships within well-established family groups mentioned above. Contrary to the results of the most recent phylogenetic analyses of Marramà and Carnevale (2017) and Marramà et al. (2019), I recovered †*Gasteroclupea branisai* (Maastrichtian – Danian of Bolivia) as a sister taxon to †*Tychemichthys dunveganensis* (Cenomanian of Alberta, Canada) within †Paraclupeidae, instead of being a sister taxon to †*Sorbinichthys*. The new Eocene †paraclupeid species, †*Eoellimmichthys superstes*, is placed more basally within †Paraclupeidae, being distantly related to †*Ellimmichthys* species. Surprisingly, the Early Cretaceous species †*Codoichthys carnavalii* from Brazil is nested within †Armigatidae, as a sister taxon to the Late Cretaceous genus †*Armigatus*, which so far is known only from the Eastern Tethys region. Two genera, †*Diplomystus* and †*Ellimmichthys*, are not recovered as monophyletic, as has been found in previous studies which have indicated that †*D. solignaci* and †*E. maceioensis* should be removed from their respective genera (Zaragüeta-Bagils, 2004; Alvarado-Ortega, 2008; Murray and Wilson, 2013; Vernygora and Murray, 2015). Interrelationships among the species of †*Armigatus* recovered in my analysis indicate close affinities of †*A. alticorpus* to the younger species, †*A. oligodentatus* and †*A. dalmaticus*. The three species together (†*A. alticorpus* (†*A. oligodentatus* + †*A. dalmaticus*)) belong to a lineage diverging from their sister-clade (†*A. brevisimus* + †*A. namourensis*).

Results of the bootstrap and Bremer support analyses did not show strong support for the clades recovered (Fig.5-7). A list of characters supporting recovered clades is provided in Appendix 5-2.

5.5 Discussion

5.5.1 Phylogenetic affinities of †*Armigatus alticorpus*

In the original description, †*Armigatus alticorpus* was characterized as being “intermediate between †*A. brevissimus* and †*A. namourensis*” (Forey et al., 2003:p. 282). This interpretation of the species is supported by meristic characters which are similar to the counts observed in the latter two species. Close resemblance in meristic values is common among congeneric species and is considered one of the major challenges in species delimitation in many groups of fishes (e.g., Alexandrino et al., 2006; Bagley et al., 2015; Brian et al., 2011; Eyo, 2003; McDowal, 2001, 2003; Ramler et al., 2017; Robinson and Parsons, 2002; Smith et al., 2009); however, without a detailed description, it is impossible to accurately assess phylogenetic relationships of species and assume its affinities. In the present study, I provide a greatly expanded description of the species based on the additional specimens which provide enough information to include †*A. alticorpus* in phylogenetic analysis and, for the first time, assess relationships among all five species of †*Armigatus*. Unexpectedly, results of my morphological examination and cladistic analysis recover close affinity of †*A. alticorpus* to †*A. oligodentatus* and †*A. dalmaticus*, two younger taxa described from the Cenomanian – Turonian of Morocco and Campanian of Croatia, respectively, rather than to †*A. namourensis* or †*A. brevissimus*, both of which are found in localities of similar age and geographic location as †*A. alticorpus*. Members of the clade (†*A. alticorpus* (†*A. oligodentatus* + †*A. dalmaticus*)) are distinguished from †*A. brevissimus* and †*A. namourensis* by a lack of the skull roof ornamentation (character 4:0). In the latter two species, the parietals and posterior part of the frontals are ornamented with prominent sinuous ridges and grooves (Forey et al., 2003) – a condition not observed in the other three species of †*Armigatus*.

Notably, there are no automorphies characteristic of †*A. alticorpus*; this species differs from its congeners by a unique combination of non-unique traits, as indicated by the diagnosis above. The most obvious distinguishing feature of the species – a high body depth to standard length ratio – is a highly variable morphometric trait that may change with growth (Khalloufi et al., 2010), state of the extant individual (gravid female) or be an artifact of preservation. Therefore, relative body depth as a character is often excluded from evolutionary analyses at higher taxonomic levels to reduce overall levels of homoplasy in data sets.

5.5.2 Updated phylogeny of †*Ellimmichthyiformes*

Results of my updated phylogenetic analysis support a monophyletic †*Ellimmichthyiformes* (Fig. 7). Fossil and extant clupeiform taxa, however, form a paraphyletic assemblage which could be a result of the fact that the character list used in the analysis is primarily constructed to resolve relationships among the members of the order †*Ellimmichthyiformes* and omits many of the clupeiform specific characters. Monophyly of †*Ellimmichthyiformes*, including †*Ornategulum* as an early diverging taxon, is supported by the medioparietal condition of the skull roof (2:0), lack of *recessus lateralis* (7:0), presence of the “basipterygoid” process (9:1) and three epurals (38:0). The first two features have been previously regarded as diagnostic characters of †*Ellimmichthyiformes* (Grande, 1985; Chang and Maisey, 2003) and justify inclusion of the problematic taxon, †*Ornategulum sardinoides* within the order. This placement of †*Ornategulum* is similar to the most recent results of phylogenetic analyses by Marramà et al. (2019) and Marramà and Carnevale (2017) who also found †*Ornategulum* as an early diverging †*ellimmichthyiform* taxon. Previously, however, De Figueiredo and Ribeiro (2017) recovered this genus as a sister taxon to Clupeomorpha.

The next clade within †Ellimmichthyiformes corresponds to the family †Scutatuspinosidae *sensu* Vernygora et al. (2016) and comprises three Early Cretaceous species, †*Ranulfoichthys dorsonudum*, †*Scutatuspinosus itapagipensis*, and †*Foreyclupea loonensis*. †*Scutatuspinosus* is the type species of the subfamily †Scutatuspinosinae Silva Santos and Silva Corr ea, 1985. These torpedo-bodied clupeomorphs are distinguished from other †ellimmichthyiforms by a set of homoplastic features – presence of a foramen in the posterior ceratohyal (14:1; this condition is not known in †*Foreyclupea loonensis*) and subrectangular dorsal process of the posttemporal (22:1). Previous studies (De Figueiredo and Ribeiro, 2017; Marram  and Carnevale, 2017; Marram  et al., 2019) recovered †*Scutatuspinosus* as a member of †Paraclupeidae, a more derived †ellimmichthyiform group; those analyses, however, did not include the other two species of †Scutatuspinosidae which makes it impossible to directly compare their phylogenetic hypotheses to my results.

The †Scutatuspinosidae, as recovered in my study, are a remarkable clade because it includes one of the oldest clupeomorphs described (Neocomian, †*Scutatuspinosus* from Brazil) as well as the oldest clupeomorph from Canada (Albian, †*Foreyclupea* from the Northwest Territories). Considering the earlier mentioned occurrence of clupeiform †*Pseudoellimma gallae* in the Barremian sediments of Brazil, these findings suggest that the two major lineages of clupeomorphs (†ellimmichthyiforms and clupeiforms) diverged prior to that time. Another intriguing fact about this basal clade of †ellimmichthyiforms is loss of dorsal scute series in †*Ranulfoichthys* and †*Foreyclupea*; this condition is not typical for †ellimmichthyiform fishes that are commonly referred to as ‘double-armored’ herrings. Loss of dorsal scutes has been regarded as a derived characteristic of clupeiforms; however, occurrence of this feature in basal

†ellimmichthyiforms suggests that this condition has evolved multiple times within different lineages of Clupeomorpha.

The rest of †Ellimmichthyiformes are recovered in a polytomy comprising three families, †Sorbinichthyidae, †Armigatidae, †Paraclupeidae, and a poorly known taxon, †*Kwangoclupea dartvellei*. The monogeneric family †Sorbinichthyidae comprises two Late Cretaceous species, †*Sorbinichthys africanus* from Morocco and †*S. elusivo* from Lebanon. This group is distinct among other †ellimmichthyiforms and is strongly supported by a number of derived characters: broad, spatula-shaped dorsal process of the posttemporal (22:2), extensive diastema between hypurals two and three (30:3), distinct subrhomboid shape of dorsal scutes (45:2 and 46:2), and posteriormost predorsal scutes with anteroposteriorly inclined lateral processes (62:1).

Phylogenetic position of †*Sorbinichthys* within †Ellimmichthyiformes has changed from being a basal clade with unresolved relationships with other †ellimmichthyiforms (Zaragüeta-Bagils, 2004), or as having a sister-group relationship with †*Diplomystus* (Alvarado-Ortega et al., 2008), to its placement within the suborder of more advanced †ellimmichthyiforms, †Ellimmichthyoidei (Murray and Wilson, 2013). More recently, several studies (De Figueiredo and Ribeiro, 2017; Marramà and Carnevale, 2017; Marramà et al., 2019) have recovered †*Sorbinichthys* close to the base of the order †Ellimmichthyiformes, outside of the suborder †Ellimmichthyoidei Murray and Wilson, 2013 similar to my results.

The rest of †ellimmichthyiforms are grouped into two major clades that most closely correspond to the families †Armigatidae and †Paraclupeidae *sensu* Murray and Wilson (2013). Present analysis recovered †*Armigatus* as a sister group to a clade containing †*Horseshoeichthys armaserratus* and four species of †*Diplomystus*. The sister-group relationship of †*Armigatus* and †*Diplomystus* was previously described in the phylogenetic hypothesis by Murray and Wilson

(2013), who proposed family rank †Armigatidae for the clade. This clade is supported by the following combination of characteristics: S-shaped cleithrum (21:1), fusion of the first uroneural and first ural centrum (34:1), and scutes in predorsal series all being the same size (49:0). Although †*Horseshoeichthys* is known only from a single partially preserved specimen, grouping of this Late Cretaceous (Santonian – Maastrichtian) †ellimmichthyiform from freshwater deposits of Canada with the species of †*Diplomystus* is not surprising – †*H. armaserratus* and †*D. dentatus* from the Eocene freshwater deposits of the Green River Formation (Wyoming, USA) share an overall similarity of the head and body shape as well as dentition and shape of the maxilla as described by Newbrey et al. (2010).

Monophyly of †*Armigatus* is supported by presence of an incomplete predorsal series of scutes (44:0). Forey (2004) recognized heart-shaped scutes forming an incomplete predorsal series as a derived and autapomorphic feature of the genus. Taking into account newly described †ellimmichthyiform taxa lacking predorsal scutes (†*Ranulfoichthys dorsonudum* and †*Foreyclupea loonensis*), it can be inferred that †ellimmichthyiforms showed some variation in degree of the development of dorsal series of scutes with a complete series being the predominant form. This trend is reversed among living clupeiforms with only a few species retaining a complete or partial series of dorsal scutes and most clupeiform fishes lacking scutes along the dorsal margin of the body. Interestingly, †*Codoichthys carnavalii* from the Early Cretaceous of Brazil is recovered as the sister taxon to †*Armigatus*; this grouping is weakly supported by a homoplastic feature of the postpelvic abdominal scutes bearing prominent and strong ventral spines (54:1). Placement of this taxon is different from that found in the most recent phylogenetic analyses by Marramà et al. (2019) and De Figueiredo and Ribeiro (2017), who recovered †*Codoichthys* as a member of †Paraclupeidae.

Sister group to †Armigatidae is a large clade that contains the rest of the †ellimmichthyiforms and is similar in species composition to the family †Paraclupeidae *sensu* Murray and Wilson, 2013 (excluding †Scutatospinosinae). This group is characterized by dorsal margin of the body forming a distinct angle at the base of the dorsal fin (1:1), presence of a neural arch on the first ural centrum (42:1), and ornamentation of dorsal scutes (50:1, this character is not observed in †*Tychoichthys dunveganensis*, but this may be a result of preservation (Hay et al., 2007)).

The monogeneric subfamily †Thorectichthyinae Murray and Wilson, 2013 is the most early diverging clade of †Paraclupeidae. The two species of †*Thorectichthys* are characterized by a short second uroneural that does not reach distal end of the first uroneural (36:1); this feature is also present in more derived †paraclupeids – †*Triplomystus*, †*Tychoichthys dunveganensis*, and †*Paraclupea chetungensis*. Recently described Eocene †paraclupeid, †*Eoellimmichthys superstes*, is recovered as a sister taxon to the rest of †Paraclupeidae; contrary to the results of Marramà et al. (2019), who found this unique taxon in a clade with †*Ellimmichtys longicostatus* and †*E. goodi*.

More derived †paraclupeids are subdivided into two clades, subfamilies †Paraclupeinae Chang and Chou, 1977 and †Triplomistinae Murray and Wilson, 2013. The subfamily †Paraclupeinae includes exclusively Early Cretaceous taxa: †*Paraclupea chetungensis*, †*Ellimma branneri*, and two species of †*Ellimmichthys*, †*E. longicostatus* and †*E. goodi*. These species share presence of distinct skull roof sculpturing with the parietals and frontals strongly ornamented with radiating ridges (4:2) and a deep hypural diastema between the second and third hypurals (30:2; also observed in †*Tychoichthys dunveganensis*). †*Ellimma branneri* along with

†*Ellimmichthys longicostatus* and †*E. goodi* are placed in a tribe †Ellimmichthyini Grande, 1982.

†Triplomystinae comprise a monophyletic †*Triplomystus* and a clade containing †*Tychoichthys dunveganensis*, †*Gasteroclupea branisai*, and †*Ellimmichthys maceioensis*. This clade, however, has weak support and is characterized by plesiomorphic characters – shape of the proximal end of the first hypural (28:1), first uroneural bearing a dorsal expansion of laminar bone (35:1, absent in †*Gasteroclupea*), and distal end of the second uroneural not reaching the distal end of the first uroneural (36:1). Results of my analysis suggest that †*E. maceioensis* does not form a monophyletic group with the type species of †*Ellimmichthys* (†*E. longicostatus*), as also found by Vernygora and Murray (2015) and probably is not a member of that genus.

5.6 Conclusions

Despite the overall similarity among the species of †*Armigatus*, the challenge of delimiting co-occurring congeneric species (†*A. alticorpus*, †*A. namourensis*, and †*A. brevissimus*) can be less daunting in practice when considering a unique combination of meristic characters in †*A. alticorpus*, as well as body depth ratio, and lack of ornamentation on the skull roof, which sets this taxon apart from the other two species commonly found in the same fossil assemblages. Overall, new fossil material indicates a greater faunal similarity between Cenomanian localities in Lebanon, Hakel and Namoura than previously reported. Relatively high species diversity and older age of fossils suggests that the eastern part of the central Tethys could be a potential centre of origin and radiation of the group. This is congruent with the temporal occurrence of taxa

which indicate early emergence of †*Armigatus* in the eastern central Tethys and successive dispersal north (†*A. dalmaticus* from Croatia) and west (†*A. oligodentatus* from Morocco).

References

- Alexandrino, P., R. Faria, D. Linhares, F. Castro, M. Le Corre, R. Sabatié, J. L. Bagliniere, and S. Weiss. 2006. Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. *Journal of Fish Biology* 69:242–259.
- Alvarado-Ortega, J., E. Ovalles-Damián, and G. Arratia. 2008. A review of the interrelationships of the order Ellimmichthyiformes (Teleostei: Clupeomorpha); pp. 257–278 in G. Arratia, H.-P. Schultze, and M.V.H. Wilson (eds.), *Mesozoic Fishes 4 – Homology and Phylogeny*. Verlag Dr. Friederich Pfeil, Munich, Germany, 502 pp.
- Bagley, J. C., F. Alda, M. F. Breitman, E. Bermingham, E. P. van den Berghe, and J. B. Johnson. 2015. Assessing species boundaries using multilocus species delimitation in a morphologically conserved group of neotropical freshwater fishes, the *Poecilia sphenops* species complex (Poeciliidae). *PloS ONE* 10:e0121139.
- Blainville, H. M. de. 1818. Sur les ichthyolites ou les poissons fossiles. Paris, *Nouvelles Dictées d'Histoire naturelle* 37:310–395.
- Chang, M. M. and C. C. Chou. 1977. On late Mesozoic fossil fishes from Zhejiang Province, China. *Memoirs, Institute of Vertebrate Palaeontology and Palaeoanthropology. Academia Sinica* 12:1–60.
- Chang, M. M. and J. G. Maisey. 2003. Redescription of *Ellimma branneri* and *Diplomystus shengliensis*, and relationships of some basal clupeomorphs. *American Museum Novitates* 3404:1–35.

- Dubertret, L. 1959. Contribution à la stratigraphie et à la paléontologie du Crétacé et du Nummulitique de la marge N. W. De la péninsule arabique. Notes et Mémoires sur le Moyen Orient, Paris 7:193–220.
- Dubertret, L. and H. Vautrin. 1937. Révision de la stratigraphie du Crétacé du Liban. Notes et Mémoires, Haut-Commissariat de la République Française en Syrie et au Liban 2:43–73.
- Dubertret, L. 1966. Liban, Syrie et bordure de pays voisins. 1^é partie: tableau stratigraphique. Notes et Mémoires sur le Moyen Orient, Paris 8:249–358.
- Eyo, J. E. 2003. Congeneric discrimination of morphometric characters among members of the Pisces Genus *Clarias* (Clariidae) in Anambra River, Nigeria. *The Zoologist* 2:1–17.
- Figueiredo, F. J. and D. R. Ribeiro. 2017. Reassessment and relationships of †*Scutatuspinosus itapagipensis* (Teleostei, Clupeomorpha, †Ellimmichthyiformes) from the Neocomian of Recôncavo Basin, Northeastern Brazil. *Anais da Academia Brasileira de Ciências* 89:799–823.
- Forey, P. L. 2004. A three-dimensional skull of a primitive clupeomorph from the Cenomanian English Chalk and implications for the evolution of the clupeomorph acustico-lateralis system; pp. 405–427 in G. Arratia, and A. Tintori (eds.), *Mesozoic Fishes 3—Systematics, Paleoenvironments and Biodiversity*. Verlag Dr. Friedrich Pfeil, Munich, 649 pp.
- Forey, P. L., L. Yi, C. Patterson, and C. E. Davies. 2003. Fossil fishes from the Cenomanian (Upper Cretaceous) of Namoura, Lebanon. *Journal of Systematic Palaeontology* 1:227–330.
- Goloboff, P. A., J. S. Farris, and K. C. Nixon. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774–786.

- Grande, L. 1982. A revision of the fossil genus *Diplomystus*, with comments on the interrelationships of clupeomorph fishes. *American Museum Novitates* 2728:1–34.
- Grande, L. 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. *Bulletin of the American Museum of Natural History* 181:231–372.
- Greenwood, P. H., D. E. Rosen, S. H. Weitzman, and G. S. Myers. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History* 131:339–455.
- Hay, M. J., S. L. Cumbaa, A. M. Murray, and A. G. Plint. 2007. A new paraclupeid fish (Clupeomorpha, Ellimmichthyiformes) from a muddy marine prodelta environment: middle Cenomanian Dunvegan Formation, Alberta, Canada. *Canadian Journal of Earth Sciences* 44:775–790.
- Hemleben, C. 1977. Rote Tiden und die oberkretazischen Plattenkalke im Libanon. *Neues Jahrbuch für Geologie und Palaontologie, Monatshefte* 1977:239–255.
- Hückel, U. 1970. Die Fischeschiefer von Hakel und Hjoula in der Oberkreide des Libanon. *Neues Jahrbuch für Geologie und Paläontologie, Abhandlungen, Stuttgart* 135:113–149.
- Hückel, U. 1974. Geochemischer Vergleich der Plattenkalke Solnhofens und des Libanon mit anderer Kalken. *Neues Jahrbuch für Geologie und Paläontologie, Abhandlungen, Stuttgart* 145:279–305.
- Khalloufi, B., R. Zaragüeta-Bagils, and H. Lelièvre. 2010. *Rhombichthys intoccabilis*, gen. et sp. nov. (Ellimmichthyiformes, Clupeomorpha, Teleostei), from the Cenomanian (Upper Cretaceous) of Ein Yabrud, Middle East: anatomical description and phylogenetic implications. *Journal of Vertebrate Paleontology* 30:57–67.

- Maddison, W. P. and D. R. Maddison. 2015. Mesquite: A Modular System for Evolutionary Analysis, version 3.6. Available at <http://mesquiteproject.org>. Accessed November 13, 2019.
- Marramà, G. and G. Carnevale. 2017. The relationships of *Gasteroclupea branisai* Signeux, 1964, a freshwater double-armored herring (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous-Paleocene of South America. *Historical Biology* 29:904–917.
- Marramà, G., A. F. Bannikov, J. Kriwet, and G. Carnevale. 2019. An Eocene paraclupeid fish (Teleostei, Ellimmichthyiformes) from Bolca, Italy: the youngest marine record of double-armoured herrings. *Papers in Palaeontology* 5:83–98.
- McDowall, R. M. 2001. Diadromy, diversity and divergence: implications for speciation process. *Fish and Fisheries* 2:278–286.
- McDowall, R. M. 2003. Shads and anadromy: implications for ecology, evolution, and biogeography; pp. 11–23 in K. E. Limburg and J. R. Waldman (eds.), *Biodiversity, Status, and Conservation of the World's Shads*. Baltimore, MD: American Fisheries Society, 370 pp.
- Müller, J. 1845. Über den Bau und die Grenzen der Ganoiden und über das natürlichen System der Fische. *Abhandlungen Akademie der Wissenschaften, Berlin* 1844:117–216.
- Murray, A. M., and M. V. H. Wilson. 2013. Two new paraclupeid fishes (Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco; pp. 267–290 in G. Arratia, H.-P. Schultze, and M. V. H. Wilson (eds.), *Mesozoic Fishes 5 – Global Diversity and Evolution*. Verlag Dr. Friederich Pfeil, Munich, Germany, 560 pp.
- Murray, A. M., O. Vernygora, S. Japundžić, J. Radovčić, M. V. H. Wilson, D. Bardack, and T. Grande. 2016. Relationships of the species of *Armigatus* (Clupeomorpha,

- Ellimmichthyiformes) and the description of a new species from the Cretaceous of Dalmatia, Croatia. *Journal of Vertebrate Paleontology* 36:p.e1226851.
- Newbrey, M. G., A. M. Murray, D. B. Brinkman, M. H. V. Wilson, and A. G. Neuman. 2010. A new articulated freshwater fish (Clupeomorpha, Ellimmichthyiformes) from the Horseshoe Canyon Formation, Maastrichtian, of Alberta, Canada. *Canadian Journal of Earth Sciences* 47:1183–1196.
- Patterson, C. and D. E. Rosen. 1977. Review of ichthyodectiform and other Mesozoic teleost fishes and the theory and practice of classifying fossils. *Bulletin of the American Museum of Natural History* 158:83–172.
- Ramler, D., A. Palandačić, G. B. Delmastro, J. Wanzenböck, and H. Ahnelt. 2017. Morphological divergence of lake and stream Phoxinus of Northern Italy and the Danube basin based on geometric morphometric analysis. *Ecology and evolution* 7:572–584.
- Robinson, B. W. and K. J. Parsons. 2002. Changing times, spaces, and faces: tests and implications of adaptive morphological plasticity in the fishes of northern postglacial lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 59:1819–1833.
- Saint Marc, P. 1974. Étude stratigraphique et micropaléontologique de l'Albien, du Cenomanien et du Turonien du Liban. *Notes et Mémoires sur le Moyen Orient, Paris* 13:8–42.
- Schram, F. R., C. H. J. Hof, and F.A. Steeman. 1999. Thylacocephala (Arthropoda: Crustacea?) from the Cretaceous of Lebanon and implications for thylacocephalan systematics. *Palaeontology* 42:769–797.
- Sidlauskas, B. L., J. H. Mol, and R. P. Vari. 2011. Dealing with allometry in linear and geometric morphometrics: a taxonomic case study in the *Leporinus cylindrifomis* group

(Characiformes: Anostomidae) with description of a new species from Suriname.
Zoological Journal of the Linnean Society 162:103–130.

Silva Santos, R., and V. L. Silva Corrêa. 1985. Contribuição ao conhecimento da paléoiçtiofaunula do Cretáceo no Brasil; pp. 169–174 in A. de Campos, C. S. Ferreira, I. M. Brito, and C. F. Viana (eds.), Coletânea de Trabalhos Paleontológicos, Série Geologia 27(2). Ministério das Minas e Energia-Departamento Nacional de Produção Mineral, Rio de Janeiro, Brazil, 656 pp.

Smith, W. D., J. J. Bizzarro, V. P. Richards, J. Nielsen, F. Márquez-Flarías, and M. S. Shivji. 2009. Morphometric convergence and molecular divergence: the taxonomic status and evolutionary history of *Gymnura crebripunctata* and *Gymnura marmorata* in the eastern Pacific Ocean. *Journal of Fish Biology* 75:761–783.

Vernygora, O. and A. M. Murray. 2016. A new species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous of Morocco, and its phylogenetic relationships. *Journal of Vertebrate Paleontology* 36:e1031342.

Zaragüeta-Bagils, R. 2004. Basal clupeomorphs and ellimmichthyiform phylogeny; pp. 391–404 in G. Arratia, G. Tintori, and A. Tintori (eds.), *Mesozoic Fishes 3 – Systematics, Paleoenvironment and Biodiversity*. Verlag Dr. Friederich Pfeil, Munich, Germany, 703 pp.

TABLE 5-1. Meristics and measurements (in mm) for the holotype and TMP 1998.65.11 specimens of †*Armigatus alticarpus*.

	Holotype	Specimen A	Specimen B	Specimen C	Specimen D
Total length (TL)	47	63	?	?	?
Standard length (SL)	36	52	?	48	47
Head length (HL)	12	18	?	16	17
Head depth (HD)		18	?	18	17
Body depth (BD)	19	24	26	24	22
Predorsal length (PD)	18	25	?	23	24
Prepelvic length (PV)	20	26	?	?	26
Preanal length (PA)	26	35	?	32	34
HL/SL	33.3	34.6	?	33.3	36.1
HD/SL		34.6	?	37.5	36.1
BD/SL	52.8	46.2	?	50.0	46.8
PD/SL	50.0	48.1	?	47.9	51.1
PV/SL	55.6	50.0	?	?	53.2
PA/SL	72.2	67.3	?	66.7	72.3
Abdominal centra	18	16	16	16	17
Caudal centra (incl. U1, U2)	21	20	20	19	20
Total centra	37	36	36	35	37
Preural centra	35	34	34	33	35
Pairs of ribs	?15	13	13	12+	?14
Predorsal scutes	?	7+	?9	?	7+
Predorsal bones	6	5	6	6	5
Abdominal scutes	24	?	20+	?	21+
Dorsal fin above centrum	15	12	14	-	14
Dorsal fin rays	15	16	17	?	16
Dorsal fin pterygiophores	15	16	17	15+	17
Anal fin rays	?	?	?	?	25+?
Anal fin pterygiophores	27	27	?	27	?26

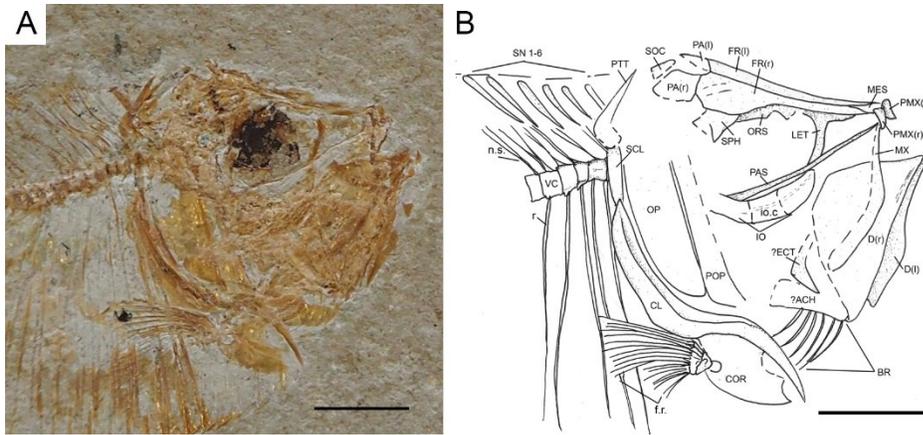


FIGURE 5-1. †*Armigatus alticarpus*, NHMUK P.63134 (holotype). A – photograph of the head; B – line drawing. Scale bar equals 5 mm.

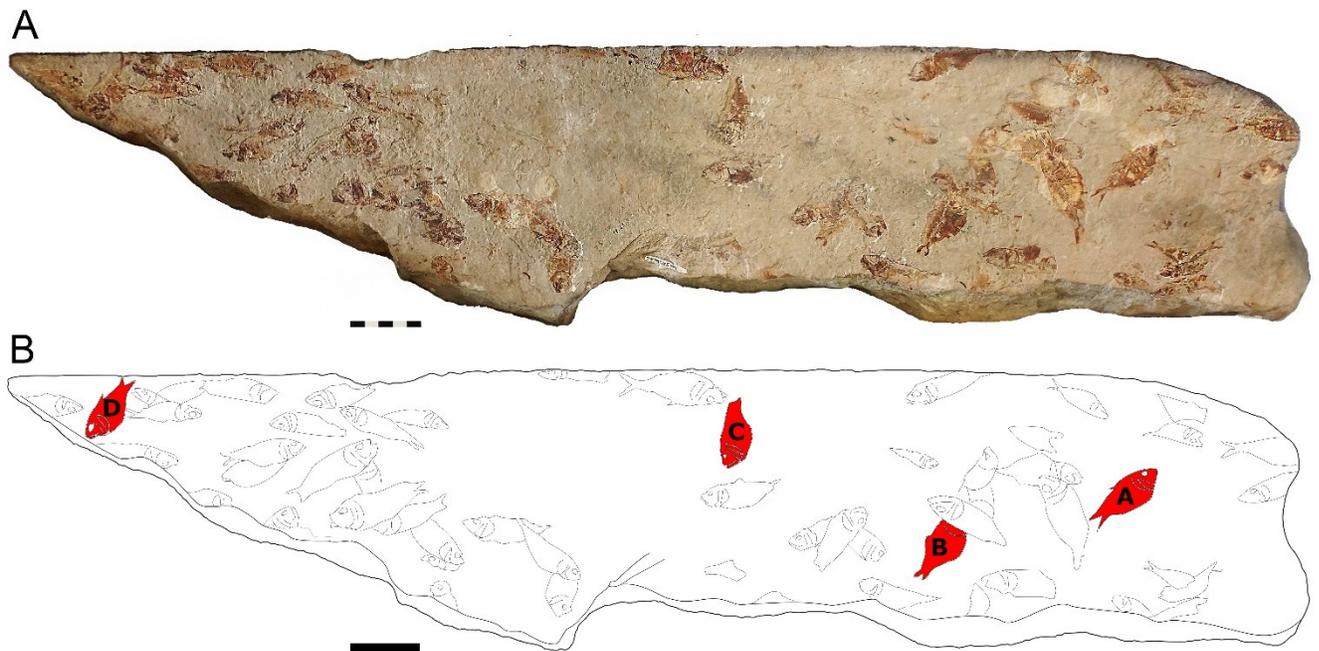


FIGURE 5-2. TMP 1998.65.11, limestone slab preserving multiple clupeomorph specimens. A – photograph and B – line drawing of the specimen block with referred specimens labeled with letters A – D corresponding to how the specimens are referred to in the study. Scale bar equals 5 cm.

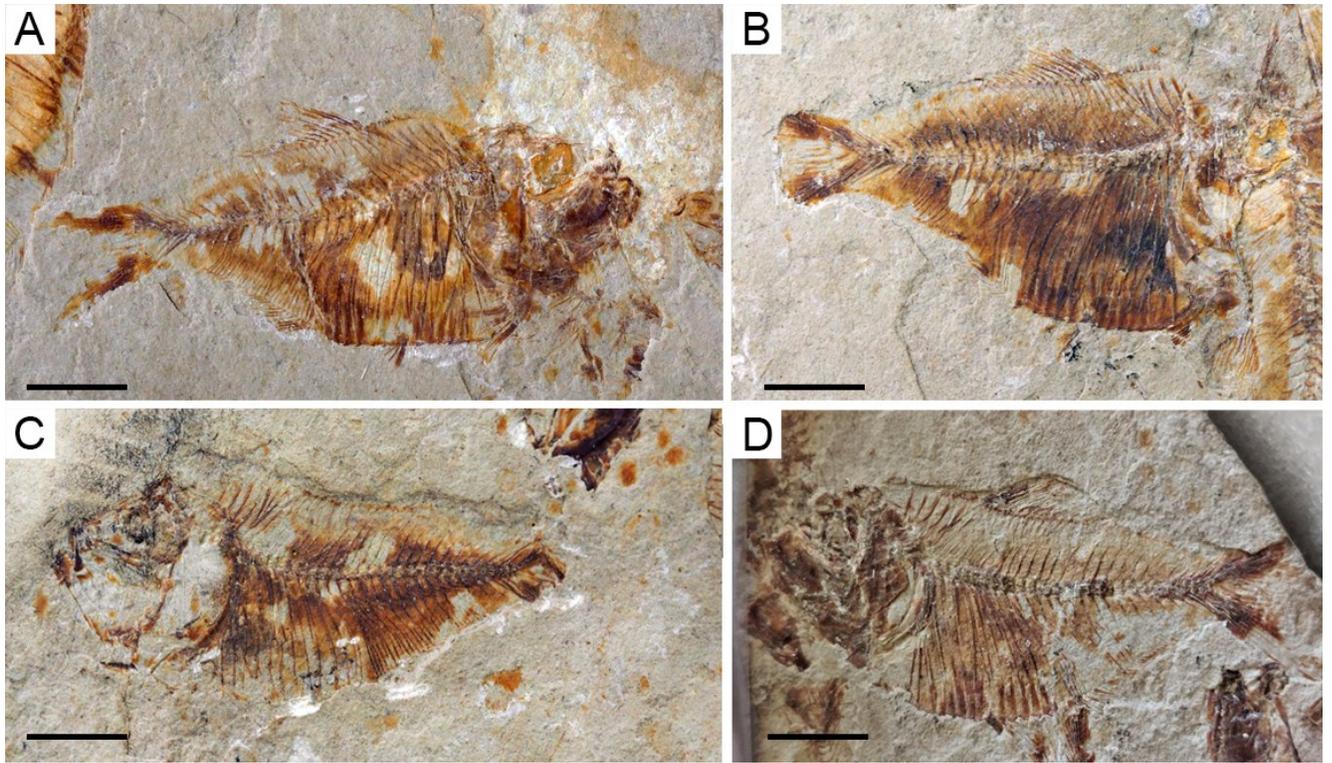


FIGURE 5-3. TMP 1998.65.11, four †*Armigatus alticorpus* specimens referred in this study.

Scale bar equals 10 mm.

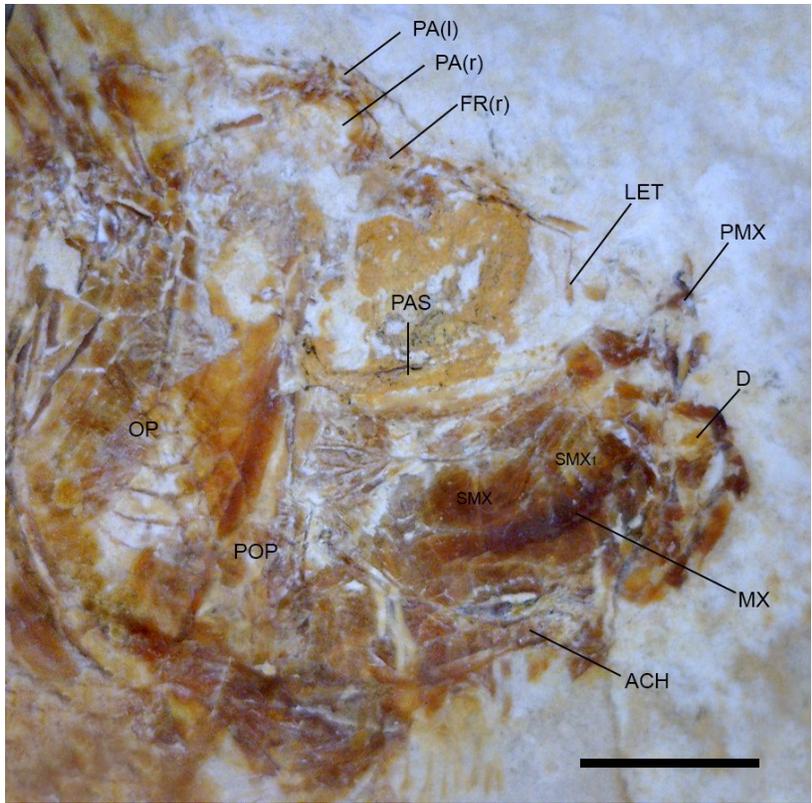


FIGURE 5-4. Close up of the head of the TMP 1998.65.11 specimen A. Scale bar equals 5 mm.

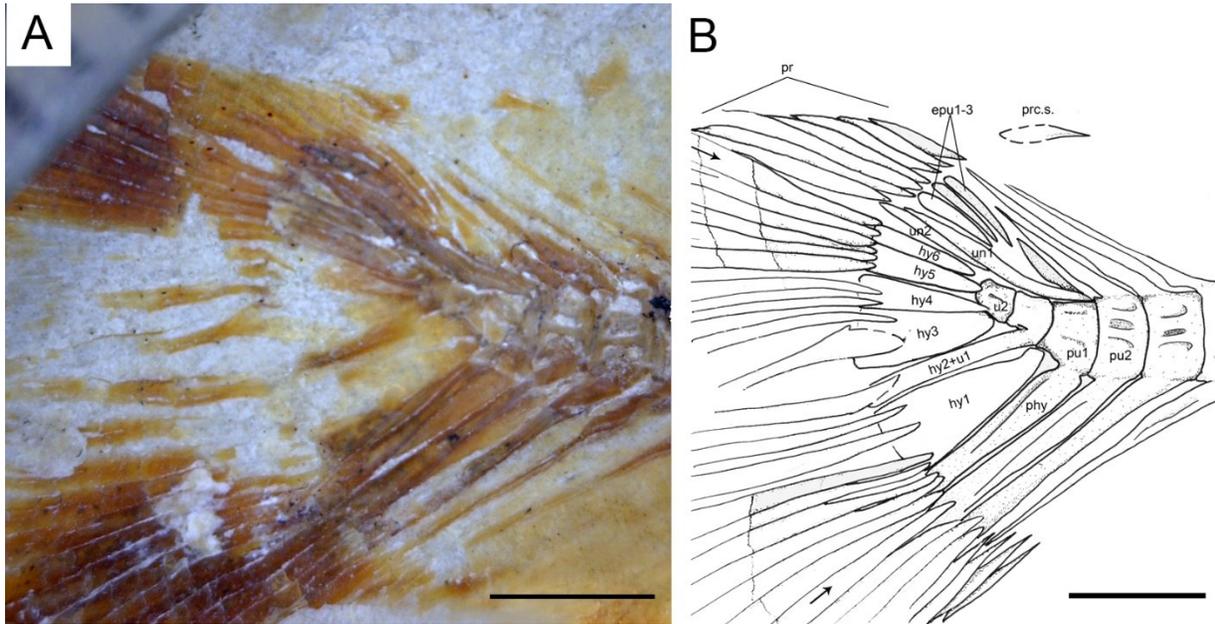


FIGURE 5-5. Caudal fin of †*Armigatus alticarpus*. A – photograph and B – line drawing, TMP 1998.65.11 specimen B. Scale bar equals 2 mm.

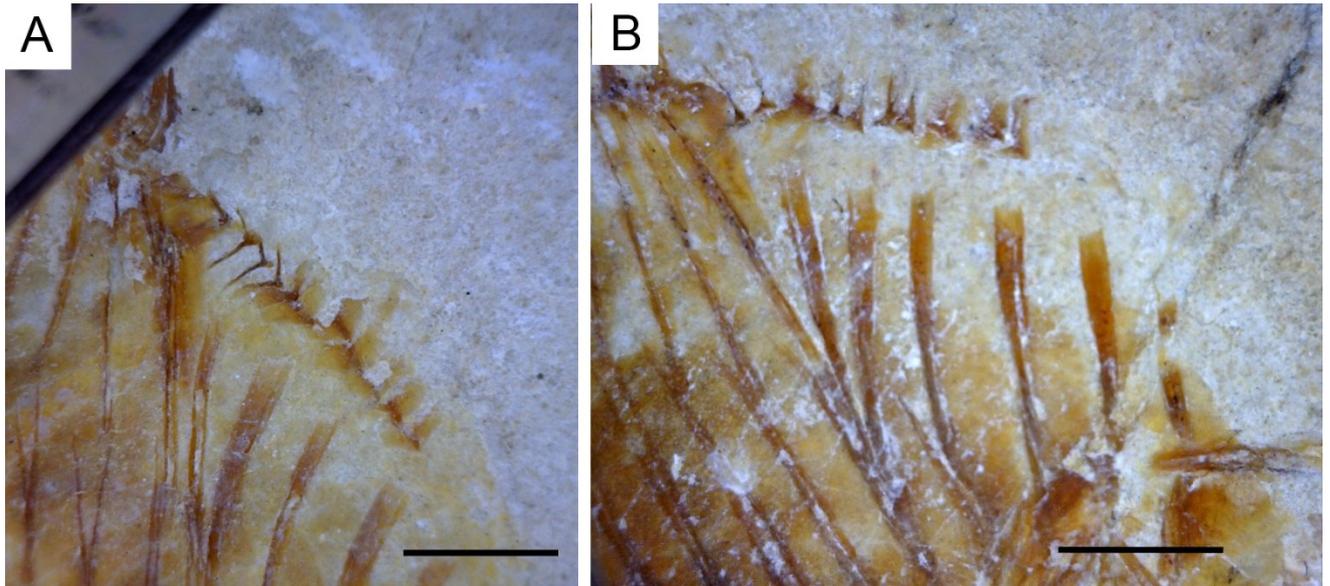


FIGURE 5-6. Dorsal series of scutes of †*Armigatus alticarpus*. A – TMP 1998.65.11 specimens

A; B – specimen D. Scale bar equals 2 mm.

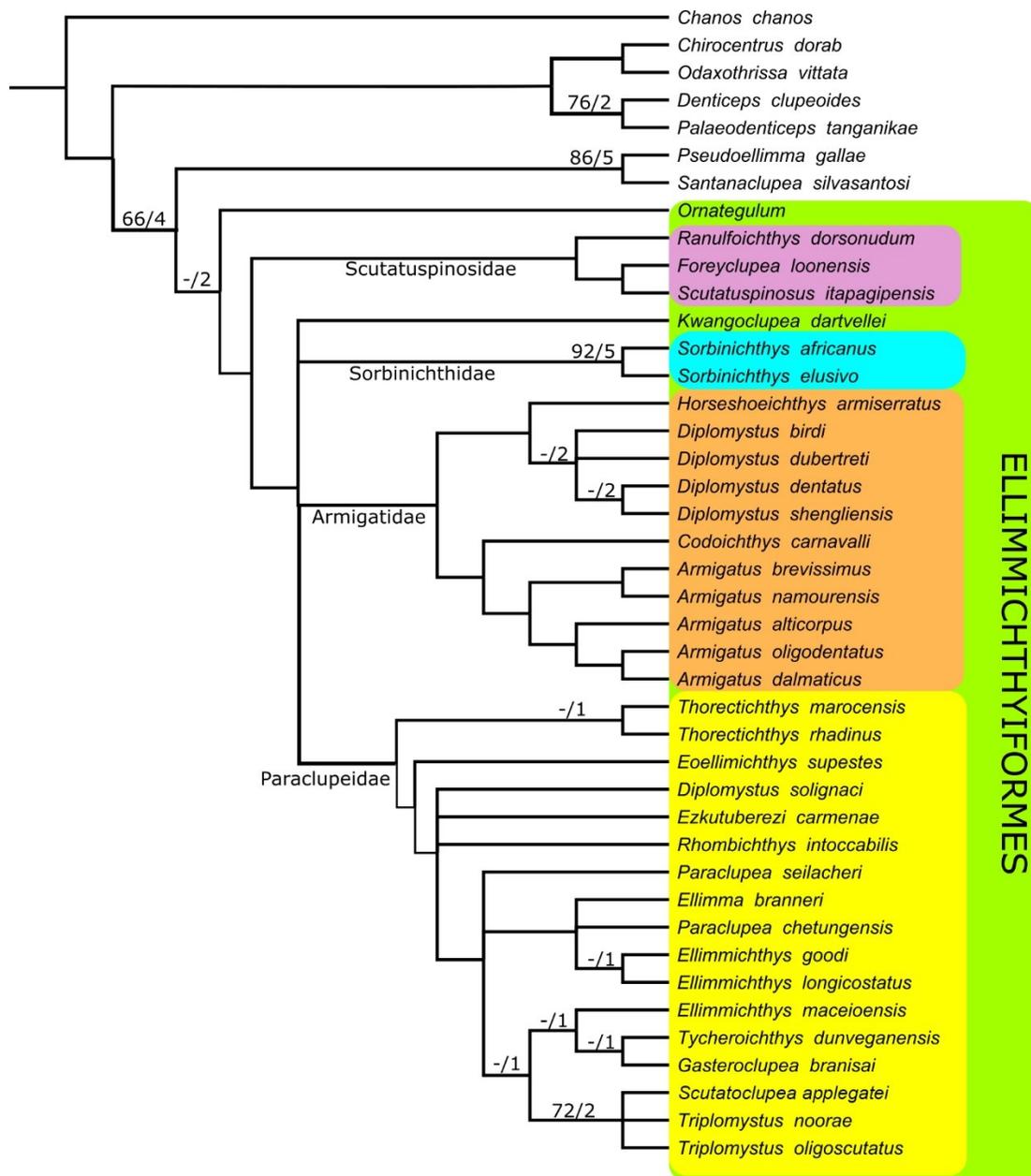


FIGURE 5-7. Majority rule consensus tree of 738 most parsimonious trees. Tree metrics: length = 271 steps, CI = 0.3358, RI = 0.6457. Numbers above branches indicate bootstrap (>50%) and Bremer support values to the left and the right of the slash, respectively.

APPENDIX 5-1. List of characters used in phylogenetic analysis of the †Ellimmichthyiformes dataset:

1. Anterior dorsal margin of body: [0] rounded and convex; [1] almost straight, forming a marked angle at the dorsal fin insertion.
2. Skull roof condition: [0] parietal bones contacting each other in the midline; [1] supraoccipital separates parietal bones.
3. Lateral profile of skull roof: [0] a straight line from anterior tip of frontal to back of skull, with no distinct angle apparent; [1] with distinct angle between anterior and posterior parts, normally in the region of the parietal.
4. Ornamentation of skull roof: [0] absent; [1] present, fine, more or less parallel ridges; [2] present, strong grooves with numerous fine, radiating ridges.
5. Posttemporal fossa: [0] absent; [1] present.
6. Cavity in the temporal region of the skull: [0] pre-epioccipital fossa (between parietal, epioccipital and pterotic bones); [1] pre-epioccipital fenestra (between the parietal, epioccipital and supraoccipital bones); [2] absence of cavity or fenestra.
7. Recessus lateralis: [0] absent; [1] present.
8. Supramaxillary bones: [0] two; [1] one or none.
9. ‘Basipterygoid’ process of parasphenoid: [0] absent; [1] present.
10. ‘Osteoglossid’ tooth patch on the parasphenoid: [0] absent; [1] present.
11. Supraorbital bone: [0] absent; [1] present.
12. Antorbital bone: [0] absent; [1] present.
13. Beryciform foramen within the anterior ceratohyal: [0] absent; [1] present.
14. Foramen in posterior ceratohyal: [0] absent; [1] present.

15. Teeth on endopterygoid: [0] absent; [1] present.
16. Total number of vertebrae excluding ural centra: [0] 30-40; [1] 41-43; [2] more than 50.
17. Halves of the neural arches of most abdominal vertebrae: [0] separate medially; [1] fused medially.
18. Pleural ribs: [0] all ribs articulate with parapophyses along the abdominal region; [1] anteriormost ribs articulate with deep pits on the lateral side of all abdominal centra and those located posteriorly articulate with well-developed parapophyses; [2] all ribs articulate with deep pits on the lateral side of all abdominal centra.
19. Epineurals and epipleurals in the caudal region: [0] absent; [1] present.
20. Epicentrals: [0] absent; [1] present.
21. Shape of cleithrum: [0] L-like (having a single angle in the bone); [1] S-like (having two angles).
22. Dorsal process of posttemporal: [0] slender and sharp; [1] sub-rectangular; [2] broad, wider at distal tip than at midpoint of bone.
23. Number of anal fin rays: [0] eight to eleven; [1] fourteen or fifteen; [2] seventeen or eighteen; [3] twenty; [4] twenty-two to thirty-two; [5] thirty-six to forty-one.
24. Number of dorsal fin rays: [0] eight to thirteen; [1] fourteen to nineteen; [2] twenty-one to twenty-five.
25. Number of hypurals: [0] seven; [1] six; [2] five.
26. Hypural 2: [0] autogenous; [1] fused to first ural centrum (diural terminology).
27. Length of hypural 1: [0] long, reaching ural centrum 1; [1] short, not reaching ural centrum 1.
(Diural terminology)

28. Proximal end of hypural 1: [0] massive and forming an upward process; [1] sharp; [2] massive but no upward process.
29. Shape of hypural 2: [0] distal end distinctly broader than proximal end; [1] very thin and stick-like.
30. Diastema between second and third hypural: [0] third hypural not expanded posteriorly leaving a gap or notch between the second and third hypural; [1] third hypural expanded posteriorly, leaving a small triangular notch between second and third hypural; [2] third hypural expanded posteriorly, leaving a deep triangular notch between second and third hypural; [3] third hypural expanded posteriorly and has a concave ventral edge forming a large concavity between second and third hypurals.
31. Size of first ural centrum (diural terminology): [0] roughly the same size (length and depth) as the preural centra; [1] much smaller than the preural centra [1].
32. Number of uroneurals: [0] three; [1] two; [2] one.
33. First uroneural: [0] extends anteriorly to reach second preural centrum; [1] does not reach second preural centrum.
34. Fusion of first uroneural and first ural centrum: [0] absent; [1] present.
35. First uroneural bearing a dorsal expansion of laminar bone: [0] absent; [1] present..
36. Distal end of second uroneural: [0] reaching the distal end of the first uroneural; [1] not reaching the distal end of the first uroneural.
37. Parhypural: [0] base/arch of bone fused with preural centrum 1; [1] autogenous.
38. Number of epurals: [0] three; [1] two; [2] none, or those present are weakly ossified, perhaps cartilaginous.

39. Position of epurals: [0] epurals fill the space between the neural spines of pu1 and pu2; [1] epurals are located far from the spine of pu2, leaving an open space between them.
40. Caudal scutes: [0] absent; [1] present.
41. Neural spine of first preural centrum: [0] large or lanceolate; [1] short or sub-rectangular.
42. Neural arch of first ural centrum: [0] absent; [1] present.
43. Predorsal scutes: [0] absent; [1] present.
44. Predorsal scute series: [0] incomplete (absent in anterior part); [1] complete.
45. Shape of scutes in anterior part of predorsal series: [0] subrectangular; [1] heart-shaped or ovoid; [2] at least some rhomboid scutes present.
46. Shape of scutes in posterior part of predorsal series: [0] subrectangular; [1] heart-shaped or ovoid; [2] at least some rhomboid scutes present.
47. Series of spines on the posterior margin of the lateral wings of the predorsal scutes: [0] absent; [1] present.
48. Prominent median strong spine on posteriormost predorsal scutes: [0] absent; [1] present.
49. Size of scutes of predorsal series: [0] all scutes of same size; [1] irregular in size, size of scutes increasing posteriorly.
50. Surface of predorsal scutes: [0] smooth; [1] ornamented with radiating grooves.
51. Number of predorsal scutes: [0] six to fourteen; [1] sixteen to nineteen; [2] twenty to forty-one.
52. Abdominal scute series: [0] absent; [1] present.
53. Complete abdominal scute series between isthmus and anus (i. e., postpelvic scutes are present): [0] absent; [1] present.

54. Postpelvic abdominal scutes bearing very prominent and strong ventral spine: [0] absent; [1] present.
55. Size of lateral wings of abdominal scutes: [0] small; [1] large, extended upward and covering the abdominal cavity laterally for at least one quarter of the distance from ventral body edge to vertebral column.
56. Shape of lateral wing of abdominal series scutes: [0] spine-like, with large spaces between wings of scutes; [1] wide or spatula-like, with wings of adjacent scutes touching for most of their length.
57. Postdorsal scute series: [0] absent; [1] present.
58. Number of abdominal scutes (in some taxa they will not be equivalent to vertebral counts): [0] fewer than 20; [1] 22-30; [2] more than 32.
59. Number of predorsal bones: [0] 10 or more; [1] 7-9; [2] 6 or fewer.
60. Predorsal bones (supraneurals) forming a fan-shaped structure with at least one anteriormost predorsal bone inclined anterodorsally-posteroventrally to meet proximal end of the next supraneural: [0] absent; [1] present.
61. Position of the pelvic fin anterior to the origin of the dorsal fin: [0] absent; [1] present.
62. Lateral projections of the most posterior predorsal scutes inclined antero-dorsally: [0] absent; [1] present.

APPENDIX 5-2. Data matrix used in the phylogenetic analysis.

<i>Chanos_chanos</i>	01000001001?0001001110011011100110111200100???????0???0?2000
<i>Chirocentrus_dorab</i>	01001?1100110002121101411112101111001100000???????1000000010
<i>Odaxothrissa_vittata</i>	0100101100110001101100311112101111001001000???????11000022000
<i>Denticeps_clupeoides</i>	01001011001101001?1100402102001101011100110???????11000032010
† <i>Palaeodenticeps_tanganikae</i>	01001??1?01?0?001???004021020011010?110???0???????1110?0?010
† <i>Armigatus_alticorpus</i>	0000?00101???000010104111011001?00000011?10110000011010031000
† <i>Armigatus_brevissimus</i>	000?1?0011?11?0001?1041110110?1000000011010110000011?10021000
† <i>Armigatus_dalmaticus</i>	0?00??0?0?1?1001101?20110110010?0000011010110000011010031000
† <i>Armigatus_namourensis</i>	00011?00111?10100010104111011001000000011010110000211??0041000
† <i>Armigatus_oligodentatus</i>	0000?00101?10000110101011011001000000011010110000011010021000
† <i>Codoichthys_carnavalli</i>	0011??00?0?1??00?110001111001201010?000?1011110000011000021000
† <i>Diplomystus_birdi</i>	0?11?20011?10??012101141110?1001010000?10?110010003111100?1100
† <i>Diplomystus_dentatus</i>	0011?2?0110000111210115011011001010000010011001000?11110051100
† <i>Diplomystus_dubertreti</i>	0011?200?????001?1?1142110?100?0?0?001???1100100031111?0?1100
† <i>Diplomystus_shengliensis</i>	0011??01?0??11121?1050010?10010?0000010?110010005111100?1100
† <i>Diplomystus_solignaci</i>	1?1???001?????11110?1?121?0?10000?0000011?11?0?0???111100?1000
† <i>Ellimma_branneri</i>	0002??00100?1?00110?111111001200000000010?11100111111111041000
† <i>Ellimmichthys_goodi</i>	1102??00?????101101111111001200010000010111100111111111061000
† <i>Ellimmichthys_longicostatus</i>	1102??00?0????101101111111001200010000110111100111111111061010
† <i>Ellimmichthys_maceioensis</i>	1?01??00?????00?????1?11?10210000011001011111001111111?051000
† <i>Eoellimichthys_supestes</i>	1000??0011??1?10110?10411??1?0?????????1111110111011110012010
† <i>Ezkutuberezi_carmenae</i>	1001??00?????00110?1?42?102100001000001??11100111?11111042000
† <i>Foreyclupea_loonensis</i>	0001??00?01??0?011101?1?????????????????0?????????1110001?010
† <i>Gasteroclupea_branisai</i>	2000??00????101011102010100112001001001?1111000100211000060010
† <i>Horseshoeichthys_armiserratus</i>	0000?100??00????02?0?1?????????????????????1?010?0?1??0????00
† <i>Kwangoclupea_dartvellei</i>	0001?10010????00?1??41?1001200000000001011110010?1111?0?1000
† <i>Ornatogulum_sardinoides</i>	0002?10010111?12001000010100100001000000100?????????0????0?2000

† <i>Paraclupea_chetungensis</i>	1002??0010?010101101111111001200000?00111?11100111211111061000
† <i>Paraclupea_seilacheri</i>	1002??0010?111100110111101001?00000000010?11100111111110061000
† <i>Pseudoellimma_gallae</i>	0101111000????000010001121001110000?011?011???01?0?11100031000
† <i>Ranulfoichthys_dorsonudum</i>	0001?000?0111100011001011100100000000001100????????11100042000
† <i>Rhombichthys_intoccabilis</i>	1001?10010101?1011001012?100100000100000011100011111111061000
† <i>Santanaclupea_silvasantosi</i>	01011110?01?100000000?012100111000100?1?0?0????????111010?1000
† <i>Scutatoclupea_bacchiae</i>	0002????00????011?1?1?2111??120??0??0?00111221100011111121000
† <i>Scutatuspinosus_itapagipensis</i>	0001???0101?1100110001001100100?000000011011120010011101022010
† <i>Sorbinichthys_africanus</i>	1010??001010??1011100240?100130000000??0001122110?411010012001
† <i>Sorbinichthys_elusivo</i>	1011??0010?01000121002411100130000000??0?01122111?411010012001
† <i>Thorectichthys_marocensis</i>	1001??0010??101011101041110010000?010001111111001?011110031000
† <i>Thorectichthys_rhadinus</i>	1000??0010??101011101041110010000?010001111111001?011110031000
† <i>Scutatoclupea_applegatei</i>	1001????100????011??1121?1011000011100011?1100011111111152000
† <i>Triplomystus_noorae</i>	1001120010?0??101111112211011000011100011?1100011?1111111?1000
† <i>Triplomystus_oligoscutatus</i>	100112001000??101111114111011000011100011?1100011111111131000
† <i>Tychoichthys_dunveganensis</i>	1010???01000??01110114111??1201001100101011000110211111061000



APPENDIX 5-3. Majority rule consensus tree with node and terminal taxa labeled

APPENDIX 5-4. List of unambiguously optimized characters (the same optimization under ACCTRAN and DELTRAN algorithms) supporting internal nodes and terminal taxa in the majority rule consensus tree with node numbers as shown in Figure S1.

<i>Chirocentrus dorab</i> :	† <i>Codoichthys carnavalli</i> :	† <i>Ellimmichthys longicostatus</i> :
Char. 18: 0 --> 2	Char. 3: 0 --> 1	Char. 39: 0 --> 1
Char. 22: 0 --> 1	Char. 23: 4 --> 1	Char. 61: 0 --> 1
Char. 53: 1 --> 0	Char. 30: 0 --> 2	
Char. 59: 2 --> 0	Char. 55: 1 --> 0	
		† <i>Ellimmichthys maceioensis</i> :
<i>Odaxothrissa vittata</i> :	† <i>Diplomystus birdi</i> :	Char. 15: 1 --> 0
Char. 23: 4 --> 3	Char. 12: 0 --> 1	Char. 28: 1 --> 2
Char. 38: 1 --> 0		
Char. 40: 0 --> 1	† <i>Diplomystus dentatus</i> :	† <i>Eoellimmichthys supestes</i> :
	No autapomorphies	Char. 4: 1 --> 0
<i>Denticeps clupeioides</i> :		Char. 10: 0 --> 1
No autapomorphies	† <i>Diplomystus dubertreti</i> :	Char. 59: 1 --> 2
<i>Palaeodenticeps tanganikae</i> :	Char. 24: 1 --> 2	Char. 61: 0 --> 1
No autapomorphies	Char. 39: 0 --> 1	
	† <i>Diplomystus shengliensis</i> :	† <i>Foreyclupea loonensis</i> :
† <i>Ornategulum sardinoides</i> :	Char. 22: 1 --> 0	Char. 20: 0 --> 1
Char. 4: 1 --> 2	Char. 25: 1 --> 0	
Char. 15: 0 --> 1		† <i>Ezkutuberezi carmenae</i> :
Char. 16: 0 --> 2	† <i>Tychoichthys dunveganensis</i> :	Char. 15: 1 --> 0
Char. 25: 1 --> 0	Char. 3: 0 --> 1	Char. 23: 1 --> 4
Char. 34: 0 --> 1	Char. 23: 1 --> 4	Char. 28: 0 --> 2
Char. 52: 1 --> 0	Char. 32: 0 --> 1	Char. 34: 0 --> 1
	Char. 42: 1 --> 0	Char. 58: 6 --> 4
† <i>Armigatus alticorpus</i> :		Char. 59: 1 --> 2
No autapomorphies	† <i>Diplomystus solignaci</i> :	† <i>Gasteroclupea branisai</i> :
† <i>Armigatus brevissimus</i> :	Char. 3: 0 --> 1	Char. 1: 1 --> 2
No autapomorphies	Char. 16: 0 --> 1	Char. 21: 1 --> 2
† <i>Armigatus namourensis</i> :	Char. 48: 1 --> 0	Char. 22: 1 --> 0
Char. 51: 0 --> 2	Char. 56: 1 --> 0	Char. 24: 1 --> 0
Char. 58: 2 --> 4		Char. 26: 1 --> 0
	† <i>Ellimma branneri</i> :	Char. 33: 0 --> 1
† <i>Armigatus oligodentatus</i> :	Char. 1: 1 --> 0	Char. 35: 1 --> 0
No autapomorphies	Char. 15: 1 --> 0	Char. 49: 1 --> 0
	Char. 58: 6 --> 4	Char. 54: 1 --> 0
† <i>Armigatus dalmaticus</i> :	† <i>Ellimmichthys goodi</i> :	Char. 55: 1 --> 0
Char. 15: 0 --> 1	No autapomorphies	Char. 56: 1 --> 0
		Char. 59: 1 --> 0

Char. 61: 0 --> 1	Char. 41: 1 --> 0	Char. 23: 0 --> 4
† <i>Horseshoeichthys armiserratus</i> :	Char. 42: 0 --> 1	Char. 34: 0 --> 1
Char. 4: 1 --> 0	Char. 43: 0 --> 1	
	Char. 52: 0 --> 1	
† <i>Kwangoclupea dartvellei</i> :	Char. 57: 0 --> 1	Node 47 :
Char. 30: 0 --> 2	Char. 59: 2 --> 1	No synapomorphies
	† <i>Scutatuspinosus itapagipensis</i> :	Node 48 :
† <i>Paraclupea chetungensis</i> :	Char. 17: 0 --> 1	Char. 24: 1 --> 0
Char. 39: 0 --> 1	Char. 19: 1 --> 0	Char. 25: 1 --> 2
Char. 51: 1 --> 2	Char. 24: 1 --> 0	Char. 29: 1 --> 0
	Char. 43: 0 --> 1	Node 49 :
† <i>Paraclupea seilacheri</i> :	Char. 56: 0 --> 1	Char. 2: 1 --> 0
Char. 12: 0 --> 1	Char. 58: 1 --> 2	Char. 7: 1 --> 0
Char. 14: 0 --> 1		Char. 9: 0 --> 1
Char. 17: 1 --> 0	† <i>Sorbinichthys africanus</i> :	Char. 38: 1 --> 0
Char. 25: 1 --> 0	Char. 4: 1 --> 0	
Char. 56: 1 --> 0	Char. 15: 0 --> 1	Node 50 :
	Char. 24: 1 --> 0	Char. 4: 0 --> 1
† <i>Pseudoellimma gallae</i> :	Char. 49: 1 --> 0	Char. 6: 0 --> 1
Char. 23: 0 --> 1		Char. 8: 1 --> 0
Char. 43: 0 --> 1	† <i>Sorbinichthys elusivo</i> :	Char. 13: 0 --> 1
	Char. 18: 1 --> 2	Char. 20: 1 --> 0
† <i>Ranulfoichthys dorsonudum</i> :		Char. 32: 1 --> 0
Char. 58: 1 --> 4	† <i>Thorectichthys marocensis</i> :	Char. 37: 1 --> 0
	No autapomorphies	Node 51 :
† <i>Rhombichthys intoccabilis</i> :	† <i>Thorectichthys rhadinus</i> :	Char. 4: 1 --> 0
Char. 35: 0 --> 1	Char. 4: 1 --> 0	Node 52 :
Char. 40: 1 --> 0		Char. 44: 1 --> 0
Char. 45: 1 --> 0	† <i>Scutatoclupea applegatei</i> :	Node 53 :
	Char. 59: 1 --> 2	Char. 12: 0 --> 1
† <i>Santanaclupea silvasantosi</i> :	† <i>Triplomystus noorae</i> :	Char. 54: 1 --> 0
Char. 19: 1 --> 0	Char. 24: 1 --> 2	
Char. 35: 0 --> 1		Node 54 :
Char. 56: 0 --> 1	† <i>Triplomystus oligoscutatus</i> :	Char. 32: 0 --> 1
	Char. 23: 2 --> 4	Char. 49: 1 --> 0
† <i>Scutatoclupea bacchiae</i> :	Node 45 :	Node 55 :
Char. 2: 1 --> 0	Char. 41: 1 --> 0	Char. 12: 1 --> 0
Char. 4: 0 --> 2		Char. 23: 0 --> 4
Char. 17: 0 --> 1	Node 46 :	Char. 43: 0 --> 1
Char. 23: 0 --> 2	Char. 17: 0 --> 1	Char. 55: 0 --> 1
Char. 26: 0 --> 1		
Char. 30: 0 --> 2		
Char. 37: 1 --> 0		

Char. 59: 2 --> 1
Node 56 :
Char. 18: 0 --> 1

Node 57 :
Char. 10: 0 --> 1

Node 58 :
Char. 24: 1 --> 0

Node 59 :
Char. 3: 0 --> 1
Char. 6: 1 --> 2

Node 60 :
Char. 11: 1 --> 0
Char. 18: 1 --> 2
Char. 22: 0 --> 1
Char. 46: 1 --> 0
Char. 47: 0 --> 1

Node 61 :
Char. 15: 0 --> 1
Char. 16: 0 --> 1
Char. 23: 4 --> 5
Char. 24: 1 --> 0

Node 62 :
Char. 4: 1 --> 0

Char. 30: 0 --> 2
Char. 50: 1 --> 0
Char. 51: 1 --> 2

Node 63 :
Char. 39: 0 --> 1
Char. 40: 1 --> 0

Node 64 :
Char. 28: 0 --> 1
Char. 35: 0 --> 1
Char. 36: 0 --> 1

Node 65 :
Char. 11: 1 --> 0
Char. 22: 0 --> 1
Char. 24: 2 --> 1

Node 66 :
Char. 23: 4 --> 1
Char. 24: 1 --> 2
Char. 46: 1 --> 0
Char. 51: 0 --> 1
Char. 56: 0 --> 1
Char. 58: 1 --> 6

Node 67 :
Char. 19: 1 --> 0
Char. 48: 0 --> 1

Node 68 :
Char. 1: 0 --> 1
Char. 42: 0 --> 1
Char. 50: 0 --> 1

Node 69 :
Char. 4: 0 --> 2
Char. 30: 0 --> 2

Node 70 :
Char. 2: 0 --> 1
Char. 34: 0 --> 1

Node 71 :
Char. 61: 0 --> 1

Node 72 :
Char. 14: 0 --> 1
Char. 22: 0 --> 1

Node 73 :
Char. 25: 1 --> 2
Char. 30: 0 --> 1
Char. 39: 0 --> 1
Char. 41: 1 --> 0
Char. 59: 2 --> 1

CHAPTER 6

Phylogenetic analysis of Clupeomorpha

6.1 Introduction

Evolutionary relationships within the superorder Clupeomorpha have been a problematic issue that has attracted the interest of many researchers. Higher and lower order clupeomorph interrelationships have been addressed by many authors using various approaches and data types (e.g., Whitehead, 1962; Greenwood et al., 1966; Forey, 1975, 2004; Grande, 1982; 1985; Chang and Maisey, 2003; Zaragüeta-Bagils, 2004; Di Dario and De Pinna, 2006; Lavoué et al. 2007, 2008, 2010; 2013, 2014; Li and Orti, 2007; Alvarado-Ortega et al., 2008; Murray and Wilson, 2013; Bloom and Lovejoy, 2014; Bloom et al., 2018). Despite major challenges of identifying reliable morphological traits and obtaining genetic data to diagnose the Clupeomorpha and to delimit natural groups within it, taxonomic research on the phylogenetic interrelationships of the group has seen substantial progress in the past five decades. Major advances in the systematics of the clupeomorph fishes include descriptions of new species (e.g., Vernygora and Murray, 2016; Murray et al., 2016; Alvarado-Ortega and Melgarejo-Damián, 2017; Marramà et al., 2019; Hata et al., 2020), detailed re-examination of previously described taxa (e.g., Figueiredo and Ribeiro, 2016; Marramà and Carnevale, 2017), increased taxonomic sampling in phylogenetic analyses (e.g., Lavoué et al. 2013; Bloom and Lovejoy, 2014; Bloom and Egan, 2018; Bloom et al., 2018), and adoption of cladistic methods of phylogenetic inference. However, a holistic analysis that would integrate morphological and molecular data for the extinct and living clupeomorph taxa in a cladistic framework to produce a unifying taxonomic classification is still lacking.

The type genus of Clupeomorpha was established by Carl Linnaeus as one of 47 fish genera listed in the 10th edition of *Systema Naturae* (Linnaeus, 1758). The genus *Clupea* was placed in the order Abdominales and contained ten species. Diagnosis of the genus as well as the order was based largely on ambiguous homoplastic traits such as position of the pelvic fins, number of branchiostegal and fin rays, and colouration. The lack of unambiguous derived diagnostic features led to a subsequent expansion of the group by inclusion of species that shared primitive features and were otherwise difficult to classify. In later works (e.g., Gill, 1872; Woodward, 1895; Goodrich, 1909; Jordan, 1923; Berg, 1940; Gosline, 1960), clupeomorphs were placed in various groups together with elopomorphs, †ichthyodectiforms, gonorynchiforms, †leptolepiforms, salmoniforms, and other taxa sharing primitive morphological features.

A major milestone in clupeomorph systematics was marked by the work of Greenwood et al. (1966), who revised the diagnosis of the superorder and established three major character complexes that distinguished members of this group: inner ear-swim bladder connection, architecture of the neurocranium (presence of the recessus lateralis, temporal foramina, pre-epiotic fossae, and auditory fenestrae), and caudal fin skeleton (Fig. 6-1). While their study was focused primarily on the higher rank classification of the extant teleostean lineages and did not provide many details about either the composition or interrelationships of the clupeomorph groups, the authors noted that Denticipitoidei are likely to be the most primitive living clupeomorph group. Subsequent workers (e.g., Patterson, 1967; Greenwood, 1968; Nelson, 1970, 1973; Patterson and Rosen, 1977; Gosline, 1980; Grande, 1982, 1985; Whitehead, 1985; Sato, 1994; Di Dario, 2002, 2009; Miyashita, 2010) refined and elaborated this classification by conducting extensive morphological examinations of the members of the group and adding fossil

taxa into the clupeomorph classification. However, interrelationships of the major lineages remained largely unresolved and to a large extent conflicted across these studies.

Nelson (1970) recognized two suborders within Clupeiformes, the Denticipitoidei and Clupeoidei comprised of four superfamilies – Chirocentroidea, Engrauloidea, Pristigasteroidea, and Clupeoidea; however, interrelationships among these lineages were left unresolved.

Patterson and Rosen (1977) assessed Mesozoic representatives of major teleostean lineages; in their classification, the fossil clupeomorph taxa †*Ornategulum sardinoides* and †‘Diplomystidae’ were placed within the order Clupeiformes, but outside of either of the two recognized suborders (Denticipitoidei and Clupeoidei).

The classification of Clupeomorpha started to take its modern form in the foundational works of Grande (1982, 1985) who conducted extensive anatomical surveys of living and fossil clupeomorphs. According to Grande (1985), Clupeomorpha included two divisions – Division 1 containing a single fossil species †*Erichalcis arcta* (originally Grande’s (1982) Division 1 included †*Ornategulum* but it was removed in the later systematic revision) and Division 2 containing all other fossil and living clupeomorphs. Within Division 2, Grande established the new order †Ellimmichthyiformes comprised of the fossil genera of double-armoured herrings †*Diplomystus* and †*Ellimmichthys*. This new order was placed in a polytomy with the traditionally recognized Clupeiformes, and a lineage of uncertain affinities that contained a single fossil genus †*Armigatus*. Subdivisions within Clupeiformes followed, in part, previous classifications recognizing two suborders, Denticipitoidea containing only two genera (*Denticeps* and †*Paleodenticeps*) and Clupeoidei comprised of the Engrauloidea, Pristigasteroidea, and Clupeoidea. Evolutionary relationships among the clupeoid lineages were left unresolved as well as the interrelationships of the five clupeid subfamilies, Clupeinae, Alosinae, Dorosomatinae,

Pellonulinae, and Dussumieriinae. Of the five clupeid subfamilies, Grande (1985) provided diagnostic morphological features only for Pellonulinae and Dussumieriinae; the monophyly of Clupeinae, Alosinae, and Dorosomatinae could not be supported by any osteological features. Although Grande's classification was established using non-cladistic methods and did not provide fully resolved ingroup relationships within Clupeomorpha, it provided the basis for subsequent phylogenetic studies of the group.

The new stage in clupeomorph systematics started with the development of molecular techniques that allowed large scale phylogenetic analyses of the group (Li and Orti, 2007; Lavoué et al., 2007, 2013; Wilson et al., 2008; Bloom and Lovejoy, 2012, 2014; Bloom et al., 2018). Li and Orti (2007) conducted one of the first molecular assessments of the clupeomorph interrelationships using two mitochondrial (12S and 16S) and two nuclear (RAG1 and RAG2) markers and 37 ingroup taxa. Their analysis recovered a non-monophyletic Clupeiformes with *Denticeps* placed as sister taxon to the outgroup ostariophysan taxa. Within Clupeoidei, Engraulidae and Pristigasteridae formed monophyletic groups, however Clupeidae formed a polyphyletic assemblage with Pristigasteridae nested within the clupeid group and Chirocentridae as the sister group to a clade containing dussumieriin taxa. Around the same time, Lavoué et al. (2007) published their phylogenetic study of clupeiform fishes based on complete mitochondrial genome sequences. Although, their taxonomic sampling was smaller than that of Li and Orti (2007), Lavoué et al. (2007) recovered a monophyletic Clupeiformes with *Denticeps clupeoides* as sister group to the remaining clupeiform taxa, the monophyletic Engraulidae was recovered as the basalmost clupeoid lineage, Pristigasteridae formed a monophyletic group, Chirocentridae was placed within the polyphyletic Clupeidae, and the clupeid subfamilies Alosinae, Clupeinae, and Dorosomatinae did not form monophyletic groups. Further molecular

phylogenetic studies (Wilson et al., 2008; Bloom and Lovejoy, 2012, 2014; Lavoué et al., 2013; Bloom et al., 2018) substantially increased taxonomic sampling of ingroup taxa and produced well-supported and, for the large part, consistent results that are best summarized in Lavoué et al. (2014): 1) Clupeiformes are a monophyletic group comprising suborders Denticipitoidei and Clupeiodei; 2) Engraulidae and Pristigasteridae are monophyletic; and 3) Clupeidae and all clupeid subfamilies (Alosinae, Clupeinae, Dorosomatinae, Pellonulinae) *sensu* Grande, 1985 are non-monophyletic assemblages. This new evidence served as the basis for revising the clupeomorph classification established by Grande (1985).

Great advances in molecular techniques, however, have shifted focus of the neontological systematics of Clupeomorpha away from anatomical examination. Few recent studies are focused on morphological characters to establish evolutionary relationships among major living clupeomorph lineages (Di Dario, 2002, 2009; Miyashita, 2010). Additionally, most molecular studies overlook fossil taxa except when fossils are used to date nodes in time-calibrated phylogenies (Lavoué et al., 2013). This has caused a breach between the systematic research of the extinct order †Ellimmichthyiformes and extant Clupeiformes, making it difficult to assess ingroup relationships within the total group Clupeomorpha. In recent phylogenetic studies of †Ellimmichthyiformes (Murray and Wilson, 2013; Vernygora et al., 2016; Murray et al., 2016; Figueiredo and Ribeiro, 2017; Marramà and Carnevale, 2017; Marramà et al., 2019), the order has consistently been recovered as monophyletic and supported by the medioparietal condition of the skull roof (parietals contacting each other in the midline), presence of the basipterygoid process of the parasphenoid, ‘beryciform’ foramen in the anterior ceratohyal, ornamentation of the skull roof, presence of three epurals, and presence of a predorsal series of scutes. These

studies, however, include very few clupeiform taxa as outgroups and do not provide a good assessment of ingroup relationships within Clupeomorpha.

Overall, the current classification of Clupeomorpha can be summarized as following (Lavoué et al., 2014, Nelson et al., 2016, Fricke et al., 2020):

Superorder Clupeomorpha

Order †Ellimmichthyiformes

Family †Armigatidae

Family †Sorbinichthyidae

Family †Paraclupeidae

Order Clupeiformes

Suborder Denticipitoidei

Family Denticipitidae

Suborder Clupeoidei

Family Engraulidae

Subfamily Engraulinae

Subfamily Coiliinae

Family Pristigasteridae

Subfamily Pristigasterinae

Subfamily Pelloninae

Family Chirocentridae

Family Spratelloididae

Family Dussumieriidae

Family Clupeidae

Subfamily Clupeinae

Subfamily Alosinae

Subfamily Ehiravinae

Subfamily Dorosomatinae

In this chapter, I perform the first comprehensive cladistic analysis of Clupeomorpha including representatives of all major extant and extinct lineages. The data matrix constructed for this study integrates morphological and molecular data, with the morphological character list thoroughly revised and updated based on previously published studies and personal observations of specimens. I analyse the newly constructed data set using multiple phylogenetic inference methods (equal and implied weights maximum parsimony, Bayesian inference) to address the following questions:

- Do morphological and molecular data produce congruent phylogenetic hypotheses for Clupeomorpha?
- What are the evolutionary relationships among the major clupeomorph lineages?
- Is the current classification of Clupeomorpha supported by combined evidence analysis?
- What morphological traits can be used to delimit major clupeomorph lineages?
- What are the times of divergence of the major clupeomorph lineages?

6.2 Materials and methods

6.2.1 Data set construction

Taxonomic sampling — A total of 107 ingroup species was sampled for the study including members of all of major traditionally recognized extinct and extant clupeomorph lineages (e.g., families Denticipitidae, Engraulidae, Chirocentridae, Pristigasteridae, Dussumieridae, Clupeidae, Sundasalangidae, †Paraclupeidae, †Armigatidae, †Scutatuspinosidae, †Sorbinichthyidae). This sampling represents approximately 20 percent of the total number of living and fossil clupeomorph species. I selected ingroup taxa to represent the taxonomic diversity and geographic range of each group. Outgroup taxa included three teleost taxa, a euteleost *Esox lucius*, and *Chanos chanos* and *Cyprinus carpio*, both members of Ostariophysii, which is the sister group to Clupeomorpha (Hughes et al., 2018). Complete lists of extant and fossil taxa used in the study are provided in Appendices 6-1 and 6-2.

Morphological data set — The morphological character list was constructed based on an unpublished PhD thesis (Sato, 1994) with character revisions and additions from previous phylogenetic studies of clupeomorph and basal teleost fishes (e.g., Grande, 1985; Whitehead, 1985; Alvarado-Ortega et al., 2008; Murray and Wilson, 2013; Murray et al., 2016) as well as personal examination of fluid-preserved and cleared-and-stained specimens, skeletal preparations, and micro-computed tomography (μ CT) data. For taxa that were not available in museum collections I had access to, I scored morphological characters based on original descriptions of the species and other complimentary literature. Taxonomic sampling for the morphological data set consisted of 104 taxa, including 3 outgroup and 101 ingroup species (60 extant and 41 fossil species). The morphological character list includes 175 characters (Appendix 6-3). Most characters (n=125) are coded as binary and 50 characters are coded as multistate. In

all analyses, except for implied weighting parsimony, characters were treated as equally weighted and unordered. Polarity of character states (e.g., state “0” is primitive vs state “1” is derived) were not defined prior to analysis; polarity of character states was determined during optimization of phylogenetic inference analysis. I scored all taxa on adult specimens to exclude ontogenetic effects.

Micro-computed tomography (μ CT-scans) — scans of specimens were performed at Friday Harbor Laboratories, University of Washington, and were either scanned personally or provided by Dr. Adam Summers. Specimens were scanned using a SkyScan 1173 scanner (Bruker, Billerica, MA, USA) at 60 kV and 133 μ A on a 2240 x 2240 pixel detector. All scans, along with the individual scan specifications, are available on the Virtual Natural History Museum website at vnhm.de.

Molecular data set — The molecular data set comprised 15 molecular markers (13 mitochondrial and 2 nuclear loci) for 63 extant taxa used in the study including the three outgroup species. For the nine ingroup taxa, *Anchoviella lepidentostole*, *Setipinna tenuifilis*, *Stolephorus indicus*, *Ilisha megaloptera*, *Ilisha melastoma*, *Pellona ditchela*, *Sardinops sagax*, *Sardinella lemuru*, and *Herklotsichthys punctatus*, for which molecular data were not available, I used sequences of congeneric species, *Anchoviella* sp., *S. taty*, *S. insularis*, *I. africana*, *I. elongata*, *P. flavipinnis*, *S. melanostictus*, *S. fijiensis*, and *H. dispilonotus*, respectively. Sequence data were downloaded from the GenBank database and aligned in MAFFT 7.245 (Katoh et al., 2002; Katoh and Standley, 2013) using the G-INS-i global alignment with iterative refinement strategy. I aligned sequences for each locus individually, trimmed them to the same length, and concatenated aligned loci into a single data set. Finally, I visually inspected the aligned data set in Mesquite 3.61 (Maddison and Maddison, 2015) and performed manual adjustment of the

alignment. The total length of the aligned molecular data set is 13967 base pairs. A complete list of molecular markers and accession numbers is provided in Appendix 6-4. To find the best partitioning scheme and model of evolution for the molecular data set analysis, I used Partition Finder 2 (Lanfear et al., 2017). Configuration settings used in the Partition Finder 2 analysis are listed in Appendix 6-5.

Finally, I combined both data sets into a single data matrix with two partitions (morphological and molecular partitions). The combined data set includes 110 taxa (three outgroups and 107 ingroups) and 14142 characters (175 morphological and 13967 molecular) and is available online at <https://dataverse.harvard.edu/api/access/datafile/3874690>.

6.2.2 Phylogenetic analyses

I analysed each data set (morphological and molecular) individually first, to examine phylogenetic signal for each data type. Morphological data was analysed using maximum parsimony (equal and implied weights), maximum likelihood, and Bayesian inference methods; molecular data was analysed using probabilistic methods of phylogenetic inference only (maximum likelihood and Bayesian inference). Finally, I used the partitioned concatenated data set to perform a combined evidence time-calibrated analysis. A summary of the analytical pipeline is shown in Figure 6-2.

Equal weights maximum parsimony — I analyzed the data sets in TNT v 1.5 (Goloboff et al., 2008) using the New Technology Search algorithm. I performed two rounds of tree search with the following modifications to default settings: (a) First round: 1000 initial Wagner trees obtained by the randomized addition sequences (RAS) algorithm and followed by Sectorial Search – random sector selection (RSS) with lower bound equal to 5 and upper bound equal to

45 taxa; constraint-based sector selection (CSS) for 100; Ratchet for 100 iterations; Drift for 100 rounds with a maximum fit of 10; and Tree Fusing for 100 rounds; all trees obtained after the first round of searching were saved and used as initial trees for the second round. In the second round, I performed 1000 iterations of Ratchet and Tree Fusing, using the initial trees saved from the first round of searches. I used the trees resulting from the second round of tree search to construct a 50% majority-rule consensus trees. Tree metrics such as tree length, consistency and retention indices were calculated in Mesquite 3.61 (Maddison and Maddison, 2015). Bootstrap and Bremer support values for the consensus topology were calculated in TNT v.1.5.

Implied weighting maximum parsimony — Analyses were performed in TNT v.1.5 (Goloboff et al., 2008) using the New Technology search settings as described above for the EWMP analysis. To investigate effects of different weight functions on the resulting topology, I performed a sensitivity analysis by testing different weight function values (K values). The starting value of the weight function was set to K=1; I repeated the analysis using K=3, 5, 7, 9, 11, 13, 15, 20, and 30 (Fig. 6-3). The best fit trees were saved and used to construct a 50% majority-rule consensus tree.

Maximum likelihood — I analysed data sets in IQ-TREE v. 1.6.11 web server (Trifinopoulos et al., 2016). For the morphological data set, I used the Mk+G4+ASC model, which is a time-reversible model that assumes equal rates of character change and equal state frequencies (Lewis, 2001), but allows rate heterogeneity across sites modeled using a gamma distribution with four discrete categories (Yang, 1995; Soubrier et al., 2012). Ascertainment bias correction (ASC option) corrects for any artificial increase in branch lengths due to the use of only variable characters in morphological data sets, which is known to cause an overestimation of divergence in phylogenetic analyses (Leache et al., 2015). The molecular data set was

analysed under the GTR+G4 model which specifies a general time reversible model with across-site rate heterogeneity sampled from a gamma distribution with four discrete categories. I conducted 1000 ultrafast bootstrap replicates (Minh et al., 2013) to calculate clade support values and saved all bootstrap replicate trees to construct a 50% majority rule consensus trees.

Bayesian inference — Non-clock Bayesian inference analysis was conducted in MrBayes v.3.2.6 (Ronquist et al., 2012) on the Graham computer clusters available through Compute Canada. The relative burn-in fraction was set to 50%. Chains were sampled every 1000 generations. The temperature parameter was set to 0.015 and 0.075 for morphological and molecular data sets, respectively. The temperature parameter determines the relative decrease (‘melting’) of the parameter landscape on a parallel ‘heated’ chain compared to the actual parameter landscape; this feature helps the search algorithm avoid being stuck at local optima in the parameter space. Molecular data was analysed under the GTR+G4 substitution model. Morphological data was analysed under a simple Mk model with ascertainment bias correction (Mkv, equivalent to Mk+ASC model in maximum likelihood analysis). Rate heterogeneity among characters was sampled from a gamma distribution with four discrete categories. The number of generations was set to 100 million generations. Four independent runs were performed for each analysis with four chains per run. Convergence of MCMCMC runs was assessed in Tracer v. 1.7 (Rambaut et al., 2018) with effective sample size (ESS) greater than 200, trace plots showing stationary phase for all independent runs, average standard deviation of split frequencies between the runs less than 0.01, and effective mixing between chains (acceptance rates 10–70%). The 50% majority-rule consensus trees were calculated after the burn-in component was removed.

Time-calibrated Bayesian inference analysis—Divergence times of major lineages within Clupeomorpha were estimated in BEAST v. 2.6.1 (Bouckaert et al., 2014) under the Fossilized Birth-Death model. I used a relaxed log-normal molecular clock and substitution models consistent with model specifications used in non-clock analysis conducted in MrBayes (GTR+G4 for molecular partition and Mkv+G4 for morphological data partition). I used the tip dating approach which accounts for uncertainty in placement of fossil taxa and avoids bound estimates for node-based age calibrations. Tip ages used for fossil taxa were sampled from uniform prior distributions defined as stratigraphic ranges of occurrence of fossils (Appendix 6-2). For extant taxa, tip ages were set to 0. I set a monophyly constraint on ingroup taxa because monophyly was confirmed in all previous analyses and enforcing this constraint helps reduce the total time of computationally intensive time-calibrated analysis. I conducted four independent MCMC analyses for 2×10^8 generations with trees sampled every 1000 generations. Convergence and stationarity of independent runs was assessed in Tracer v. 1.7, checking for effective sample size for each parameter to be greater than 200 and each run reaching a stationary phase when the likelihood scores of sampled topologies and model parameters were concentrated around the same mean value without any trend in decrease or increase of that mean value. Results of the time-calibrated analysis were summarized in a maximum clade credibility tree constructed using the TreeAnnotator v. 2.6.1 application available with BEAST2 software. This type of summary tree selects the single topology with the highest product of clade probability values among all trees in the posterior sample. The selected topology is then annotated with ranges for sampled parameter values (e.g., divergence dates, evolutionary rates, branch lengths, clade support values) from the posterior sample.

6.3 Results

6.3.1 Morphological data set analyses

Equal weights maximum parsimony—the parsimony analysis recovered 462 most parsimonious trees (MPTs) of 816 steps. The 50% majority rule consensus tree (Fig. 6-4) has a length of 840 steps with consistency (CI) and retention (RI) indices equal to 0.288 and 0.716, respectively.

In the consensus tree, the two outgroup ostariophysan taxa (*Chanos chanos* and *Cyprinus carpio*) form a monophyletic group with a sister-group relationship to ingroup clupeomorph taxa. Within the monophyletic Clupeomorpha, relationships are somewhat unexpected with the extinct order †Ellimmichthyiformes nested within Clupeiformes as sister group to Clupeidae *sensu* Grande, 1985. At the family level, most currently recognized lineages are recovered as monophyletic. Denticipitidae are the clupeomorph clade leading to the monophyletic and well-supported families Engraulidae, Pristigasteridae, and Chirocentridae. Clupeidae are recovered as paraphyletic with the round herring family Dussumieriidae nested within a clade that includes two dorosomatid species, *Limnothrissa miodon* and *Stolothrissa tanganyicae*. Of the four commonly recognized clupeid subfamilies (Clupeinae, Alosinae, Dorosomatinae, and Ehiravinae), only Alosinae are recovered as a monophyletic group, comprising genera *Alosa*, *Brevoortia*, *Sardina*, and *Sardinops*. Polyphyletic Clupeinae are split into two groups – a clade comprising type genus *Clupea* and genus *Sprattus*, and a clade that includes genera *Potamalosa* and *Hyperlophus* with a sister-group relationship to fossil taxon †*Knightia eocaena*. Within polyphyletic Dorosomatinae, two major clades are recovered – a clade containing type genus *Dorosoma* along with *Escualosa*, *Konosirus*, *Anodontostoma*, *Nematalosa*, and †*Chasmoclupea*, and a clade comprising genera *Amblygaster*, *Opisthonema*, *Sardinella*, *Harengula*, *Lile*, and

Herklotsichthys. Members of the extinct lineage traditionally defined as order

†Ellimmichthyiformes are grouped in a sister clade to Clupeidae *sensu* Grande, 1985. This clade also includes two fossil taxa previously identified as clupeiform species of uncertain affinities as its stem members, †*Santanaclupea silvasantosi* and †*Pseudoellimma gallae*. Interrelationships of †ellimmichthyiform taxa are poorly resolved in the tree. The three major clades recovered within the †ellimmichthyiform group comprise the basalmost clade comprised of genera †*Armigatus*, †*Codoichthys*, and †*Ornategulum*, a monogeneric group that includes two species of genus †*Thorectichthys*, and a large polytomy consisting of members of †Paraclupeidae, monogeneric †Sorbinichthyidae, and genera †*Diplomystus* and †*Horseshoeichthys*.

Implied weighting maximum parsimony—Sensitivity analysis using different implied weighting schemes revealed several major patterns of interrelationships within Clupeomorpha (Fig. 6-5). Regardless of the weighting scheme, (1) Denticipitidae were consistently recovered as the basalmost clupeomorph lineage; (2) †ellimmichthyiform taxa formed a monophyletic group with two stem fossil taxa of uncertain affinities, †*Santanaclupea silvasantosi* and †*Pseudoellimma gallae*; and (3) Engraulidae, Pristigasteridae, and Chirocentridae are monophyletic groups. Interrelationships among the major monophyletic groups, and to a lesser degree within each group, varied substantially depending on the strength of the weighting scheme used. Under strong homoplasy downweighting schemes, $K = 1-3$, (1) pristigasterids were recovered as the sister group to an extinct group including †ellimmichthyiforms, †*Santanaclupea silvasantosi*, and †*Pseudoellimma gallae*; (2) Dussumieriidae are sister to a clade comprising monophyletic Chirocentridae and Engraulidae; (3) clade (Dussumieriidae (Chirocentridae, Engraulidae)) is sister to Clupeidae (Fig. 6-5 A). As the weighting scheme was relaxed to $K=5-7$, ingroup relationships approached the traditional classification scheme with the

dichotomy between the fossil †ellimmichthyiform lineage and Clupecoidei. Within the clupeoid clade, Pristigasteridae was recovered as the basalmost lineage followed by a larger clade (Dorosomatidae (Chirocentridae, Engraulidae)) that is sister to the Clupeidae (Fig. 6-5 B). Relaxing the weighting function to $K = 9-11$, effected ingroup relationships within the two major clades while maintaining the overall dichotomy between the extinct †ellimmichthyiform lineage and the clupeoid group (Fig. 6-5 C). At $K=13, 15$, and 20 , Engraulidae were recovered more basally within Clupeomorpha as a lineage diverging after the basalmost Denticipitidae. Pristigasteridae were the next lineage to diverge. Notably, Chirocentridae was recovered at the base of the fossil clade comprising †ellimmichthyiforms, †*Santanaclupea silvasantosi*, and †*Pseudoellimma gallae* (Fig. 6-5 D). At $K=30$, results of IWMP analysis converged on EWMP results, indicating that the concavity function reached a nearly linear form with the weighting function resembling the equal-weights scheme (Fig. 6-5 E).

Maximum likelihood—The bootstrap majority-rule consensus tree offers a poorly resolved topology that generally agrees with the results of parsimony analyses (Fig. 6-6). Denticipitidae are the basalmost clupeomorph lineage followed by a major polytomy comprised of monophyletic Pristigasteridae, Engraulidae, Chirocentridae, unresolved polytomic Clupeidae and Dussumieriidae, and a large clade grouping all fossil †ellimmichthyiforms and †*Santanaclupea silvasantosi*, and †*Pseudoellimma gallae*.

Bayesian inference—In the Bayesian inference consensus tree, ingroup relationships within Clupeomorpha follow the general pattern recovered by maximum likelihood analysis (Fig. 6-7). Denticipitidae are the earliest diverging clupeomorph lineage followed by a dichotomy between a fossil clade comprising all †ellimmichthyiform taxa plus †*Santanaclupea silvasantosi* and †*Pseudoellimma gallae*, and a large polytomic clade that includes monophyletic

Pristigasteridae, Chirocentridae as sister to Engraulidae, and a polyphyletic Clupeidae.

Relationships within Dussumieriidae are unresolved, placing this group in a polytomy with the clade (Chirocentridae + Engraulidae). Consistent with the results of all other analyses, Alosinae is the only monophyletic clupeid subfamily. The fossil clupeiform species †*Chasmoclupea aegyptica* is nested with a dorosomatin clade composed of genera *Escualosa*, *Dorosoma*, *Konosirus*, *Anodontostoma*, and *Nematalosa*.

6.3.2 Molecular data set analyses

Maximum likelihood—In the bootstrap consensus tree (Fig. 6-8), Denticipitidae represented by the extant genus *Denticeps* is the basalmost clupeomorph lineage followed by a well-supported monophyletic Engraulidae. Pristigasteridae are the next clade to diverge within Clupeomorpha; it is the sister group to a large polytomy comprising Chirocentridae as sister to the dussumieriid taxa (*Dussumieria acuta* (*Etrumeus sadina*, *Etrumeus micropus*)) and Clupeidae with unresolved relationships between monophyletic Clupeinae and a large clade including monophyletic subfamilies Alosinae, Dorosomatinae, and Ehiravinae.

Bayesian inference—the analysis recovered a topology (Fig. 6-9) that is largely consistent with the results of ML analysis with the important exception that Chirocentridae are the sister group to Pristigasteridae, and Dussumieriidae are a monophyletic group nested within a paraphyletic Clupeidae. All four clupeid subfamilies are recovered as monophyletic.

6.3.3 Combined data set time-calibrated analysis

The maximum clade credibility topology recovered in combined evidence analysis is overall consistent with the results of maximum likelihood, Bayesian inference, and implied

weighting parsimony under K=5–7 analyses –Denticipitidae are the basalmost clupeomorph group, the sister groups †Ellimmichthyiformes and Clupeiformes are monophyletic, the clupeiform families Pristigasteridae, Engraulidae, Chirocentridae, and Dussumieriidae are natural groups, but Clupeidae is a polyphyletic assemblage (Fig. 6-10). The estimated age of the total group Clupeomorpha is around 214 million years (mya) which is younger than the estimate provided for the age of Clupeiformes in the recent time calibrated analysis by Bloom and Lovejoy (2014), ~ 225 mya, but older than that provided by Lavoué et al. (2013), ~ 140 mya. The chronogram showing 95% age credibility intervals is shown in Appendix 6-6.

6.4 Discussion

The higher-level relationships of clupeomorph fishes recovered in the present phylogenetic study reveal common patterns across different data types and multiple methods of phylogenetic inference. These patterns are summarized based on the results of combined evidence analysis (Fig. 6-10), with adjustments based on alternative topologies recovered, and are discussed below. It is important to note that obtaining a fully resolved summary tree for time calibrated analyses is necessary for meaningful annotation of divergence times because annotation of polytomic nodes results in averaging across ages of all collapsed internal nodes, which does not have evolutionary meaning. However, a potential disadvantage of using a maximum clade credibility tree is that a single fully resolved topology does not summarize uncertainty in the inference of phylogenetic relationships and may include weakly supported and incorrectly inferred clades. This artifact can be assessed with posterior probability support values annotated for each node in the maximum clade credibility tree and by comparing the maximum clade credibility tree topology to results

recovered in other analyses. A revised classification of the Clupeomorpha based on synthesis of the phylogenetic results is provided at the end of the Discussion section.

Clupeomorpha were recovered as monophyletic in all analyses regardless of the data type and method of phylogenetic inference used. Morphological support for the group is consistent with the diagnostic features proposed by Greenwood et al. (1966) and adopted with some modifications in subsequent systematic works on clupeomorph fishes (e.g., Nelson, 1973; Grande, 1982, 1985; Whitehead, 1985; Sato, 1994; Di Dario, 2004). These traits include the otophysic connection between the swim bladder and inner ear involving intercranial incapsulation of the swim bladder diverticulum vesicles in the prootic and often in the pterotic bullae, presence of the pre-epiotic fossa, at least one abdominal scute, a predorsal series of scutes, and fusion between the second hypural and the first ural centrum (Grande, 1985; Nelson et al., 2016). The ingroup relationships, however, differ considerably from the traditional classification of the Clupeomorpha that divides it into two major groups –extinct †Ellimmichthyiformes and the Clupeiformes that include all living members of the group. Instead, the results of combined evidence analysis and analyses of morphological data that include fossil taxa, indicate that Denticipitidae are the basalmost clupeomorph lineage and †Ellimmichthyiformes are a sister group to Clupeiformes. This arrangement of ingroup lineages indicate that the total group Clupeomorpha is also supported by the presence of the recessus lateralis (22:1) consistent with Greenwood et al. (1966) but later used as a diagnostic feature restricted to order Clupeiformes (Grande, 1985). The recessus lateralis is a complex intracranial feature that involves convergence of cranial sensory canals in a common chamber in the otic region of the skull (Grande, 1985; Di Dario, 2004). This trait has often been treated in systematic studies as a simple presence/absence character; however, a thorough assessment of this

anatomical feature by Di Dario (2004) revealed a variety of morphological patterns associated with the number and location of sensory canal openings into the intercranial chamber. Therefore, omitting these intricacies of the structure of the recessus lateralis may lead to loss of important phylogenetic information. Another challenge of using the recesses lateralis as a character in phylogenetic analyses is associated with scoring presence or absence of the recessus lateralis in fossil taxa. Because most clupeomorph fossils are extremely laterally compressed with crushed skulls, observation of intracranial features including the recessus lateralis chamber is nearly impossible in most specimens. In practice, this feature is often inferred based on osteological proxies such as the degree of dermosphenotic development, and proximity of preopercular and infraorbital sensory canals (Patterson, 1967; Grande, 1985; Di Dario, 2004). Using the degree of the dermosphenotic development is, however, problematic not only because this is a relative measure that can be difficult to compare across taxa but also because of ambiguity in identifying this bone. Grande (1985) noted that the last bone in the infraorbital series is designated as the dermosphenotic; usually it is either the fifth or sixth infraorbital bone. However, homology of the last infraorbital bone in the series may be difficult to establish due to differences in patterns of reduction of the infraorbital series. For example, reduction of number of infraorbital bones due to fusion between the third and fourth infraorbital bones has been reported in *Potamalosa richmondia* (Grande, 1985) and *Hyperlophus vittatus* (Yabumoto and Uyeno, 1982); therefore, the last bone in the series is the original infraorbital six. It is important to note that Sato (1994) reported a complete series of six infraorbital bones to be present in the same species (*P. richmondia* and *H. vittatus*) which indicates possible plasticity of this character. In members of the genera *Jenkinsia* and *Spratelloides*, reduction of the series occurs as a result of loss of the sixth infraorbital (Grande, 1985; Sato, 1994); therefore, the last bone in the series in these

species is the fifth infraorbital that is often larger than the sixth infraorbital of other taxa. Establishing the homology of the dermosphenotic in taxa with a reduced number of infraorbital bones, therefore, requires a careful investigation of the ontogenetic series. In fossil specimens, infraorbital bones are often crushed and poorly preserved, which further complicates assessment of the number and degree of development of the dermosphenotic. It is therefore recommended to score presence or absence of the recessus lateralis as ‘unknown’ in taxa with no direct evidence of the condition, as has been done by some authors.

6.4.1 Ingroup relationships

Combined evidence analysis recovered three major lineages within Clupeomorpha – basalmost Denticipitoidei and two sister clades, †Ellimmichthyiformes and Clupeoidei. This pattern of ingroup relationships is consistent across all analyses except for equal-weights maximum parsimony (Fig. 6-4) and implied-weights parsimony under very stringent ($K = 1-3$) or relaxed ($K = 13-30$) weighting schemes (Fig. 6-5 A, D, E). This difference in results can be attributed to the morphological data set having an overall relatively high level of homoplasy ($CI = 0.288$). In such cases, stringent weighting against homoplasy results in the tree search being driven primarily by very few characters with high consistency indices, which creates bias in determining the best fitting topology. On the other hand, relaxed weighting schemes do not penalize highly homoplastic characters strongly enough and lead to recovering clades supported by highly homoplastic traits. It is therefore important to assess levels of homoplasy in data sets and perform sensitivity analysis in order to determine the best weighting scheme for any given data matrix.

6.4.1.1 Denticipitoidei

This lineage consists of two monotypic genera, *Denticeps* and †*Palaeodenticeps*.

Greenwood et al. (1966) recognized *Denticeps* as the most primitive living clupeomorph based on the type of caudal skeleton that is more similar to that of elopiform fishes and larval clupeomorphs than to the more consolidated caudal skeleton of other adult clupeomorphs, suggesting its ancestral condition. Among other morphological features suggesting a basal position of Denticipitoidei, Greenwood et al. (1966) mentioned the structure of the syncranium and jaws as well as presence of a complete lateral line system. While a simplified caudal skeleton and lack of ossification of some cranial elements in *Denticeps* may suggest a possible developmental truncation in this small clupeomorph species with an average total length up to 5 cm (Daget, 1984; Nelson et al., 2016; but Teugels (2003) reported a maximum total length of 15 cm), presence of a complete lateral line system in *Denticeps* and †*Palaeodenticeps* is not observed at any ontogenetic stage in any other living clupeomorph including some miniature paedomorphic taxa (e.g., *Sundasalanx*, *Amazonsprattus*). A complete lateral line is present in some †ellimmichthyiforms (e.g., †*Armigatus brevissimus*, †*Diplomystus dentatus*, †*Ornategulum sardinioides*) and basal clupeiforms (†*Santanachupea silvasantosi*) with the trunk scales preserved; this indicates that presence of a complete lateral line system is a retention of the plesiomorphic condition present in other teleostean lineages. In the present study, monophyly of Denticipitoidei is supported by contact between the wings of the lateral ethmoid and the anterior portion of mesethmoid ([6:1], this condition is also present in Engraulidae and Chirocentridae), presence of a short vomer that is wider than it is long ([10:0], this condition is also present in Chirocentridae and some engraulids), presence of greatly expanded nasals [41:3] and denticles covering dermal bones of the skull [173:1]. The basalmost placement of Denticipitoidei is further

supported by the results of molecular data set analyses that consistently recover this group as the earliest diverging lineage among clupeomorphs and, more importantly, indicate a substantial degree of genetic divergence from Clupeoidei. The long internal branch connecting the two extant groups may be an artifact of the missing clade, whereas in the combined evidence analysis the extinct †ellimmichthyiform clade is recovered as a sister group to Clupeoidei, resulting in more proportionate evolutionary distances among the groups. It is important to point out, however, that the basalmost placement of Denticipitoidei creates an extremely long ghost lineage going from the origin of Clupeomorpha (~ 214 mya) to the origin of the clade (*Denticeps*, †*Palaeodenticeps*) dated at ~ 68 mya (Fig. 6-10, Appendix 6-6). Without additional information on the fossil record of the stem denticipitoid taxa, it is not clear whether this ghost lineage is an artifact of an incorrect phylogenetic inference or is, in fact, an evolutionary lineage with a long history for which we are missing fossil evidence.

6.4.1.2 †Ellimmichthyiformes

Monophyly of †Ellimmichthyiformes was supported in all analyses. This well-supported group was established by Grande (1985) to include fossil non-clupeiform clupeomorphs characterized by the presence of dorsal scutes with expanded subrectangular lateral wings. The composition and diagnosis of the group have changed considerably over the years. Currently, †Ellimmichthyiformes include about 40 species in 20 genera with their temporal range spanning the Early Cretaceous through to the middle Eocene (Alvarado-Ortega and Melgarejo-Damian, 2017; Figueiredo and Ribeiro, 2017; Marramà et al., 2019). In recent comprehensive analyses of the group (Murray and Wilson, 2013; Vernygora and Murray, 2016; Marramà and Carnevale, 2017; Marramà et al., 2019) the monophyly of †Ellimmichthyiformes was supported by the

medioparietal condition of the skull roof, ornamentation of the skull roof bones, presence of the ‘basipterygoid’ process of the parasphenoid, ‘beryciform’ foramen in the anterior ceratohyal, lack of the recess lateralis, fusion between the first preural centrum and parhypural, presence of three epurals, and a predorsal series of scutes.

In combined evidence analysis, the monophyly of †Ellimmichthyiformes is supported by the lack of the recessus lateralis [22:0], medioparietal condition of the skull roof [40:1], fusion between the parhypural and first ural centrum [127:1], precaudal vertebrae ornamented with three lateral pits [153:1], and presence of the dorsal series of scutes [154:1]. Contrary to previous results, presence of the ‘beryciform’ foramen within the anterior ceratohyal was not recovered as a diagnostic feature of the †Ellimmichthyiformes. This feature, however, is vaguely defined when applied to clupeomorph fishes (Fig. 6-11). McAllister (1968) defined the beryciform foramen as a perforation in the dorsal half of the anterior ceratohyal that also enters the groove for the hyoidean artery along the lateral wall of the anterior ceratohyal. Presence of such a foramen has been considered to be a diagnostic feature of †Ellimmichthyiformes (Grande, 1985) but assessment of the foramen in fossil taxa often does not involve its relationship to the hyoidean duct that, in clupeomorphs, can be bone-enclosed and impossible to locate in fossil specimens (pers. observ.). This feature, therefore, has been consistently scored as present in fossil specimens if an opening in the anterior ceratohyal is present. In his revision of the extant clupeoid interrelationships, Sato (1994) noted that a large opening observed in some species of the genera *Alosa*, *Clupea*, *Harengula*, *Hyperlophus*, *Lile*, *Opisthonema*, *Sardinella*, *Sardinops*, and *Sprattus* resembles the beryciform condition, but is “one step away from it” because the foramen does not perforate both walls of the hyoidean artery duct; however, in some taxa (*Alosa*, *Clupea*, *Hyperlophus*, *Sardinops*, *Sprattus*) with a partially reduced lateral wall of the duct, the

opening perforates both sides and, according to Sato (1994), differs from the beryciform condition only in the degree of the perforation. Grande et al. (2013) noted that the beryciform foramen is present in several groups of teleost fishes including beryciforms, basal clupeomorphs, †ctenothrissiforms, and zeiforms, but the degree of development of the foramen varies among and within taxa. In some zeiforms, the foramen is represented only by a deep groove and shows evidence of ontogenetic reduction (Grande et al., 2013). This variation in degree of beryciform foramen development in other groups suggests that the condition observed in some clupeoid taxa is comparable to that found in zeiforms and beryciforms, although, establishing homology of these conditions requires a more detailed examination of the developmental series. More important for the aims of the present study is to establish a consistent scoring scheme for this character in both living and fossil taxa. As noted above, presence of an opening within the anterior ceratohyal is traditionally scored as presence of the ‘beryciform’ foramen in fossil clupeomorph taxa. The use of this term with quotation marks around ‘beryciform’ suggests that the observed condition may not, in fact, be homologous to that found in beryciforms, but rather refers to the overall similarity of the presence of an opening in the anterior ceratohyal. To score this character consistently across all taxa in the present study, I scored it as present in taxa that showed a complete perforation of the ceratohyal regardless of the size of the opening.

Within †Ellimmichthyiformes, two major lineages are identified, †Armigatoidei and †Ellimmichthyoidei. †Armigatoidei include a single family †Armigatidae which is subdivided into two clades. The first group comprises the genera †*Armigatus*, †*Codoichthys*, and †*Ornategulum* and is weakly supported by homoplastic characters including the lack of premaxillary dentition [64:0] and absence of prominent median spines on the postpelvic abdominal scutes [166:0]. Two of these genera (†*Armigatus* and †*Codoichthys*) were recovered

as sister taxa in the phylogenetic analysis presented in Chapter 5; however, some previous studies (Figueiredo and Ribeiro, 2016; Marramà et al., 2019) recovered †*Codoichthys* as a member of a more derived †ellimmichthyiform group, †Paraclupeidae. These results, however, are difficult to compare because of considerable differences in taxonomic sampling and characters used to infer phylogenetic relationships in these studies.

Placement of †*Ornategulum sardinoides* as sister taxon to the genus †*Armigatus* is unexpected because most previous analyses have recovered this taxon as the most basal †ellimmichthyiform outside of †Armigatidae *sensu* Murray and Wilson (2013). Evolutionary relationships of †*Ornategulum sardinoides* have been controversial ever since the original description of the taxon by Forey (1973) who designated it *Clupeomorpha insertae sedis*. Subsequent studies showed that this taxon belongs within †Ellimmichthyiformes based on a number of diagnostic features including the medioparietal condition of the skull roof, parhypural fused to the first preural centrum, and heavy ornamentation of the skull roof bones. This fossil species, however, lacks any scute development, which makes classification of this taxon especially difficult.

The second group within †Armigatidae contains genera †*Diplomystus*, †*Horseshoeichthys*, †*Kwangoclupea*, †*Ranulfoichthys*, and †*Foreyclupea*, which are grouped into two clades. The two Early Cretaceous species, †*Ranulfoichthys dorsonudum* and †*Foreyclupea loonensis*, form a clade supported by the lack of the predorsal series of scutes [154:0]. The lack of predorsal scutes is rare among †ellimmichthyiforms that are commonly known as “double armoured herrings” for the presence of the dorsal series of scutes in addition to the abdominal series of scutes characteristic of clupeomorph fishes. The only other †ellimmichthyiform taxon that has no dorsal or abdominal scutes is the problematic species †*Ornategulum sardinoides* mentioned above.

Overall, the pattern of dorsal scute series development among clupeomorph fishes suggests that the loss of predorsal scutes is not unique to clupeoid taxa and happened multiple times in various clupeomorph lineages. The second clade is comprised of genera †*Horseshoeichthys*, †*Kwangoclupea*, and †*Diplomystus*, with the exception of †*D. solignaci*. Members of this group have a predorsal series of bones with at least one supraneural that is anteriorly inclined [121:1]. In species of the genus †*Diplomystus*, supraneurals form a peculiar fan-shaped structure [122:1].

The rest of †Ellimmichthyiformes are grouped in a clade similar in composition to suborder †Ellimmichthyoidei *sensu* Murray and Wilson, 2013. These are heavily armoured fishes with well-developed abdominal and predorsal series of scutes [168:1]. Derived members of this clade (†*Triplomystus* and †*Scutatoclupea*) have an additional third series of scutes behind the dorsal fin. The suborder is diagnosed by presence of abdominal scutes with prominent median spines ([162:1], but absent in members of the †Sorbinichthyidae) and a predorsal series of scutes of irregular size [163:1] with ornamentation on the surface of the scutes [164:1]. Interrelationships within †Paraclupeidae are not well resolved in majority rule consensus topologies and the clades recovered in the extended consensus tree have an overall low support. The family includes a monotypic subfamily †Scutatuspinosinae as its basalmost taxon. The type and only species of the subfamily, †*Scutatuspinosus itapagipensis*, is described from Neocomian deposits of Brazil (Silva Santos and Silva Corrêa, 1985; Figueiredo and Ribeiro, 2017) and is one of the oldest known clupeomorph taxa diagnosed by the presence of short [167:0] and broad [168:1] lateral wings of abdominal scutes and a slender almost straight maxillary blade [68:0].

The next †ellimmichthyiform clade includes genera †*Sorbinichthys* and †*Thorectichthys*. This group is weakly supported in combined evidence analysis and is collapsed in a polytomy in most consensus topologies of morphological data analyses. Members of genus †*Sorbinichthys*

are traditionally placed in a monogeneric family †Sorbinichthyidae that is diagnosed by a number of derived features including the posttemporal with a wedge-shaped dorsal process [104:2], size of the third hypural roughly equal to that of the fourth hypural [135:1], diastema between second and third hypurals formed by a large concavity in the third hypural [141:3], distinct subrhomboid dorsal scutes [159:2 and 161:2], and the presence of anteroposteriorly inclined lateral processes of posteriormost predorsal scutes [170:1]. While monophyly of †Sorbinichthyidae has been consistently supported in all phylogenetic studies of †Ellimmichthyiformes, the phylogenetic placement of the group has not been established with certainty. Several recent studies (Vernygora and Murray, 2016; Marramà and Carnevale, 2017; and Marramà et al., 2019) recovered †Sorbinichthyidae as a basal clade with a sister-group relationship to the rest of †ellimmichthyiforms excluding †*Ornategulum sardinioides*. Other studies suggest a more derived placement of †Sorbinichthyidae as sister to †Paraclupeidae (Murray and Wilson, 2013). In the maximum clade credibility tree recovered by combined evidence analysis, †*Sorbinichthys* is the sister group to another Late Cretaceous genus from Morocco, †*Thorectichthys*. The sister-group relationship of these two genera is supported by the vertically oriented ventral limb of the hyomandibula [58:1] and presence of radial striations on the opercle [63:1].

The remaining †ellimmichthyiforms are grouped into the family †Paraclupeidae that can be divided into †Ellimminae that comprise †*Ellimma branneri*, †*Ellimmichthys maceioensis*, and †*Tychoichthys dunveganensis*; †Paraclupeinae, containing †*Ellimmichthys longicostatum*, †*E. goodi*, †*Paraclupea chetungensis*, and †*Ezkutuberezi carmenae*; and †Triplomystinae that include derived †ellimmichthyiforms in genera †*Triplomystus*, †*Scutatoclupea*, †*Rhombichthys*, †*Gasteroclupea*, and two species †‘*Diplomystus*’ *solignaci* and †‘*Paraclupea*’ *seilacheri* with a

characteristic concave dorsal profile of the frontals [30:1] and a postdorsal series of scutes [169:1] in †*Triplomystus* and †*Scutatoclupea*.

6.4.1.3 Clupeiformes

The sister group to †Ellimmichthyiformes is a large clade containing all clupeoid species as well as two Early Cretaceous taxa from Brazil, †*Pseudoellimma gallae* and †*Santanaclupea silvasantosi*, as stem taxa. These fossil clupeiform species are among the oldest clupeomorphs and show a mosaic of primitive and derived features that indicate divergence between †ellimmichthiform and clupeiform lineages was already underway in the Barremian (129.4–125 mya). Similar to Denticipitoidei and some †ellimmichthyiforms, †*Santanaclupea silvasantosi* has a complete series of lateral line scales [174:1] (this condition is not known in †*Pseudoellimma* because of the preservation of the fossil material). However, like clupeoids, both stem taxa have a lateroparietal condition of the skull roof [40:0], presence of the recessus lateralis [22:1] (present in †*Santanaclupea* that has a well-preserved three-dimensional cranial neurocranium (Maisey, 1993), but this condition is inferred in †*Pseudoellimma* and therefore scored as unknown in the present data matrix), and articulation of ribs with parapophyses of abdominal vertebrae [146:0]. Additionally, †*Santanaclupea* has an autogenous parhypural not fused to the first preural centrum [127:0]; the alternate condition (parhypural fused with first preural centrum) is present in most clupeoids and is considered to be one of the diagnostic features of the group (Grande, 1982, 1985).

The interrelationships recovered within Clupeoidei are overall consistent with the current classification of the suborder. The major traditionally recognized clupeoid lineages include

monophyletic families Engraulidae, Pristigasteridae, Chirocentridae, and Dussumierriidae, and a polyphyletic Clupeidae *sensu* Grande, 1985.

Engraulidae are a diverse group that includes about 164 species of anchovies and relatives (Fricke et al., 2020). This group has received much attention in the literature, due primarily to the economic importance of these fishes in the global fisheries (Nelson, 1983, 1984, 1986; Grande and Nelson, 1985; Whitehead et al., 1988; Grant et al., 2010; Lavoué et al., 2010; Bloom and Lovejoy, 2012). Peruvian anchovetta, *Engraulis ringens*, is the most harvested species of fish in the world, with annual catches of over 3 million tonnes (FAO Fisheries Department, 2019). Consistent with the results of previous morphological and molecular studies (Grande, 1985; Grande and Nelson, 1985; Grant et al., 2010; Lavoué et al., 2010, 2013; Bloom et al., 2018), my phylogenetic analysis recovered Engraulidae as a well-supported monophyletic group subdivided into two subfamilies, Engraulinae and Coiliinae. This family is one of the most easily recognizable clupeomorph groups, characterized by peculiar morphology of the snout and jaws. Grande (1985) defined this family based on backward inclination of the suspensorium [57:0] and presence of a greatly expanded mesethmoid projecting forward beyond the anterior tip of the vomer [4:2]. In addition to these classical diagnostic features, engraulids possess several other morphological traits that distinguish them from other clupeoids – contact between lateral ethmoid wings and the anterior portion of the mesethmoid [6:1], contact between the first and third infraorbitals [44:1], presence of a shallow dentary with a low coronoid process [74:1], and a long, slender anterior ceratohyal [80:1]. The origin of the group is estimated around 80 mya (Fig. 6-10, Appendix 6-6) which is consistent with the molecular clock divergence time estimation of Lavoué et al., 2013; Bloom and Lovejoy, 2014). In a recent study, Capobianco et al. (2020) reported two putative stem engraulid taxa, †*Clupeopsis straeleni* Casier, 1946 and †*Monosmilus*

chureloides Capobianco et al., 2020, from the early and middle Eocene (~ 44 – 54 mya) of Belgium and Pakistan, respectively. These taxa differ from other known engraulids by their large size, cranial morphology, and peculiar fang-like dentition including a single vomerine fang. These early diverging taxa also do not show well-developed morphological features diagnostic of crown-group engraulids which led authors to suggest stem placement of these Eocene species. Capobianco et al. (2020) also indicated that the fang-like dentition indicative of piscivorous diets may represent an ancestral condition for Engraulidae and reinforce previously suggested sister-group relationships between engraulids and Chironcetridae, a family of piscivorous herrings with prominent fang-like teeth.

Pristigasteridae are another well-supported monophyletic group of clupeomorph fishes commonly known as ‘longfin herrings’ for their distinct long anal fins. This group comprises 38 species in nine genera that are of high local economic significance (Whitehead, 1985; Nelson et al., 2016; Fricke et al., 2020). Morphologically, pristigasterids are characterized by a single pair of lateral processes of the lateral ethmoid (also present in the Engraulidae) [7:0], posterior ceratohyal with a large and distinctly upturned dorsal process [84:1], supraneurals inclined anterodorsally or oriented vertically [121:1], lack of a notch in the third hypural [134:0], and lack of the interzygapophysal articulation [152:0]. Interrelationships within Pristigasteridae are not well studied. Grande (1985) recognized a suborder Pristigasteroidei with two families – Pellonulidae and Pristigasteridae; the latter was further subdivided into Pristigasterinae and Pristigasterinae division A. The former subdivision was adopted in later studies (Lavoué et al., 2014; Nelson et al., 2016) and recognized as two subfamilies, Pellonulinae and Pristigasterinae; however, according to a recent molecular analysis of Clupeiformes (Bloom et al., 2018), these groups are not monophyletic. Although taxonomic sampling of Pristigasteridae is limited in the

present study, which was not designed to resolve interrelationships within the family, the recovered phylogenetic pattern indicates that Pellonulinae are not monophyletic. The origin of Pristigasteridae is estimated in my calibrated analysis to be around 53 mya (Fig. 6-10) which is consistent with divergence time estimates of Lavoué et al. (2013), Bloom and Lovejoy (2014), Bloom et al. (2018). Previously, it was thought that the group could date back as far as the Late Cretaceous based on the occurrence of stem pristigasterid fossil taxon †*Gasteroclupea branisai* in Santonian (83.5 – 85.8 mya) deposits of Bolivia (Signeux, 1964; Grande, 1985). This taxon has since been re-described in detail by Marramà and Carnevale (2017) and found to be a member of †Ellimmichthyiformes; results of the present study also confirm placement of †*Gasteroclupea branisai* within †Ellimmichthyiformes. Therefore, Pristigasteridae have no known fossil record.

Chirocentridae are a monogeneric family containing two very similar species, *Chirocentrus dorab* and *C. nudus*, both occurring in the Indo-West Pacific. These species are of high economic value for local fisheries in India, Pakistan, Malaysia, Tanzania, Singapore, and other countries in the region (Whitehead, 1985; FAO Fisheries Department, 2019). Chirocentrids are commonly known as ‘wolf herrings’, referring to their prominent fang-like dentition. Monophyly of the group is strongly supported by contact between the wings of the lateral ethmoid and the anterior portion of the mesethmoid [6:1], length of the vomer being less than twice the width [10:0], contact between the first and third infraorbital bones [44:1], anterior margin of the metapterygoid located anterior to the quadrate [55:1], presence of large, fang-like teeth on the premaxilla [64:2] and dentary [76:2] (fang-like dentition is found in other clupeiforms such as the pristigasterid *Chirocentrodon*, engraulid *Lycothrissa*, and pellenulid *Odaxothrissa*), presence of a single supraneural [119:1], and presence of a U-shaped pelvic scute [158:2]. The evolutionary history

of Chirocentridae is very poorly understood. Fossil taxa previously suggested to be members of the family have since been reclassified as either †ichthyodectiforms or †ellimmichthyiforms (Grande, 1985; Nelson et al., 2016); therefore, the origin of the group remains uncertain. It has been suggested (Whitehead, 1963; Nelson, 1970) that Chirocentridae are a primitive clupeomorph lineage retaining some ancestral features such as their fang-like dentition, lack of abdominal scutes, and presence of a spiral valve in the intestine; however, as pointed out by Grande (1985) these features also occur in other clupeomorph lineages and do not necessarily represent plesiomorphic conditions.

In the present study, with only two closely related extant species sampled, the age of Chirocentridae is estimated to be 2 mya, but divergence of the family from its sister group Pristigasteridae is estimated to be as early as the earliest Late Cretaceous (~ 101 mya) creating a long ghost lineage leading to crown chirocentrids (Fig. 6-10). The phylogenetic placement of Chirocentridae within Clupeiformes is also uncertain. In recent molecular studies, the family was recovered as sister to either Spratelloididae (Lavoué et al., 2007; Li and Orti, 2007; Wilson et al., 2008), Dussumieriidae (Lavoué et al., 2013), Engraulidae (Bloom and Lovejoy, 2014; Bloom and Egan, 2018, mitochondrial markers), or Pristigasteridae (Bloom and Egan, 2018, nuclear markers). Some morphological studies of chirocentrids suggest their close affinities to clupeids and pristigasterids based on the unique W-shaped occipital articulation (Miyashita, 2010) while other studies indicate close evolutionary relationships between chirocentrids and engraulids based on several features of the suspensorium, branchial arches, and mandibular dentition (Di Dario, 2009; Capobianco et al., 2020). Resolving this uncertainty in the phylogenetic placement of Chirocentridae among other clupeomorphs requires a large scale phylogenomic analyses as well as discovery and description of fossil chirocentrid taxa that could provide invaluable

information about character evolution in this peculiar clupeomorph group and help fill in missing pieces along the long ghost lineage leading to modern species of *Chirocentrus*.

Dussumieriidae are another clupeomorph family that has undergone considerable systematic rearrangements in the past few decades. Members of this group are commonly known as round herrings and are characterized by the presence of an unkeeled W-shaped pelvic scute (Whitehead, 1963, 1972; Grande, 1985) [158:1]. This group was previously placed within Clupeidae (Grande, 1985; Whitehead, 1985; Nelson et al., 2016) as a subfamily subdivided into two tribes, Dussumieriini, diagnosed by a high number of branchiostegal rays [87:1] (Grande, 1985), and Spratelloidini, characterized by an expansion of the fourth hypural [135:0], reduced number of epurals [137:0], fusion between the first preural and first ural centrum [129:1], and absence of the sixth infraorbital [46:0]. Results of recent molecular analyses indicate, however, that Dussumieriini and Spratelloidini do not form a monophyletic group and are not members of Clupeidae (Lavoué et al., 2007, 2008, 2013; Li and Orti, 2007; Wilson et al., 2008; Bloom and Lovejoy, 2014; Bloom and Egan, 2018). Instead, these groups represent independent lineages, both of which have been assigned to family rank (Fricke et al., 2020). In the present study, combined evidence analysis recovered well-supported monophyly for Dussumieriidae comprised of genera *Dussumieria* and *Etrumeus* and placed as the sister group to the clade (Chirocentridae, Pristigasteridae) with divergence time estimated to be 107 mya. In addition to traits mentioned above, dussumieriids are among the very few clupeiforms that have the parhypural fused to the first preural centrum [127:1]. This condition is considered primitive in clupeomorphs and is found in †Ellimmichthyiformes, Denticipitidae, and clupeiform genera *Amblygaster*, *Anodontostoma*, *Opisthonema*, and *Odontognathus*. This condition is not observed in family Spratelloididae, which include small round herrings belonging to the genera *Spratelloides* and

Jenkinsia. Similar to dussumieriids, chirocentrids, and some engraulids, members of Spratelloididae do not have prepelvic scutes [156:0]. This trait, together with the presence of a peculiar W-shaped pelvic scute, pulled Dussumieriidae and Spratelloididae together to form a clade in morphological data set analyses; however, molecular and combined evidence analyses recovered Spratelloididae as the sister group to the clupeid subfamily Ehiravinae, represented by the genera *Sundasalanx* and *Clupeichthys*. Interestingly, Whitehead (1963) recognized ehiravins as members of Spratelloidinae. He later revised his classification by removing ehiravins from his Dussumieriidae, justifying this decision by indicating that the lack or poor development of the abdominal scute series was not a satisfactory character to group ehiravins and spratelloidins (Whitehead, 1972). In the present study, I recognize Ehiravinae and Spratelloidinae as two subfamilies within the Spratelloididae. This group also includes two fossil taxa, a round herring †*Trollichthys bolcensis* from the early Eocene of Italy, and †*Chasmoclupea aegyptica* from the Oligocene of Egypt. Similar to other spratelloidin taxa, †*Trollichthys bolcensis* has an autogenous parhypural not fused to the first preural centrum and highly consolidated caudal fin support with the first preural centrum fused with the first and second ural centra. There are, however, no unambiguous morphological traits supporting placement of †*Chasmoclupea aegyptica* within Ehiravinae; however, in analyses of morphological data this fossil clupeiform taxon has been consistently recovered as a member of the Dorosomatinae *sensu* Lavoué et al., 2014 mostly based on presence of a relatively short maxilla compared to the length of the premaxilla [63:1]. Considering the results of morphological data analyses, which recovered a better supported placement of †*Chasmoclupea aegyptica*, I classify this taxon as a member of Dorosomatinae *sensu* Lavoué et al., 2014.

Clupeidae sensu Grande (1985) are the least understood family among clupeiforms. This group includes almost 200 species, many of which are of high economic value such as Atlantic and Pacific herring (*Clupea harengus* and *C. pallasii*), European sprat (*Sprattus sprattus*), and American, Allis, and Twaite shads (*Alosa sapidissima*, *A. alosa* and *A. fallax*). Considering their high commercial value, systematics of these fishes is of primary importance for conservation and sustainable stock management. However, establishing a stable and well-supported classification of the group has been a challenging task. There is no support for monophyly of Clupeidae sensu Grande (1985) who defined this group by the presence of two elongate rod-like postcleithra. This condition is also observed in the dussumieriid genus *Etrumeus*, but not in other members of Dussumieriidae. More importantly, Clupeidae was not recovered as a natural monophyletic group in any analyses conducted in the present study except for implied weighting parsimony under the stringent weighting scheme (K=1-3). Instead, results of most analyses indicate that the pellonulins *Limnothrissa miodon* and *Stolothrissa tanganyicae* as well as clupeins *Clupea harengus*, *C. pallasii*, *Hyperlophus vittatus*, *Potamalosa richmondia*, and a fossil clupeid taxon †*Knightsia eocaena* form a clade that is more closely related to Dussumieriidae, Chirocentridae, and Pristigasteridae than to the rest of the clupeid taxa. This group is weakly supported and there are no unambiguous morphological synapomorphies supporting this clade. The distinguishing combination of non-unique characters for this clade is defined by the poorly developed ridge between the orbital wall and the levator arcus palatini fossa on the frontal and sphenotic ([31:0], this ridge is very pronounced in all other taxa traditionally assigned to Clupeidae, as well as in *Chirocentrus*, *Etrumeus*, and *Denticeps*), the pterotic excluded from the subepiotic fossa ([37:0], although it is included in *Potamalosa*), lack of the ectopterygoid dentition ([49:0], but present in *Potamalosa*), and lack of the mediopharyngobranchial cartilage ([93:0], mediobranchial cartilage

is present in all other clupeids except *Brevoortia* and *Odaxothrissa*). Close evolutionary relationships among genera *Clupea*, *Sprattus*, *Hyperlophus*, and *Potamalosa* have been consistently recovered in molecular studies (Wilson et al., 2008; Lavoué et al., 2013; Bloom and Lovejoy, 2014), indicating the evolutionary integrity of this clupeid lineage. I assign these taxa to the family Clupeidae with a new, more restricted usage. Inclusion of the two pellonulins *Limnothrissa miodon* and *Stolothrissa tanganicae* in the Clupeidae (new usage), however, seems doubtful considering that this placement was not recovered in any other analysis performed in this study and is likely to be an artifact of using a fully resolved tree from a Bayesian inference analysis which can lead to retaining poorly supported and incorrectly inferred clades. The phylogenetic placement of *Limnothrissa* and *Stolothrissa* should be further assessed with more thorough taxonomic sampling of pellonulin taxa. The Eocene clupeomorph †*Knightsia eocaena* was also recovered as a member of the redefined Clupeidae clade. This fossil species is the sister taxon to genera *Hyperlophus* and *Potamalosa*. Interestingly, all three taxa have previously been placed in clupeid subfamily Pellonulinae (Grande, 1985) based on loss of the anterior supramaxilla [70:0]; however, Grande (1985) suggested that these genera, along with two others (*Sierrathrissa* and *Clupeoides*) not included in the present study, should be removed from the subfamily because they do not possess other diagnostic features of the group. The close evolutionary relationships among †*Knightsia*, *Hyperlophus*, and *Potamalosa* are also highlighted by the presence of a complete series of dorsal scutes in these taxa [155:2]. Dorsal scutes are lost in *Clupea* and *Sprattus*.

The rest of the clupeoid taxa form a clade that can be subdivided into two well-supported monophyletic groups traditionally recognized as Alosinae *sensu* Lavoué et al., 2014 and Dorosomatinae *sensu* Lavoué et al., 2014. To be consistent with taxonomic ranks in the sister

clade that includes clupeiform families Pristigasteridae, Chirocentridae, Dussumieriidae, and Clupeidae (new usage), I assign family rank to the other two clupeoid groups as well: Alosidae (new rank) and Dorosomatidae (new rank). Alosidae comprise genera *Alosa*, *Brevoortia*, *Sardina*, and *Sardinops*. Monophyly of this group is strongly supported by molecular and morphological data and has been consistently recovered in all previous phylogenetic studies of Clupeiformes (e.g., Lavoué et al., 2007, 2013; Wilson et al., 2008; Bloom and Lovejoy, 2014; Bloom and Egan, 2018). Morphologically, members of Alosidae are distinguished by presence of metapterygoid–quadrate fenestra ([54:1], absent in *Sardina*), anterior margin of the metapterygoid located anterior to the quadrate ([55:1], this condition is also present in Engraulidae, Chirocentridae, and *Clupea*), conspicuous radial striations on the opercle ([62:1]; this condition was not observed in any other clupeoid species examined in this study, although, it is present in the †ellimmichthyiform taxa †*Sorbinichthys*, †*Thorectichthys*, †*Foreyclupea*, †*Rhombichthys*, †*Tycheoichthys*, †*Ellimma branneri*, †*Paraclupea seilacheri*). Two genera, *Alosa* and *Brevoortia*, share more morphological and ecological similarities with each other than with the other two genera included in the family. Both genera have a characteristic median notch in the upper jaw [73:1] that distinguishes them from all other clupeiforms with a similar appearance. The fossil record of Alosidae is poorly known with the oldest putative member of the group, †*Pugliaclupea nolardi*, described from the Late Cretaceous (~ 74 mya) of Italy (Taverne, 2004). This taxon, however, is known only from fragmentary material and requires a thorough reassessment based on more complete material and a revised morphological diagnosis of the family that would allow a more confident phylogenetic placement of this fossil taxon. Overall, the age of Alosidae was estimated to be early Eocene (~ 54 mya) with divergence between *Alosa* and *Brevoortia* happening in the late Oligocene (~ 24 mya) (Fig. 6-7). These time

estimates are consistent with previous results of time-calibrated analyses of Clupeiformes (Lavoué et al., 2013; Bloom and Lovejoy, 2014).

Dorosomatidae (new rank) are a diverse assemblage of clupeomorph fishes that include species distributed in tropical and subtropical waters worldwide. Previous systematic revisions of Clupeomorpha and dorosomatins specifically (Miller, 1960; Nelson and Rotman, 1973; Grande, 1985) indicated that this group is highly heterogeneous, and there is no unique morphological trait that would delimit this group. In the present analysis, Dorosomatidae are supported by two homoplastic characters – number of recessus lateralis foramina (most dorosomatids have three openings [23:1], however, *Konosirus*, *Dorosoma*, *Nematalosa*, and *Opisthonema* have four [23:0], and *Odaxothrissa* has a single opening for the recessus lateralis) and presence of a well-developed dorsal blade of the urohyal ([89:0], this condition is also present in Pristigasteridae, Chirocentridae, some clupeids and alosids, but those species can be easily distinguished from dorosomatids by their respective diagnostic features). Among other most noticeable morphological traits that characterize some members of Dorosomatidae are the presence of a long filamentous last dorsal fin ray [125:1] and a relatively short maxilla [63:1] without expansion of the distal blade [67:1] and with its distal end pointing ventrally [68:2], which are present in members of genera *Dorosoma*, *Konosirum*, *Opisthonema*, and *Nematalosa*. Although identifying synapomorphies for Dorosomatidae still requires more studies, results of molecular studies consistently support monophyly of this group (e.g., Lavoué et al., 2007, 2013; Li and Orti, 2007; Wilson et al., 2008; Bloom and Lovejoy, 2014; Bloom and Egan, 2018). Notably, these studies recover the same well-supported sister group relationship between Dorosomatidae and Alosidae.

A revised classification of Clupeomorpha based on genera examined in this study is as follows (Fig. 6-12):

Superorder Clupeomorpha Greenwood et al., 1966

Order Denticipitiformes Clausen, 1959, new rank

Suborder Denticipitoidei Clausen, 1959

Family Denticipitidae Clausen, 1959

Genus *Denticeps* Clausen, 1959

Genus †*Palaeodenticeps* Greenwood, 1960

Order †Ellimmichthyiformes Grande, 1985

Suborder †Armigatoidei Murray and Wilson, 2013

Family †Armigatidae Murray and Wilson, 2013

Genus †*Armigatus* Grande, 1982

Genus †*Codoichthys* Silva Santos, 1994

Genus †*Diplomystus* Cope, 1877

Genus †*Foreyclupea* Vernygora et al., 2016

Genus †*Ornategulum* Forey, 1973

Genus †*Horseshoeichthys* Newbrey et al., 2010

Genus †*Kwangoclupea* Taverne, 1997

Genus †*Ranulfoichthys* Alvarado-Ortega, 2014

Suborder †Ellimmichthyoidei Grande, 1985

Family †Scutatuspinosidae Silva Santos & Silva Corrêa, 1985

Genus †*Scutatuspinosus* Silva Santos and Silva Corrêa, 1985

Family †Sorbinichthyidae Bannikov and Bacchia, 2000

Genus †*Sorbinichthys* Bannikov and Bacchia, 2000

Family †Thorectichthyidae Murray and Wilson, 2013, new rank

Genus †*Thorectichthys* Murray and Wilson, 2013

Family †Paraclupeidae Chang and Chou, 1977

Subfamily †Ellimminae Murray and Wilson, 2013

Genus †*Ellimma* Jordan, 1913

Genus †*Tycheroichthys* Hay et al., 2007

† '*Ellimmichthys*' *maceioensis* Malabarba et al., 2004

Subfamily †Paraclupeinae Chang and Chou, 1977

Genus †*Ellimmichthys* Cope, 1886

†*Ezcutuberezi carmenae* Poyato-Ariza et al., 2000

†*Paraclupea chetungensis* Sun, 1956

unranked

†*Eoellimmichthys surpestes* Marammà et al., 2019

Subfamily Triplomystinae Murray and Wilson 2013, new rank

Genus †*Gasteroclupea* Signeux, 1964

Genus †*Scutatoclupea* Bannikov, 2015

Genus †*Triplomystus* Forey et al., 2003

Genus †*Rhombichthys* Khalloufi et al., 2010

† '*Diplomystus*' *solignaci* Gaudant and Gaudant, 1971

† '*Paraclupea*' *seilacheri* Alvarado-Ortega and Melgarejo-Damián, 2017

Order Clupeiformes Bleeker, 1859

unranked

†*Pseudoellimma gallae* De Figueiredo, 2009

†*Santanaclupea silvasantosi* Maisey, 1993

Suborder Clupeioidi Bleeker, 1859

Family Engraulidae Gill, 1861

Subfamily Engraulinae Gill, 1861

Genus *Anchoviella* Fowler, 1911

Genus *Cetengraulis* Günther, 1868

Genus *Engraulis* Cuvier, 1816

Genus *Lycengraulis* Günther 1868

Genus *Stolephorus* Lacepède, 1803

Subfamily Coiliinae Bleeker, 1872

Genus *Coilia* Gray, 1830

Genus *Setipinna* Swainson, 1839

Family Chirocentridae Bleeker, 1849

Genus *Chirocentrus* Cuvier 1816

Family Pristigasteridae Bleeker, 1872

Genus *Chirocentrodon* Günther, 1868

Genus *Ilisha* Richardson, 1846b

Genus *Odontognathus* Lacepède, 1800

Genus *Pellona* Valenciennes in Cuvier and Valenciennes, 1847

Family Dussumieriidae Gill, 1861

Genus *Dussumieria* Valenciennes in Cuvier and Valenciennes, 1847

Genus *Etrumeus* Bleeker, 1853b

Family Spratelloididae Jordan, 1925

Subfamily Spratelloidinae Jordan, 1925

Genus *Jenkinsia* Jordan and Evermann, 1896

Genus *Spratelloides* Bleeker, 1851

Genus †*Trollichthys* Marramà and Carnevale, 2015

Subfamily Ehiravinae Deraniyagala, 1929

Genus *Clupeichthys* Bleeker, 1855

Genus *Sundasalanx* Roberts, 1981

Family Clupeidae Cuvier, 1816

Genus *Clupea* Linnaeus, 1758

Genus *Hyperlophus* Ogilby, 1892

Genus *Potamalosa* Ogilby, 1897

Genus *Sprattus* Girgensohn, 1846

Genus †*Knightia* Jordan, 1907

Family Alosidae Svetovidov, 1952, new rank

Genus *Alosa* Linck, 1790

Genus *Brevoortia* Gill, 1861

Genus *Sardina* Antipa, 1904

Genus *Sardinops* Hubbs, 1929

Family Dorosomatidae Gill, 1861

Genus *Amblygaster* Bleeker, 1849a

Genus *Anodontostoma* Bleeker, 1849b

Genus †*Chasmoclupea* Murray et al., 2005

Genus *Dorosoma* Rafinesque, 1820

Genus *Escualosa* Whitley, 1940

Genus *Harengula* Valenciennes in Cuvier and Valenciennes, 1847

Genus *Herklotsichthys* Whitley, 1951

Genus *Konosirus* Jordan and Snyder, 1900

Genus *Lile* Jordan and Evermann, 1896

Genus *Limnothrissa* Regan, 1917a

Genus *Nematalosa* Regan, 1917b

Genus *Odaxothrissa* Boulenger, 1899

Genus *Opisthonema* Gill, 1861

Genus *Pellonula* Günther, 1868

Genus *Sardinella* Valenciennes in Cuvier and Valenciennes, 1847

Genus *Stolothrissa* Regan, 1917a

Genus *Tenualosa* Fowler, 1934

6.5 Conclusions

Results of the multiple phylogenetic analyses performed in the present study indicate a substantial degree of agreement in the phylogenetic signal between morphological and molecular data when the data sets were analysed separately. Although relatively high homoplasy in morphological data resulted in considerable variation in the arrangement of major clupeomorph groups, as indicated by the results of implied weighting parsimony analyses, a common pattern of evolutionary interrelationships was observed across summary topologies. This pattern was corroborated by combined evidence analysis. This synthesis of multiple phylogenetic results

suggests that Clupeomorpha are a monophyletic group that can be subdivided into three orders, Denticipitiformes, †Ellimmichthyiformes, and Clupeiformes, with Denticipitidae being the basalmost clupeomorph lineage. Interrelationships of higher level taxa within †Ellimmichthyiformes and Clupeiformes are overall consistent with the current classification. Because the family Clupeidae in its current usage does not represent a monophyletic group, I assign a family rank to previously recognized clupeid subfamilies that have been consistently recovered as monophyletic groups, Clupeidae (new usage), Dorosomatidae, and Alosidae. Divergence time estimation analysis recovered two Early Cretaceous taxa as stem clupeiforms suggesting that divergence between †Ellimmichthyiformes and Clupeiformes happened before the Barremian with the age of the total group Clupeomorpha estimated to be around 214 mya.

References

- Alvarado-Ortega, J. 2014. Ancient herring from the Tlayúa Quarry (Cretaceous, Albian) near Tepexi de Rodríguez, Puebla State, central Mexico, closing the gap in the early diversification of Clupeomorpha. *Cretaceous Research* 50:171–186.
- Alvarado-Ortega, J. and E. Ovalles-Damián. 2008. *Triplomystus applegatei*, sp.nov. (Teleostei: Ellimmichthyiformes), a rare “triple armored herring” from El Espinal Quarry (Early Cretaceous), Chiapas, southeastern Mexico. *Journal of Vertebrate Paleontology* 28:53–60.
- Alvarado-Ortega, J., E. Ovalles-Damián, and G. Arratia. 2008. A review of the interrelationships of the order Ellimmichthyiformes (Teleostei: Clupeomorpha); pp. 257–278 in G. Arratia, H.-P. Schultze, and M.V.H. Wilson (eds.), *Mesozoic Fishes 4 – Homology and Phylogeny*. Verlag Dr. Friederich Pfeil, Munich, Germany, 502 pp.
- Alvarado-Ortega, J. and M. Melgarejo-Damián. 2017. *Parachupea seilacheri* sp. nov., a double armored herring (Clupeomorpha, Ellimmichthyiformes) from the Albian limestones of Tlayúa quarry, Puebla, Mexico. *Revista Mexicana de Ciencias Geológicas* 34:234–249.
- Antipa, G. 1904. Die Clupeinen des westlichen Teiles des Schwarzen Meeres und der Donaumündungen. *Anzeiger der Kaiserlichen Akademie der Wissenschaften, Wien, Mathematisch-Naturwissenschaftliche Classe* 41:299–303.
- Bannikov, A. F. 2015. A new genus for the Cenomanian ellimmichthyiform fishes from Lebanon and Mexico. *Bollettino della Societa Paleontologica Italiana* 54:211–218.
- Bannikov, A. F. and F. Bacchia. 2000. A remarkable clupeomorph fish (Pisces, Teleostei) from a new Upper Cretaceous marine locality in Lebanon. *Senckenbergiana Lethaea* 80:3–11.

- Bennett, E. T. 1832. Observations on a collection of fishes from the Mauritius, presented by Mr. Telfair, with characters of new genera and species. *Proceedings of the Committee of Science and Correspondence of the Zoological Society of London* 31:165–169.
- Bennett, E. T. 1835. A letter to the Secretary by Keith E. Abbott, Esq., a collection of skins of mammals and birds, and of preserved reptiles, fishes, and insects. *Proceedings of the Zoological Society of London* 3:89–92.
- Berg, L. S. 1940. *Sistema ryboobraznykh i ryb, nye zhivushchikh i iskopaemykh: Classification of fishes, both recent and fossil.* Izd-vo Akademii nauk SSSR, 517 pp.
- Blainville, H. M. de. 1818. Sur les ichthyolites ou les poissons fossiles. *Paris, Nouvelles Dictées d'Histoire naturelle* 37:310–395.
- Bleeker, P. 1849a. A contribution to the knowledge of the ichthyological fauna of Celebes. *Journal of the Indian Archipelago and Eastern Asia (Singapore)* 31:65–74.
- Bleeker, P. 1849b. Bijdrage tot de kennis der ichthyologische fauna van het eiland Madura, met beschrijving van eenige nieuwe soorten. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen.* 22:1–16.
- Bleeker, P. 1851. Nieuwe bijdrage tot de kennis der ichthyologische fauna van Celebes. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 2:209–224.
- Bleeker, P. 1853a. Diagnostische beschrijvingen van nieuwe of weinig bekende vischsoorten van Batavia. Tiental I-VI. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 4:451–516.
- Bleeker, P. 1853b. Nalezingen op de ichthyologie van Japan. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen* 25:1–56.
- Bleeker, P. 1854. *Faunae ichthyologicae japonicae species novae.* *Natuurkundig Tijdschrift voor Nederlandsch Indië* 6:395–426.

- Bleeker, P. 1855. Nalezingeren op de vischfauna van Sumatra. Visschen van Lahat en Sibogha. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 9:257–280.
- Bleeker, P. 1857. Zesde bijdrage tot de kennis der vischfauna van Sumatra. Visschen van Padang, Troessan, Priaman, Sibogha en Palembang. *Acta Societatis Regiae Scientiarum Indo-Neêrlandicae* 3:1–50.
- Bleeker, P. 1859. Enumeratio specierum piscium hucusque in Archipelago indico observatarum adjectis habitationibus citationibusque, ubi descriptiones earum recentiores reperiuntur, nec non speciebus Musei Bleekeriani Bengalensibus, Japonicis, Capensibus Tasmanicisque. *Acta Societatis Regiae Scientiarum Indo-Neerlandicae* 6:1–276.
- Bleeker, P. 1972. Atlas ichthyologique des Indes Orientales Neerlandaises 6:79–143.
- Bloch, M. E. and J. G. Schneider. 1801. *Systema Ichthyologiae Iconibus. Illustratum. Post obitum auctoris opus inchoatum absolvit, correxit, interpolavit Jo. Gottlob Schneider, Saxo. Berolini. Sumtibus Auctoris Impressum et Bibliopolio Sanderiano Commissum.* 110 pp.
- Bloom, D. D. and N. R. Lovejoy. 2012 Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (Family Engraulidae). *Journal of Evolutionary Biology* 25:701–715.
- Bloom, D. D. and N. R. Lovejoy. 2014. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proceedings of the Royal Society B: Biological Sciences* 281:2013–2081.
- Bloom, D. D., M. D. Burns, and T. A. Schriever. 2018. Evolution of body size and trophic position in migratory fishes: a phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies). *Biological Journal of the Linnean Society* 125:302–314.

- Bloom, D. D. and J. P. Egan. 2018. Systematics of Clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. *Neotropical Ichthyology* 16:p.e180095.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* 10:e1003537.
- Boulenger, G. A. 1899. Matériaux pour la faune du Congo. Zoologie.--Série I. Poissons nouveaux du Congo. Quatrième Partie. Polyptères, Clupes, Mormyres, Characins. *Annales du Musée du Congo (Ser. Zoology)* 1:59–96.
- Boulenger, G. A. 1906. Fourth contribution to the ichthyology of Lake Tanganyika.--Report on the collection of fishes made by Dr. W. A. Cunnington during the Third Tanganyika Expedition, 1904-1905. *Transactions of the Zoological Society of London* 17(1):537–601.
- Capobianco, A., H. T. Beckett, E. Steurbaut, P. D. Gingerich, G. Carnevale, and M. Friedman. 2020. Large-bodied sabre-toothed anchovies reveal unanticipated ecological diversity in early Palaeogene teleosts. *Royal Society Open Science* 7:192260.
- Casier, E. 1946 La faune ichthyologique de l'Yprésien de la Belgique. *Mémoires du Musée Royal d'Histoire Naturelle de Belgique* 104:1–267.
- Casier, E. 1965. Poissons fossils de la Série du Kwanga (Congo). *Sciences Geologiques. Memoire* 50:1–64.
- Castelnau, F. L. 1875. Researches on the fishes of Australia. *Philadelphia Centennial Expedition of 1876. Intercolonial Exhibition Essays* 2:1–52.

- Chang, M.-M., and C.-C. Chou. 1977. On late Mesozoic fossil fishes from Zhejiang Province, China. *Memoirs, Institute of Vertebrate Paleontology and Paleoanthropology, Academia Sinica* 12:1–60.
- Chang, M.-M. and J. Maisey. 2003. Redescription of *Ellimma branneri* and *Diplomystus shengliensis*, and relationships of some basal clupeomorphs. *American Museum Novitates* 3404:1–35.
- Clausen, H. S. 1959. Denticipitidae, a new family of primitive Isospondylous teleosts from West African fresh-water. *Videnskabelige Meddelelser Naturhistorisk Forening, Kjøbenhavn* 121:141–151.
- Cope, E. D. 1877. A contribution to the knowledge of the ichthyological fauna of the Green River Shales. *Bulletin of United States Geological and Geographical Survey* 3:807–819.
- Cope, E. D. 1886. A contribution to the vertebrate paleontology of Brazil. *Proceedings of the American Philosophical Society* 23:3–4.
- Cuvier, G. 1816. *Le Règne Animal distribué d'après son organisation pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Les reptiles, les poissons, les mollusques et les annelids*, 532 pp.
- Cuvier, G. 1829. *Le Règne Animal, distribué d'après son organisation, pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Edition 2*. 406 pp.
- Daget, J. 1984. Denticipitidae; p. 10 in J. Daget, J. P. Gosse, D. F. E. Thys van den Audenaerde (eds.), *Check-list of the freshwater fishes of Africa. Cloffa 1*. Paris, Orstom/Tervuren, MRAC, 410 pp.

- De Figueiredo, F. J. 2009. A new Clupeiform fish from the Lower Cretaceous (Barremian) of Sergipe-Alagoas Basin, northeastern Brazil. *Journal of Vertebrate Paleontology* 29:993–1005.
- De Kay, J. E. 1842. *Zoology of New-York, or the New-York fauna; comprising detailed descriptions of all the animals hitherto observed within the state of New-York, with brief notices of those occasionally found near its borders, and accompanied by appropriate illustrations. Part IV. Fishes.* W. & A. White & J. Visscher, Albany, 415 pp.
- Deraniyagala, P. E. P. 1929. Ceylon sardines. *Spolia Zeylanica (The Ceylon Journal of Science, Section B. Zoology and Geology)* 15:31–47.
- Di Dario, F. 2002. Evidence supporting a sister-group relationship between Clupeoidea and Engrauloidea (Clupeomorpha). *Copeia* 2002:496–503.
- Di Dario, F. 2004. Homology between the recessus lateralis and cephalic sensory canals, with the proposition of additional synapomorphies for the Clupeiformes and the Clupeoidei. *Zoological Journal of the Linnean Society* 141:257–270.
- Di Dario, F. 2009. Chirocentrids as engrauloids: evidence from suspensorium, branchial arches, and infraorbital bones (Clupeomorpha, Teleostei). *Zoological Journal of the Linnean Society* 156:363–383.
- Di Dario, F. and M. C. De Pinna. 2006. The supratemporal system and the pattern of ramification of cephalic sensory canals in *Denticeps clupeoides* (Denticipitoidei, Teleostei): additional evidence for monophyly of Clupeiformes and Clupeoidei. *Papéis Avulsos de Zoologia* 46:107–123.
- Durand, J. 1940. Notes sur quelques poissons d'espèces nouvelles ou peu connues des eaux douces cambodgiennes. *Institut Océanographic de l'Indochine Note* 36:1–40.

- Eastman, C. R. 1912. Tertiary fish-remains from Spanish Guinea in West Africa. *Annals of the Carnegie Museum* 8:370–378.
- Eichwald, C. E. von. 1838. Faunae Caspii maris primitiae. *Bulletin de la Société Impériale des Naturalistes de Moscou* 11:125–174.
- Figueiredo, F. J. and D. R. Ribeiro. 2016. Relationships of †*Codoichthys carnavalii* Santos, 1994 (Teleostei, Clupeomorpha, † Ellimmichthyiformes) from the Late Aptian of São Luís-Grajaú Basin, NE Brazil. *Anais da Academia Brasileira de Ciências* 88:1277–1307.
- Figueiredo, F. J. and D. R. Ribeiro. 2017. Reassessment and Relationships of †*Scutatuspinosus itapagipensis* (Teleostei, Clupeomorpha, † Ellimmichthyiformes) from the Neocomian of Recôncavo Basin, Northeastern Brazil. *Anais da Academia Brasileira de Ciências* 89:799–823.
- FAO Fisheries department. 2019. FAO yearbook. Fishery and Aquaculture Statistics 2017/FAO annuaire. Statistiques des pêches et de l'aquaculture 2017/ FAO anuario. Estadísticas de pesca y acuicultura 2017. Rome/Roma, 109 pp.
- Forey, P. L. 1973. A primitive clupeomorph fish from the middle Cenomanian of Hakel, Lebanon. *Canadian Journal of Earth Sciences* 10:1302–1318.
- Forey, P. L. 1975. A fossil clupeomorph fish from the Albian of the Northwest Territories of Canada, with notes on cladistic relationships of clupeomorphs. *Journal of Zoology* 175:151–177.
- Forey, P. L. 2004. A three-dimensional skull of a primitive clupeomorph from the Cenomanian English Chalk and implications for the evolution of the clupeomorph acustico-lateralis system; pp. 405–427 in G. Arratia, and A. Tintori (eds.), *Mesozoic Fishes 3—Systematics, Paleoenvironments and Biodiversity*. Verlag Dr. Friedrich Pfeil, Munich, 649 pp.

- Forey, P. L., L. Yi, C. Patterson, and C. E. Davies. 2003. Fossil fishes from the Cenomanian (Upper Cretaceous) of Namoura, Lebanon. *Journal of Systematic Palaeontology* 1:227–330.
- Forsskål, P. S. 1775. *Descriptiones animalium, avium, amphibiorum, piscium, insectorum, vermium; quae in itinere orientali observavit. Post mortem auctoris edidit Carsten Niebuhr. Hauniae.* 164 pp.
- Fowler, H. W. 1911. Notes on clupeoid fishes. *Proceedings of the Academy of Natural Sciences of Philadelphia* 63:204–221.
- Fowler, H. W. 1934. Descriptions of new fishes obtained 1907 to 1910, chiefly in the Philippine Islands and adjacent seas. *Proceedings of the Academy of Natural Sciences of Philadelphia* 85:233–367.
- Fricke, R., W. N. Eschmeyer, and J. D. Fong. 2020. Eschmeyer's catalog of fishes: species by family/subfamily. <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>. Electronic version accessed 28 May 2020.
- Froese, R. and D. Pauly. 2019. FishBase. World Wide Web electronic publication. www.fishbase.org. Electronic version 12/2019 accessed 28 May 2020.
- Gill, T. 1872. Arrangement of the families of fishes, or Classes Pisces, Marsipobranchii, and Leptocardii. *Smithsonian Miscellaneous Collections* 11:1–49.
- Gill, T. N. 1861. Synopsis of the subfamily of Clupeinae, with descriptions of new genera. *Proceedings of the Academy of Natural Sciences of Philadelphia* 13:33–38.
- Girgensohn, O. G. L. 1846. Anatomie und Physiologie des Fisch-Nervensystems. *Mémoires de l'Académie Impériale des Sciences de Saint Pétersbourg* 5:275–589.

- Goloboff, P., S. Farris, and K. Nixon. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774–786.
- Goode, G. B. 1878. A revision of the American species of the genus *Brevoortia*, with a description of a new species from the Gulf of Mexico. *Proceedings of the United States National Museum* 1:30–42.
- Goodrich, E. S. 1909. *Vertebrata Craniata (First Fascicle: Cyclostomes and Fishes)*. In *A Treatise on Zoology*; in R. Lankester (ed.) Part IX. Black, London, 518 pp.
- Gosse, P. H. 1851. *A naturalist's sojourn in Jamaica*. Longman, Brown, Green and Longmans, London, 508 pp.
- Gosline, W. A. 1960. Contributions toward a classification of modern isospondylous fishes. *Bulletin of the British Museum of Natural History* 6:325–365.
- Gosline, W. A. 1980. The evolution of some structural systems with reference to the interrelationships of modern lower teleostean fish groups. *Japanese Journal of Ichthyology* 27:1–28.
- Grande, L. 1982. A revision of the fossil genus *Diplomystus*, with comments on the interrelationships of clupeomorph fishes. *American Museum Novitates* 2728:1–34.
- Grande, L. 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. *Bulletin of the American Museum of Natural History* 181:231–372.
- Grande, T., W. C. Borden, and W. L. Smith. 2013. Limits and relationships of *Paracanthopterygii*: a molecular framework for evaluating past morphological hypotheses. *Mesozoic fishes* 5:385–418.

- Grande, L. and G. J. Nelson. 1985. Interrelationships of fossil and Recent anchovies (Teleostei: Engrauloidea) and description of a new species from the Miocene of *Cyprus*. *American Museum Novitates* 2826:1–16.
- Grant, W.S., F. Lecomte and B.W. Bowen. 2010. Biogeographical contingency and the evolution of tropical anchovies (genus *Cetengraulis*) from temperate anchovies (genus *Engraulis*). *Journal of Biogeography* 37:1352–1362.
- Gray, J. E. 1830. Illustrations of Indian zoology; chiefly selected from the collection of Major-General Hardwicke, F.R.S. 1:1–202.
- Greenwood, P. H. 1960. Fossil denticipitid fishes from east Africa. *Bulletin of British Museum of Natural History* 5:1–11.
- Greenwood, P. H. 1968. The osteology and relationships of the Denticipitidae, a family of clupeomorph fishes. *Bulletin of the British Museum of Natural History (Zoology)* 16:215–273.
- Greenwood, P. H., D. E. Rosen, S. H. Weitzman, and G. S. Myers. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History* 131:339–455.
- Günther, A. 1867. On the fishes of the states of Central America, founded upon specimens collected in fresh and marine waters of various parts of that country by Messrs. Salvin and Godman and Capt. J. M. Dow. *Proceedings of the Zoological Society of London* 1866:600–604.
- Günther, A. 1868. Catalogue of the fishes in the British Museum. Catalogue of the Physostomi, containing the families Heteropygii, Cyprinidae, Gonorhynchidae, Hyodontidae,

- Osteoglossidae, Clupeidae,... Halosauridae, in the collection of the British Museum. 7:1–512.
- Gaudant, M., and J. Gaudant. 1971. Une nouvelle espèce de *Diplomystus* (Poisson téléostéen) dans le Crétacésupérieur du Sudtunisien. *Bulletin de la Société géologique de France* 13:156–159.
- Hamilton-Buchanan, F. 1822. An account of the fishes found in the river Ganges and its branches. Edinburgh and London, 405 pp.
- Hata, H., S. Lavoué, and H. Motomura. 2020. *Stolephorus babarani*, a new species of anchovy (Teleostei: Clupeiformes: Engraulidae) from Panay Island, central Philippines. *Zootaxa* 4718:509–520.
- Hay, M. J., S. L. Cumbaa, A. M. Murray, and A. G. Plint. 2007. A new paraclupeid fish (Clupeomorpha, Ellimmichthyiformes) from a muddy marine prodelta environment: middle Cenomanian Dunvegan Formation, Alberta, Canada. *Canadian Journal of Earth Sciences* 44:775–790.
- Hubbs, C. L. 1929. The generic relationships and nomenclature of the California sardine. *Proceedings of the California Academy of Sciences* 18:261–265.
- Hughes, L. C., G. Ortí, Y. Huang, Y. Sun, C. C. Baldwin, A. W. Thompson, D. Arcila, R. Betancur-R, C. Li, L. Becker, and N. Bellora. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences* 115:6249–6254.
- Jenyns, L. 1842. Fish. In: *The zoology of the voyage of H. M. S. Beagle, under the command of Captain Fitzroy, R. N., during the years 1832 to 1836*. London, 172 pp.

- Jordan, D. S. 1907. The fossil fishes of California: with supplementary notes on other species of extinct fishes 5:95–145.
- Jordan, D. S. 1910. Description of a collection of fossil fishes from the bituminous shales at Riaco Doce, state of Alaçôas, Brazil. *Annals of the Carnegie Museum* 7:23–24.
- Jordan, D. S. 1913. *Ellimma*, a genus of fossil herrings. *Proceedings of the Biological Society of Washington* 26:1–79.
- Jordan, D. S. 1923. A Classification of fishes including families and genera as far as known. Stanford University Publications, University Series, Biological Sciences 3:77–243.
- Jordan, D. S. and C. H. Gilbert. 1882a. Catalogue of the fishes collected by Mr. John Xantus at Cape San Lucas, which are now in the United States National Museum, with descriptions of eight new species. *Proceedings of the United States National Museum* 5:353–371.
- Jordan, D. S. and C. H. Gilbert. 1882b. Descriptions of thirty-three new species of fishes from Mazatlan, Mexico. *Proceedings of the United States National Museum* 4:338–365.
- Jordan, D. S. and B. W. Evermann. 1896. Description of a new species of shad (*Alosa alabamae*) from Alabama; pp. 203–205 in B. W. Evermann, Report to the United States Commission on Fisheries. G.P.O., Washington, 610 pp.
- Jordan, D. S. and J. O. Snyder. 1900. A list of fishes collected in Japan by Keinosuke Otaki, and by the United States steamer Albatross, with descriptions of fourteen new species. *Proceedings of the United States National Museum* 23:335–380.
- Katoh, K., K. Misawa, K. I. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30:3059–3066.

- Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30:772–780.
- Khalloufi, B., R. Zaragüeta-Bagils, R. Lelièvre, and H. Lelièvre. 2010. *Rhombichthys intoccabilis*, gen. et sp. nov. (Ellimmichthyiformes, Clupeomorpha, Teleostei), from the Cenomanian (Upper Cretaceous) of Ein Yabrud, Middle East: anatomical description and phylogenetic implications. *Journal of Vertebrate Paleontology* 30:57–67.
- Klunzinger, C. B. 1879. Die v. Müller'sche Sammlung australischer Fische. *Anzeiger der Kaiserlichen Akademie der Wissenschaften, Wien, Mathematisch-Naturwissenschaftliche Classe* 16:254–261.
- Lacepède, B. G. E. 1800. *Histoire naturelle des poissons*. 2:1–632.
- Lacepède, B. G. E. 1803. *Histoire naturelle des poissons*. Tome cinquième. Chez Plassan, Paris, 803 pp.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular biology and evolution* 34:772–773.
- Latrobe, B. H. 1802. A drawing and description of the Clupea Tyrannus and Oniscus Prægustator. *Transactions of the American Philosophical Society* 5:77–81.
- Lavoué, S., M. Miya, K. Saitoh, N. B. Ishiguro. 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Molecular Phylogenetics and Evolution* 43:1096–1105.
- Lavoué, S., M. Miya, A. Kawaguchi, T. Yoshino, and M. Nishida. 2008. The phylogenetic position of an undescribed pedomorphic clupeiform taxon: mitogenomic evidence. *Ichthyological research* 55:328–334.

- Lavoué, S., M. Miya, and M. Nishida. 2010. Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced miniaturization in the order Clupeiformes. *Molecular Phylogenetics and Evolution* 56:480–485.
- Lavoué, S., M. Miya, P. Musikasinthorn, W.-J. Chen, and M. Nishida. 2013. Mitogenomic evidence for an Indo-West Pacific Origin of the Clupeioidi (Teleostei: Clupeiformes). *PLoS ONE* 8:e56485.
- Lavoué, S., P. Konstantinidis, and W. J. Chen. 2014. Progress in clupeiform systematics; pp. 3–42 in K. Ganiyas (ed.), *Biology and ecology of sardines and anchovies*, CRC Press, 394 pp.
- Leache, A. D., B. L. Banbury, J. Felsenstein, A. N. M. de Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64:1032–1047.
- Lesueur, C. A. 1818. Descriptions of several new species of North American fishes. *Journal of the Academy of Natural Sciences, Philadelphia* 1:222–235.
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Li, C. and G. Orti. 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolutions* 44:386–398.
- Linck, H. F. 1790. Versuch einer Eintheilung der Fische nach den Zähnen. *Magazin für das Neueste aus der Physik und Naturgeschichte*, Gotha 6:28–38.
- Linnaeus, C. 1758. *Systema Naturae per Regna Tria Naturae, secundum Classes, Ordines, Genera, Species, cum Characteribus, Differentiis, Synonymis, Locis*. Tonus I. Editio Decima, Reformata. Laurentii Salvii, Holmiae, 824 pp.

- Macleay, W. 1879. On the Clupeidae of Australia. Proceedings of the Linnean Society of New South Wales 4:363–385.
- Maddison, W. P. and D. R. Maddison. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.6 <http://mesquiteproject.org>
- Maisey, J. G. 1993. A new clupeomorph fish from the Santana Formation (Albian) of NE Brazil. American Museum Novitates 3076:1–15.
- Malabarba, M. C., D. A. do Carmo, I. Gómez-Pérez, and J. V. de Queiroz-Neto. 2004. A new clupeomorph fish from the Cretaceous Maceió Formation, Alagoas Basin, NE Brazil. Neues Jahrbuch für Geologie und Paläontologie – Abhandlungen 233:255–274.
- Marrama, G. and G. Carnevale. 2015. Eocene round herring from Monte Bolca, Italy. Acta Palaeontologica Polonica 60:701–710.
- Marramà, G. and G. Carnevale. 2017. The relationships of *Gasteroclupea branisai* Signeux, 1964, a freshwater double-armored herring (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous-Paleocene of South America. Historical Biology 29:904–917.
- Marramà, G., A. F. Bannikov, J. Kriwet, and G. Carnevale. 2019. An Eocene paraclupeid fish (Teleostei, Ellimmichthyiformes) from Bolca, Italy: the youngest marine record of double-armoured herrings. Papers in Palaeontology 5:83–98.
- Minh, B. Q., M. A. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution 30:1188–1195.
- Miller, R. R. 1960. Systematics and biology of the gizzard shad (*Dorosoma cepedianum*) and related fishes. Fisheries 1:371–392.
- Mitchill, S. L. 1814. Report, in part, of Samuel L. Mitchill, M. D., Professor of Natural History, on the fishes of New-York. D. Carlisle, New York, 28 pp.

- Miyashita, T. 2010. Unique occipital articulation with the first vertebra found in pristigasterids, chirocentrids, and clupeids (Teleostei: Clupeiformes: Clupeoidei). *Ichthyological Research* 57:121–132.
- Murray, A. M., E. L. Simons, and Y. S. Attia. 2005. A new clupeid fish (Clupeomorpha) from the Oligocene of Fayum, Egypt, with notes on some other fossil clupeomorphs. *Journal of Vertebrate Paleontology* 25:300–308.
- Murray, A. M. and M. V. H. Wilson. 2011. A new species of *Sorbinichthys* (Teleostei: Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco. *Canadian Journal of Earth Sciences* 48:1–9.
- Murray, A. M., and M. V. H. Wilson. 2013. Two new paraclupeid fishes (Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco; pp. 267–290 in G. Arratia, H.-P. Schultze, and M. V. H. Wilson (eds.), *Mesozoic Fishes 5 – Global Diversity and Evolution*. Verlag Dr. Friederich Pfeil, Munich, Germany, 560 pp.
- Murray, A. M., O. Vernygora, S. Japundžić, J. Radovčić, M. V. H. Wilson, D. Bardack, and T. Grande. 2016. Relationships of the species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) and the description of a new species from the Cretaceous of Dalmatia, Croatia. *Journal of Vertebrate Paleontology* 36:e1226851.
- Nelson, G. J. 1970. The hyobranchial apparatus of teleostean fishes of the families Engraulidae and Chirocentridae. *American Museum Novitates* 2410:1–30.
- Nelson, G. J. 1973. Relationships of clupeomorphs, with remarks on the structure of the lower jaw in fishes; pp.333–349 in P. H. Greenwood, R. S. Miles, and C. Patterson (eds.), *Interrelationships of fishes*. Academic Press, London, 536 pp.

- Nelson, G. J. 1983. *Anchoa argentivittata*, with notes on other eastern Pacific anchovies and the Indo-Pacific genus *Encrasicholina*. *Copeia* 1983:48–54.
- Nelson, G. J. 1984. Notes on the rostral organ of anchovies (Family Engraulidae). *Japanese Journal of Ichthyology* 31:86–87.
- Nelson, G. J. 1986. Identity of the Anchovy *Engraulis clarki* with notes on the species-groups of *Anchoa*. *Copeia* 1986:891–902.
- Nelson, G. J. and M. N. Rothman. 1973. The species of gizzard shads (Dorosomatinae) with particular reference to the Indo-Pacific region. *Bulletin of the American Museum of Natural History* 150:131–206.
- Nelson, J.S., T. C. Grande, and M. V. Wilson. 2016. *Fishes of the World*, fifth edition. John Wiley and Sons, Toronto, 752 pp.
- Newbrey, M. G., A. M. Murray, D. B. Brinkman, M. V. H. Wilson, and A. G. Neuman. 2010. A new articulated freshwater fish (Clupeomorpha, Ellimmichthyiformes) from the Horseshoe Canyon Formation, Maastrichtian, of Alberta, Canada. *Canadian Journal of Earth Sciences* 47:1183–1196.
- Ogilby, J. D. 1892. On some undescribed reptiles and fishes from Australia. *Records of the Australian Museum* 2:23–26.
- Ogilby, J. D. 1897. Rough-backed herrings, in "Notes and Exhibits" section. *Proceedings of the Linnean Society of New South Wales* 21:504–505.
- Patterson, C. 1967. Are the teleosts a polyphyletic group. *Colloques Internationaux du Centre National de la Recherche Scientifique* 163:93–109.

- Patterson, C. and D. E. Rosen. 1977. Review of ichthyodectiform and other Mesozoic teleost fishes, and the theory and practice of classifying fossils. *Bulletin of the American Museum of Natural History* 158:81–172.
- Pictet, F. J. 1850. Description de quelques poissons fossiles du Mont Liban. Jules-Guillaume Fick, 59 pp.
- Poey, F. 1865. Peces nuevos de la Isla de Cuba. *Repertorio Fisico-Natural de la Isla de Cuba* 1:181–192.
- Poey, F. 1867. Peces Cubanos especies nuevas. *Repertorio Fisico-Natural de la Isla de Cuba* 2:229–245.
- Poyato-Ariza, F. J., M. A. López-Horgue, M. A. Garcia-Garmilla, and F. Garcia-Garmilla. 2000. A new Early Cretaceous clupeomorph fish from the Arratia Valley, Basque Country, Spain. *Cretaceous Research* 21:571–585.
- Rafinesque, C. S. 1820. Fishes of the Ohio River. *Western Review and Miscellaneous Magazine* 2:169–177.
- Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology* 67:901–904.
- Regan, C. T. 1917a. A revision of the clupeid fishes of the genus *Pellonula* and of related genera in the rivers of Africa. *Annals and Magazine of Natural History* 19:198–207.
- Regan, C. T. 1917b. A revision of the clupeoid fishes of the genera *Pomolobus*, *Brevoortia* and *Dorosoma* and their allies. *Annals and Magazine of Natural History* 19:297–316.
- Richardson, J. 1846a. Ichthyology of the voyage of H. M. S. Erebus & Terror, under the command of Captain Sir James Clark Ross, R. N., F. R. S.; pp. 1–139 in: J. Richardson &

- J. E. Gray (eds.): The zoology of the voyage of H. M. S. Erebus & Terror, under the command of Captain Sir J. C. Ross, R. N., F. R. S., during the years 1839 to 1843. E. W. Janson, London, 139 pp.
- Richardson, J. 1846b. Report on the ichthyology of the seas of China and Japan. Report of the British Association for the Advancement of Science 15th meeting 1845:187–320.
- Roberts, T. R. 1981. Sundasalangidae, a new family of minute freshwater salmoniform fishes from southeast Asia. Proceedings of the California Academy of Sciences 42:295–302.
- 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 61:539–542.
- Rüppell, W. P. E. S. 1837. Neue Wirbelthiere zu der Fauna von Abyssinien gehörig. Fische des Rothen Meeres. Siegmund Schmerber, Frankfurt am Main, 148 pp.
- Sato, Y. 1994. Cladistic-phenetic relationships and the evolutionary classification of the suborder Clupeoidei (Pisces, Teleostei). Unpublished PhD thesis. Tokyo Metropolitan University, Tokyo, 186 pp.
- Signeux, J. 1951. Notes paléoichthyologiques. *Diplomystus dubertreti*, une nouvelle espèce du Sénonien du Liban. Bulletin du Muséum National d'Histoire Naturelle 23:692–693.
- Signeux, J. 1964. *Gasteroclupea branisai*, clupéidé nouveau du Crétacé supérieur de Bolivie. Bulletin du Muséum National d'Histoire Naturelle 1964:291–297.
- Silva Santos, R. 1994. Ictiofaunula da Formação Codó, Cretáceo Inferior, com a descrição de um novo taxon – *Codoichthys carnavalii* (Pisces-Teleostei). Anais da Academia brasileira de Ciências 66:131–144.

- Silva Santos, R., and V. L. Silva Corrêa. 1985. Contribuição ao conhecimento da paléioictiofauna do Cretáceo no Brasil; pp. 169–174 in A. de Campos, C. S. Ferreira, I. M. Brito, and C. F. Viana (eds.), *Coletânea de Trabalhos Paleontológicos, Série Geologia* 27(2). Ministério das Minas e Energia-Departamento Nacional de Produção Mineral, Rio de Janeiro, Brazil, 656 pp.
- Soubrier, J., M. Steel, M. S. Y. Lee, C. Der Sarkissian, S. Guindon, S. Y. W. Ho, and A. Cooper. 2012. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution*. 29:3345–3358.
- Spix, J. B. von and L. Agassiz. 1829. *Selecta genera et species piscium quos in itinere per Brasiliam annis MDCCCXVII-MDCCCXX jussu et auspiciis Maximiliani Josephi I collegit et pingendos curavit Dr J. B. de Spix. Monachii*, 138 pp.
- Steindachner, F. 1876. *Ichthyologische Beiträge (V). Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Classe* 74:49-240.
- Sun, A.-L. 1956. *Paraclupea* – A genus of double-armored herrings from Chekiang. *Acta Palaeontologica Sinica* 4:413–418.
- Svetovidov, A. N. 1952. *Fauna of the USSR, Fishes. Herrings (Clupeidae)*. Academy of Sciences of USSR publishing, Moscow-Leningrad, 331 pp.
- Swainson, W. 1839. *On the natural history and classification of fishes, amphibians, and reptiles, or monocardian animals*. Spottiswoode & Co., London, 452 pp.
- Taverne, L. 1997. *Les clupéomorphes (pisces, teleostei) du cénomanien (crétace) de kipala (Kwango, Zaire): Ostéologie et phylogénie*. *Belgian journal of zoology* 127:75–98.

- Taverne, L. 2004. Les poissons cre'tace's de Nardo`. 18°. *Pugliachupea nolardi* gen. et sp. nov. (Teleostei, Clupeiformes, Clupeidae). Bollettino del Museo Civico di Storia Naturale di Verona 28:17–28.
- Temminck, C. J. and H. Schlegel. 1846. Pisces; pp. 173–269 in: Siebold, P. F. de (ed.), Fauna Japonica, sive descriptio animalium, quae in itinere per Japoniam suscepto annis 1823-1830 collegit, notis, observationibus et adumbrationibus illustravit Ph. Fr. de Siebold. Lugduni Batavorum, 323 pp.
- Teugels, G. G. 2003. Denticipitidae; pp. 122–124 in D. Paugy, C. Lévêque, and G.G Teugels (eds.), The fresh and brackish water fishes of West Africa Volume 1. Coll. faune et flore tropicales 40. Institut de recherche de développement, Paris, France, Muséum national d'histoire naturelle, Paris, France and Musée royal de l'Afrique Central, Tervuren, Belgium, 457 pp.
- Than-Marchese, B. A., J. Alvarado-Ortega, W. A. Matamoros, and E. Velázquez-Velázquez. 2020. *Scombrochupea javieri* sp. nov., an enigmatic Cenomanian clupeomorph fish (Teleostei, Clupeomorpha) from the marine deposits of the Cintalapa Formation, Ocozocoautla, Chiapas, southeastern Mexico. Cretaceous Research 112:10.1016/j.cretres.2020.104448.
- Trifinopoulos, J., L. T. Nguyen, A. Von Haeseler, and B. Q. Minh. 2016. IQ-TREE web server: fast and accurate phylogenetic trees under maximum likelihood. Nucleic Acids Research 44:232–235.
- Valenciennes, A. 1847. In Cuvier, G. and A. Valenciennes. Histoire naturelle des poissons. Tome vingtième. Livre vingt et unième. De la famille des Clupéoïdes. Histoire Naturelle Des Poissons, 472 pp.

- Valenciennes, A. 1848. In Cuvier, G. and A. Valenciennes. Histoire naturelle des poissons. Tome vingt et unième. Suite du livre vingt et unième et des Clupéoïdes. Livre vingt-deuxième. De la famille des Salmonoïdes. 21:1–536.
- Van Hasselt, J. C. 1823. Uittreksel uit een'brief van Dr. JC van Hasselt, aan den Heer CJ Temminck. Algemeene Konsten Letter-Bode, II Deel 35:130–133.
- Vernygora, O. V. 2015. Two new species and a revised phylogeny of the Ellimmichthyiformes (Teleostei: Clupeomorpha). Unpublished MSc thesis. University of Alberta, Edmonton, 181 pp.
- Vernygora, O. and A. M. Murray. 2016. A new species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous of Morocco, and its phylogenetic relationships. *Journal of Vertebrate Paleontology* 36:e1031342.
- Vernygora, O., A. M. Murray, and M. V. Wilson. 2016. A primitive clupeomorph from the Albian Loon River Formation (Northwest Territories, Canada). *Canadian Journal of Earth Sciences* 53:331–342.
- Walbaum, J. J. 1792. Petri Artedi sueci genera piscium. In quibus systema totum ichthyologiae proponitur cum classibus, ordinibus, generum characteribus, specierum differentiis, observationibus plurimis. Redactis speciebus 242 ad genera 52. *Ichthyologiae pars III*. Ant. Ferdin. Rose, Grypeswaldiae. 723 pp.
- Whitehead, P. J. P. 1962. A contribution to the classification of clupeoid fishes. *Journal of Natural History* 5:737–750.
- Whitehead, P. J. P. 1963. A contribution to the classification of clupeoid fishes. *Annals and Magazine of Natural History* 5:737–750.

- Whitehead, P. J. P. 1972. A synopsis of the clupeoid fishes of India. *Journal of Marine Biology Association of India* 14:160–256.
- Whitehead, P. J. P. 1985. *FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part I. Chirocentridae, Clupeidae and Pristigasteridae.* *FAO Fisheries Synopsis* 125:1–303.
- Whitehead, P. J. P., G. J. Nelson, and T. Wongratana. 1988. *FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part 2. Engraulididae.* *FAO Fisheries Synopsis* 125:305–579.
- Whitley, G. P. 1940. Illustrations of some Australian fishes. *Australian Zoologist* 9:397–428.
- Whitley, G. P. 1951. New fish names and records. *Proceedings of the Royal Zoological Society of New South Wales v. for 1949-50*:61–68.
- Wilson, A. 1811. *Clupea* Heading. In *Reese's New Cyclopaedia. American edition. v. 9.* No pagination.
- Wilson, A. B., G. G. Teugels, and A. Meyer. 2008. Marine incursion: The freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. *PLoS ONE* 3:e1979.
- Wongratana, T. 1983. Diagnoses of 24 new species and proposal of a new name for a species of Indo-Pacific clupeoid fishes. *Japanese Journal of Ichthyology* 29:385–407.
- Woodward, A. S. 1895. *Catalogue of the Fossil Fishes in the British Museum (Natural History), Cromwell Road, S. W. Part III. Containing the Actinopterygian Teleostomi of the Orders Chondrostei (concluded), Protospondyli, Aetheospondyli, and Isospondyli (in part).* *British Museum (Natural History), London, 706 pp.*

- Yabumoto, Y. and T. Uyeno. 1982. Osteology of clupeiform fish, genus *Hyperlophus* (II).
Bulletin Kitakyushu Museum of Natural History 4:77–102.
- Yang, Z. 1995. A space-time process model for the evolution of DNA sequences. *Genetics*
139:993–1005.
- Zhang, M. M., J. J. Zhou, and D. R. Qing. 1985. Tertiary fish fauna from coastal region of Bohai
Sea. Academia Sinica Institute of Vertebrate Paleontology and Paleoanthropology
Memoirs 17:1–60.
- Zaragüeta-Bagils, R. 2004. Basal clupeomorphs and ellimmichthyiform phylogeny; pp. 391–404
in G. Arratia, G. Tintori, and A. Tintori (eds.), *Mesozoic Fishes 3 – Systematics,
Paleoenvironment and Biodiversity*. Verlag Dr. Friederich Pfeil, Munich, Germany, 703
pp.

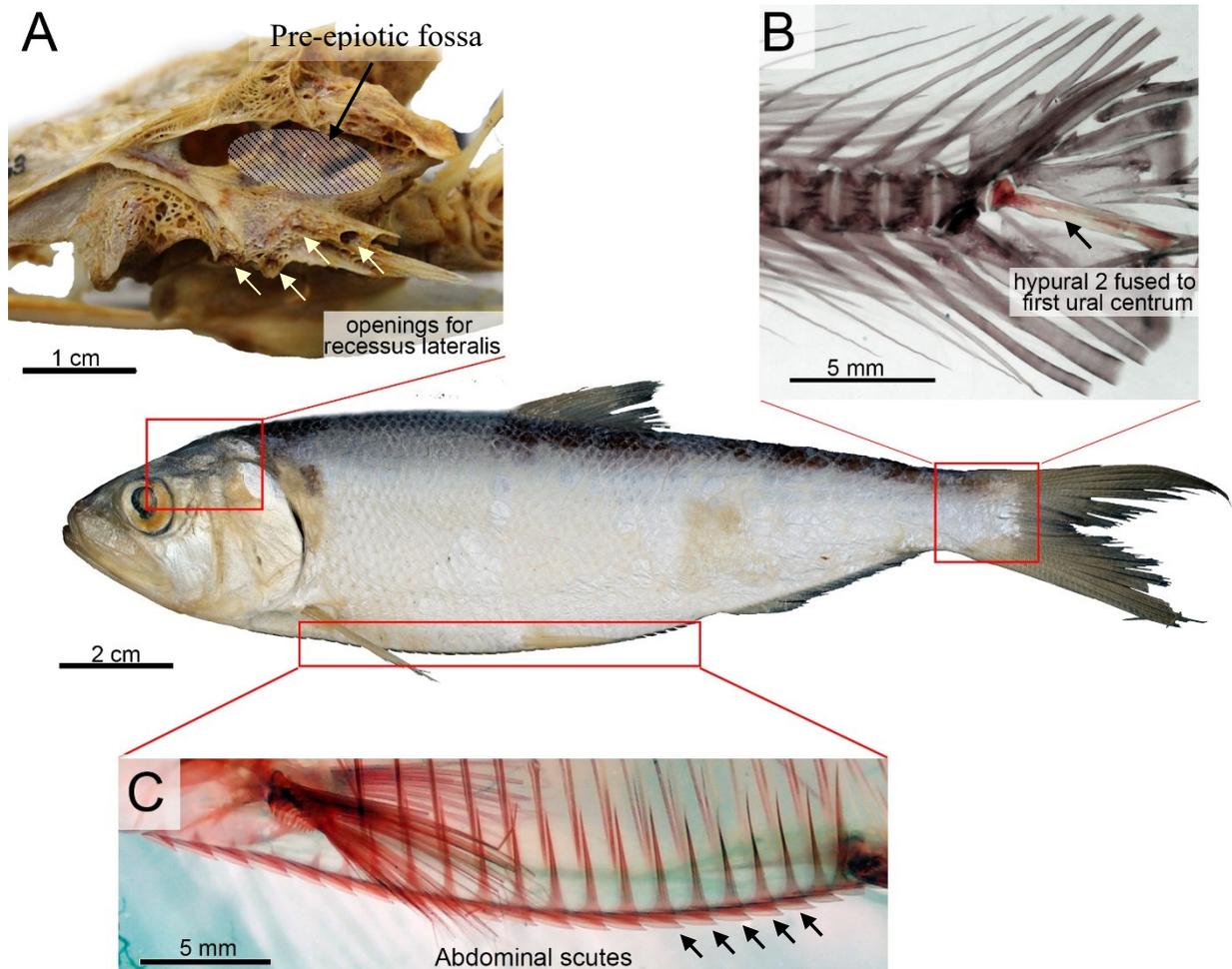


FIGURE 6-1. Some diagnostic features of Clupeomorpha. A – pre-epiotic fossa (stripped area) and openings for the recessus lateralis characteristic of Clupeiformes (white arrows); B – caudal skeleton showing fusion between the second hypural and the first ural centrum; C – abdominal scutes (black arrows).

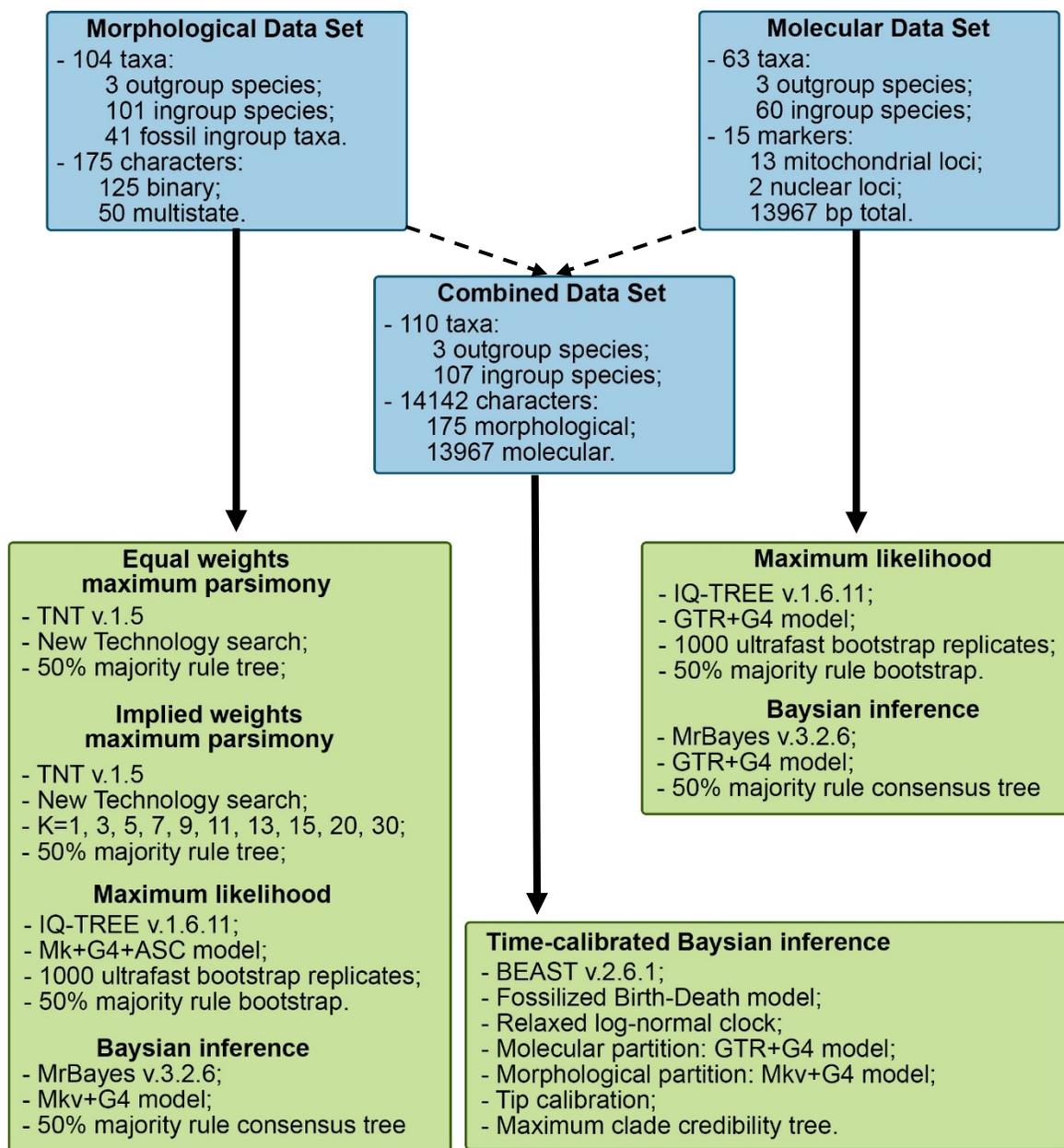


FIGURE 6-2. Flow chart summarizing the analytical pipeline of this study.

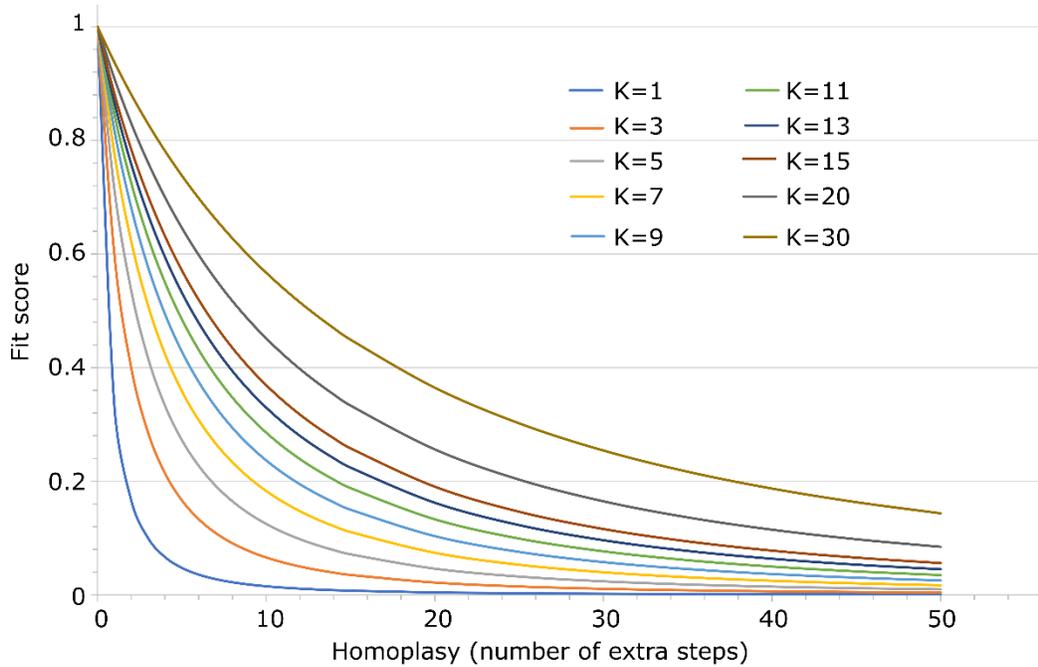


FIGURE 6-3. Effects of weighting strength function (K) on fit score of individual characters depending on their level of homoplasmy (number of extra steps required to explain a given topology).

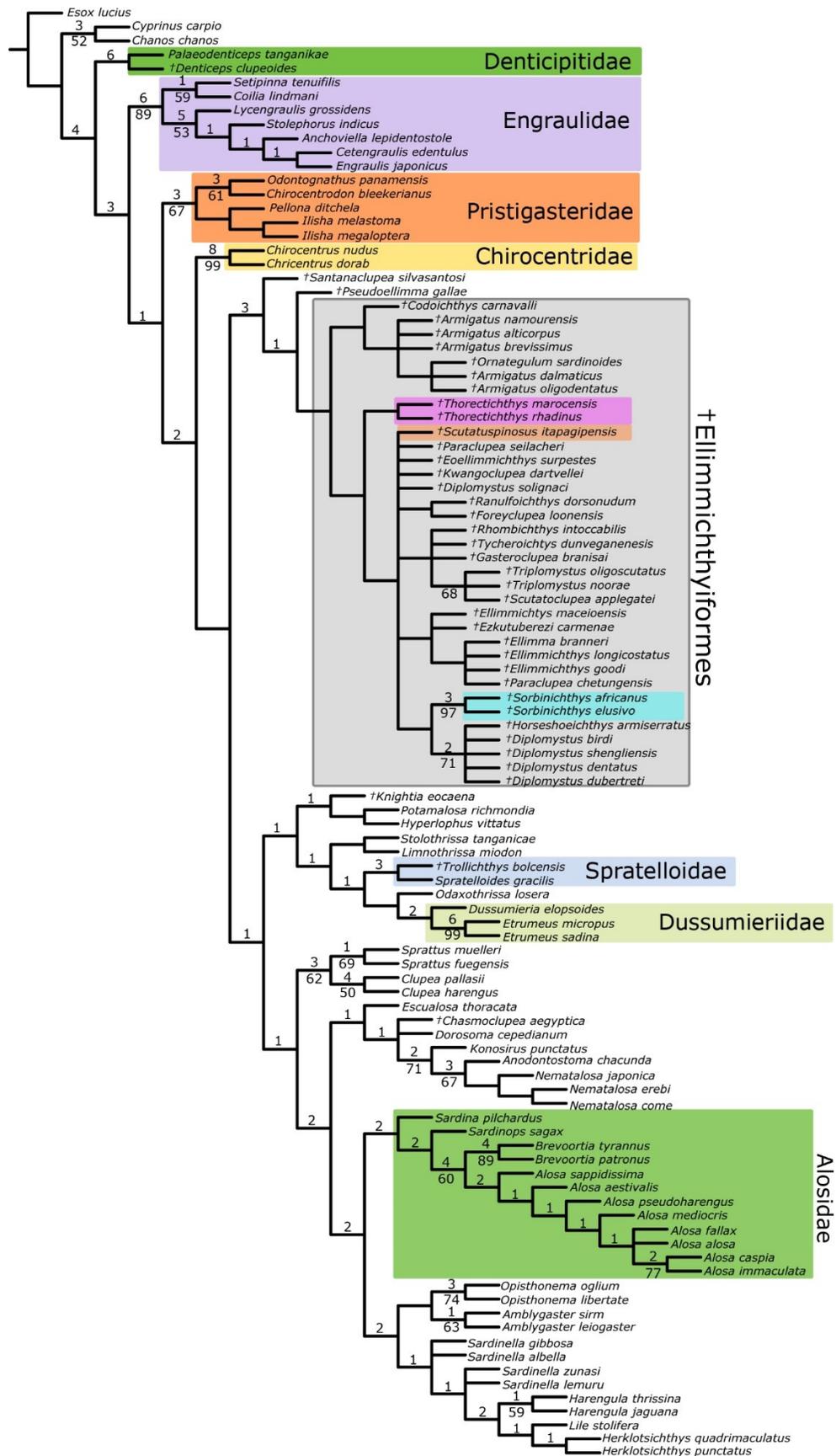


FIGURE 6-4. Equal weights maximum parsimony analysis of morphological data; 50% majority rule consensus tree of 462 most parsimonious trees; tree length = 840 steps, C. I. = 0.288, R. I. = 0.716. Bremer support values are indicated above branches; bootstrap support values >50% are shown below branches.

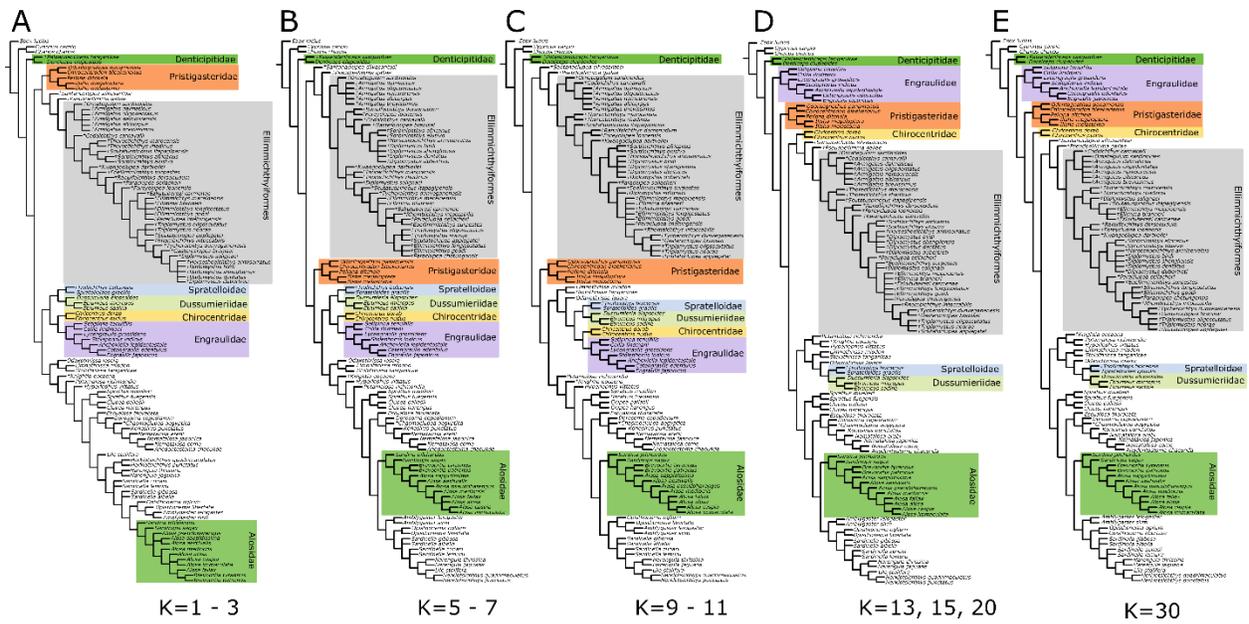


FIGURE 6-5. Implied weights parsimony analyses of morphological data. 50% majority rule consensus trees of the most fit trees recovered in each analysis for specified concavity values of the weights function: A - K=1-3; B - K=5-7; C - K=9-11; D - K=13, 15, 20; E - K=30.

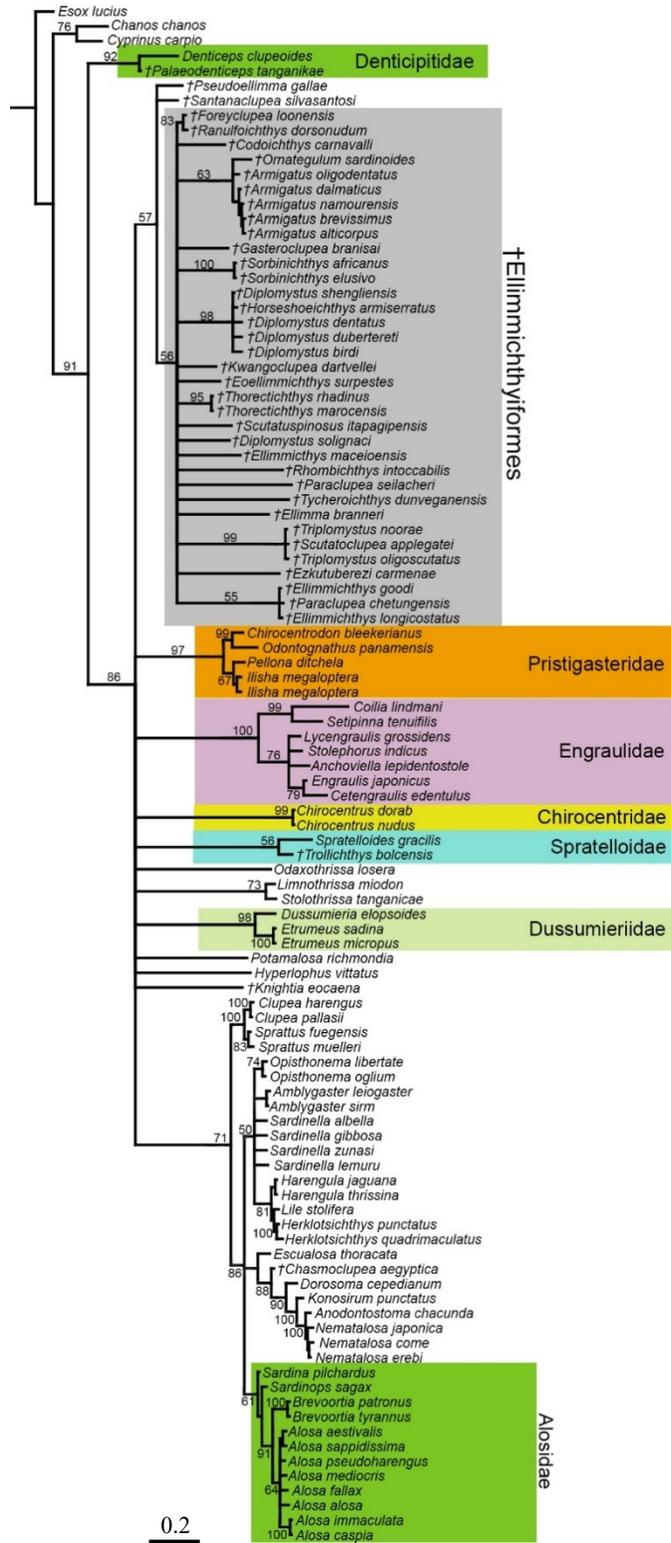


FIGURE 6-6. Maximum likelihood analysis of morphological data; 50% majority rule consensus tree of 1000 ultrafast bootstrap replicate trees. Numbers at nodes are ultrafast bootstrap support values (>50%). Branch lengths are proportional to the inferred number of substitutions. Scale bar indicates number of substitutions per site.

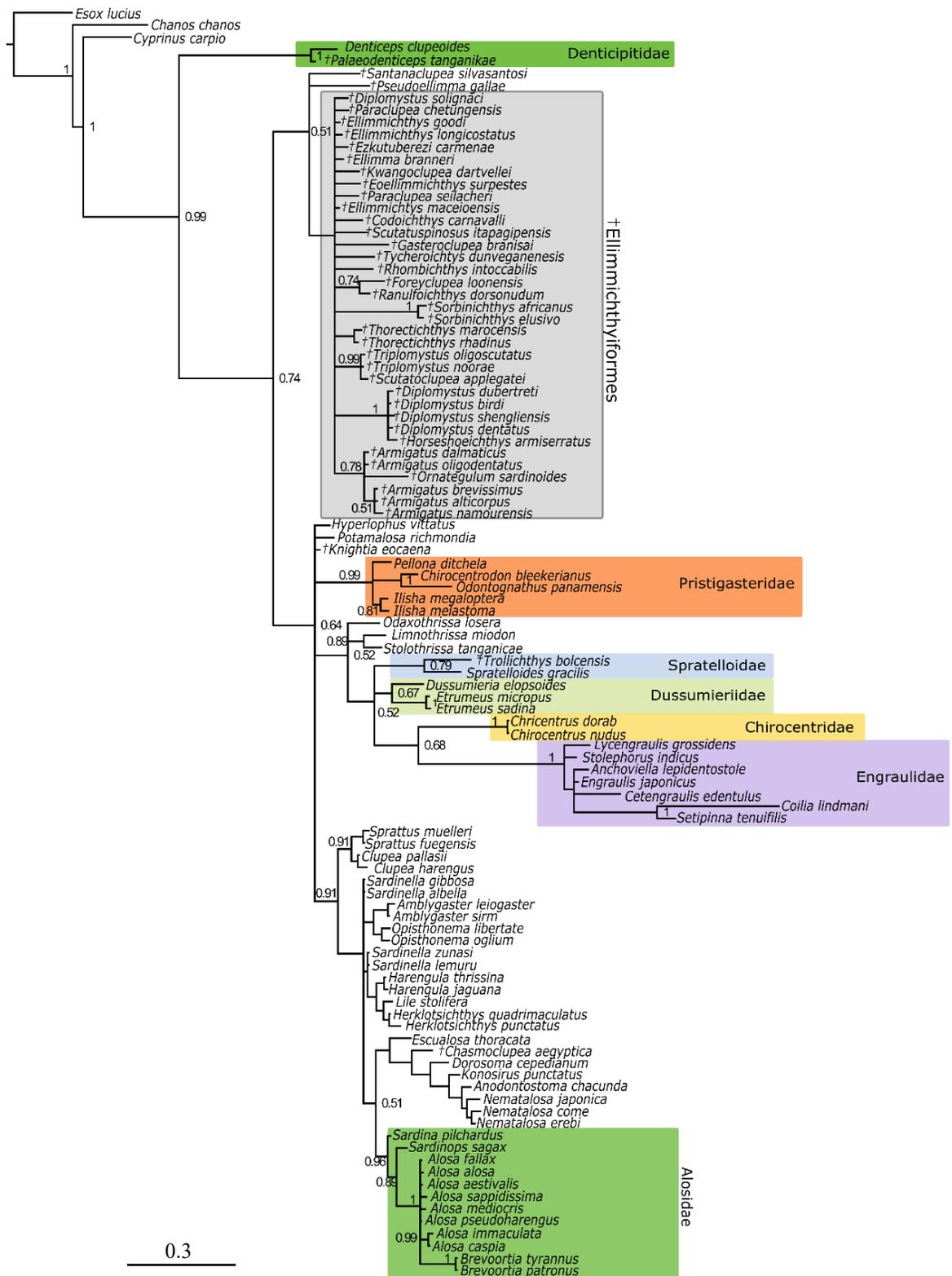


FIGURE 6-7. Bayesian inference analysis of morphological data; 50% majority rule consensus tree. Posterior probability values are indicated at nodes. Branch lengths are proportional to inferred number of substitutions. Scale bar indicates number of substitutions per site.

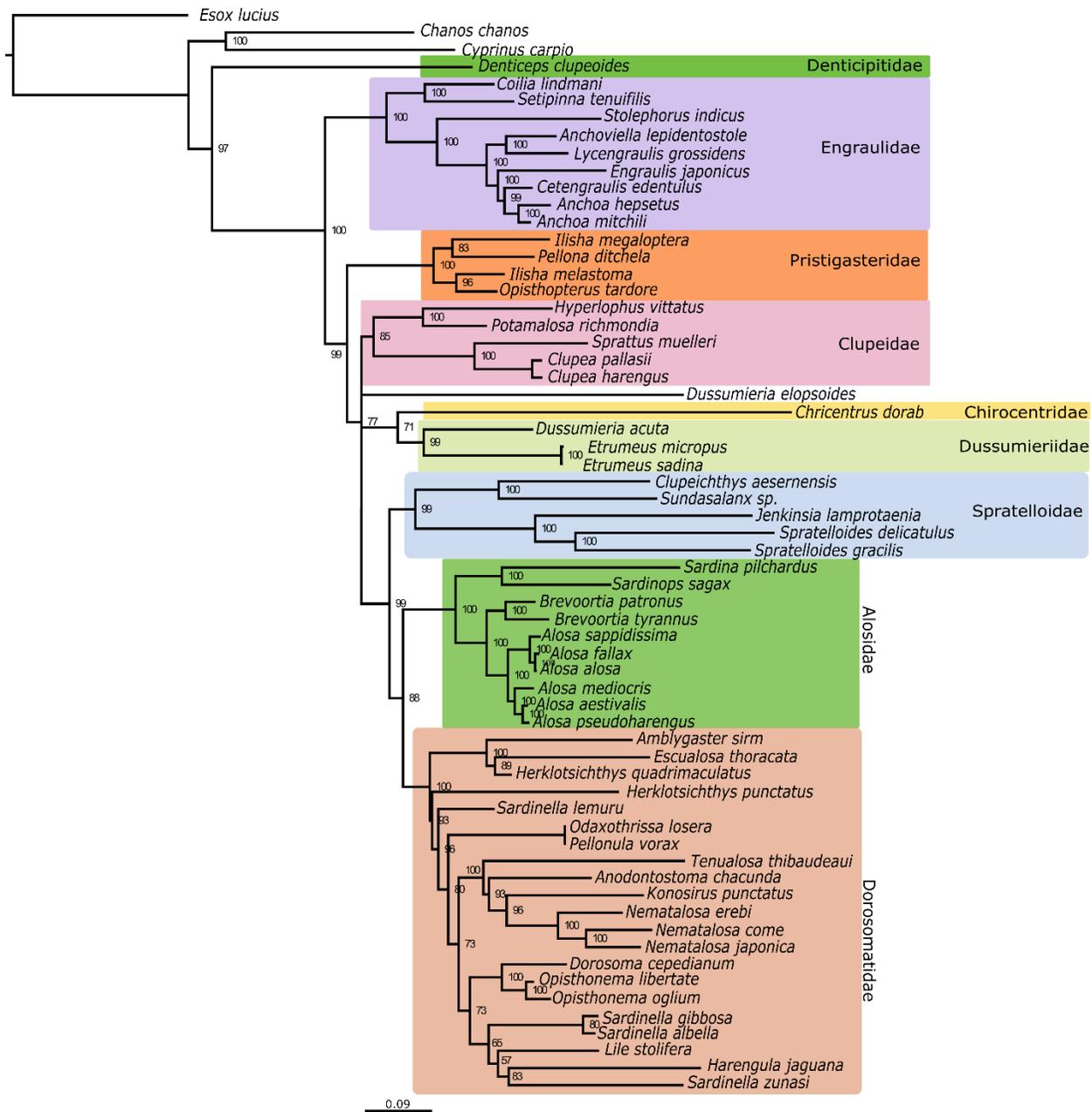


FIGURE 6-8. Maximum likelihood analysis of molecular data; 50% majority rule consensus tree of 1000 ultrafast bootstrap replicate trees. Numbers at nodes are ultrafast bootstrap support values (>50%). Scale bar indicates estimated number of substitutions per site.



FIGURE 6-9. Bayesian inference analysis of molecular data; 50% majority rule consensus tree. Posterior probability values are indicated at nodes. Branch lengths are proportional to inferred number of substitutions. Scale bar indicates number of substitutions per site.

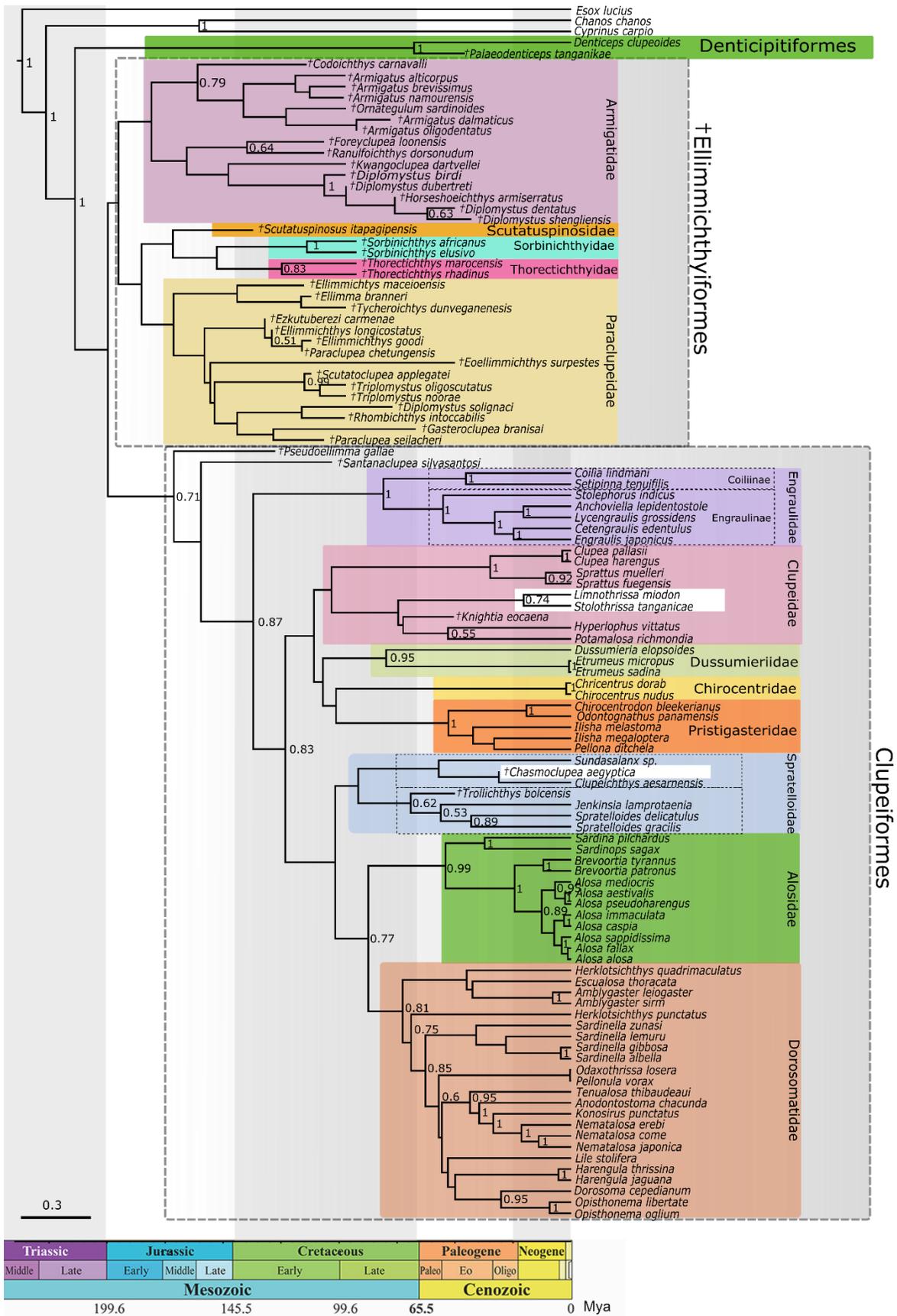


FIGURE 6-10. Time-calibrated phylogeny of Clupeomorpha. Maximum clade credibility tree recovered in Bayesian inference analysis of combined data using BEAST v.2.6.1. Posterior probability values are indicated at nodes (>0.5). Branch lengths are proportional to time. Scale bar indicates inferred number of substitutions per site. White boxes indicate taxa that were classified as members of groups different from clades where they appear in the maximum clade credibility tree.

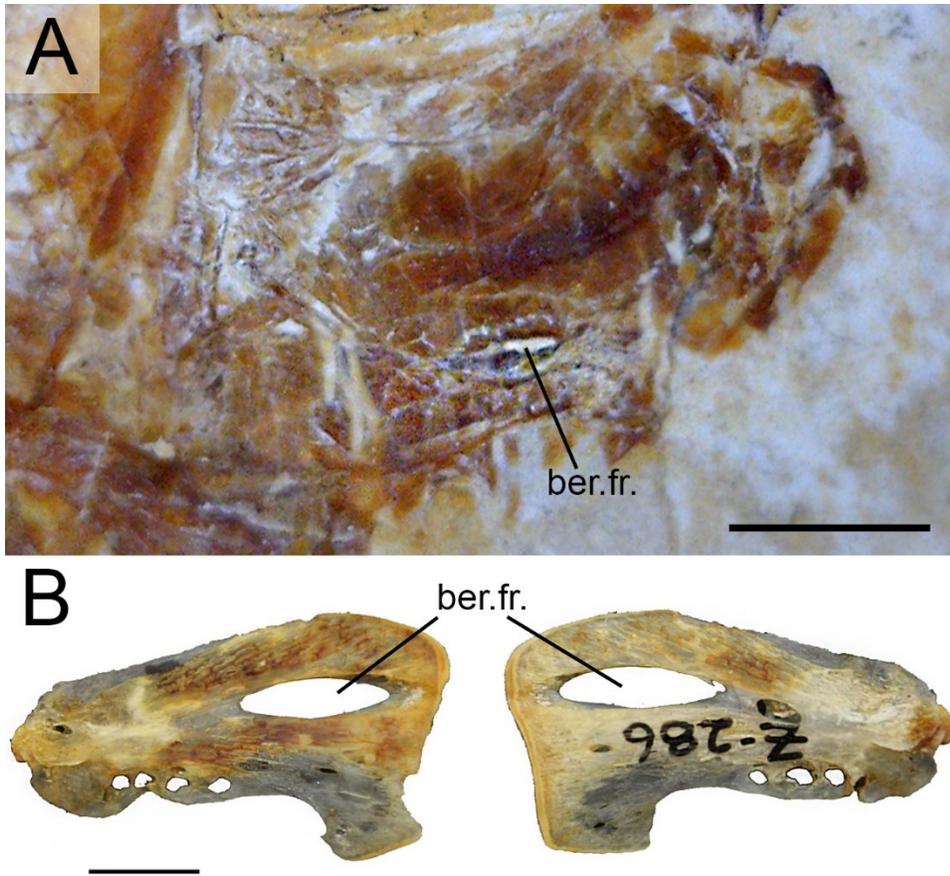


FIGURE 6-11. ‘Beryciform’ foramen (ber.fr.) in clupeomorph fishes. A – foramen in the anterior ceratohyal of †*Armigatus alticorpus* (TMP 1998.65.11); B – anterior ceratohyal of *Brevoortia tyrannus* (CMN Z-286) showing a large foramen in median (left) and lateral (right) views. Scale bar equals 5 mm.

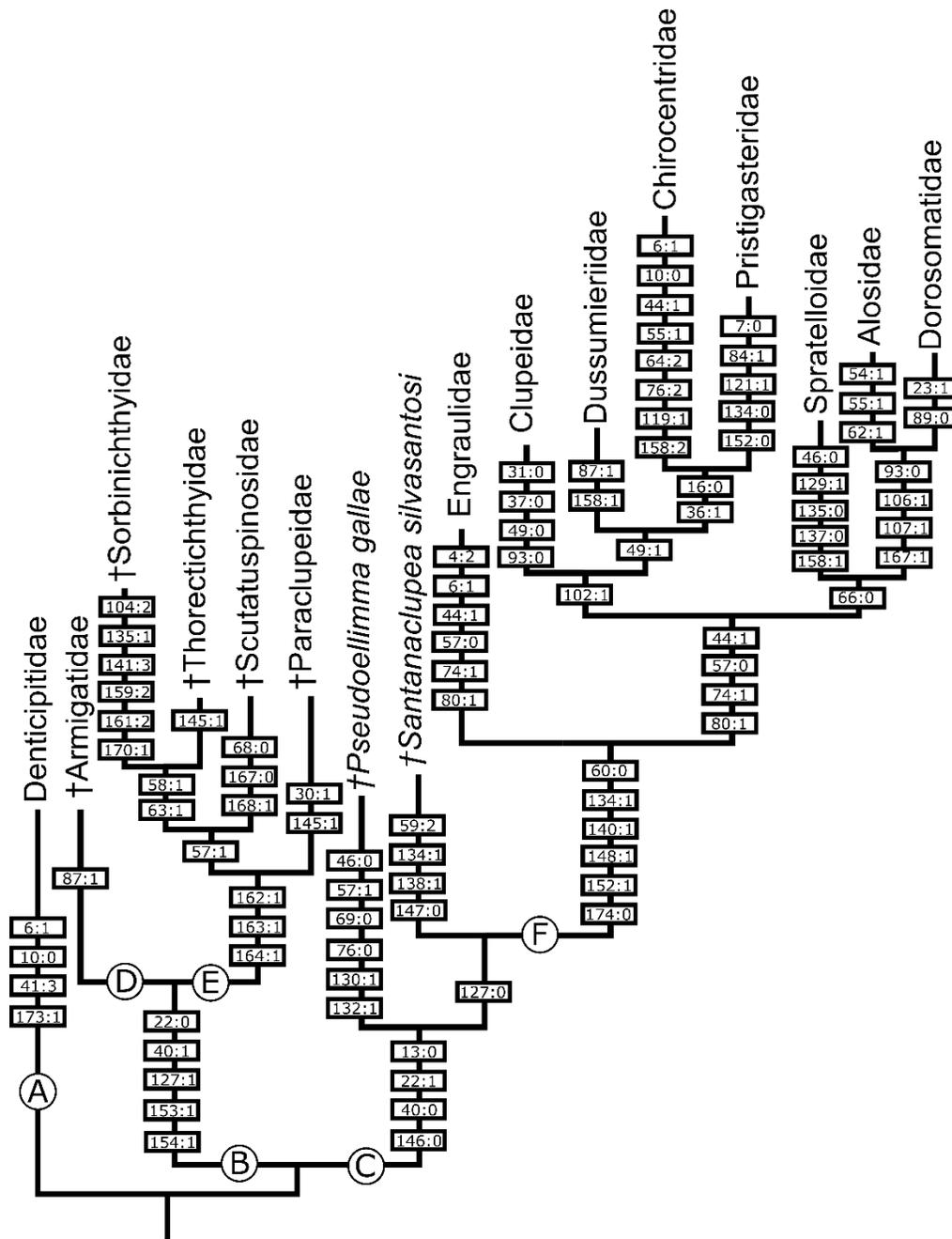


FIGURE 6-12. Simplified cladogram of Clupeomorpha based on combined evidence analysis.

Character states supporting each clupeomorph lineage are indicated in boxes along branches.

Letters indicate major clupeomorph lineages: A – Denticipitiformes; B – Ellimmichthyiformes;

C – Clupeiformes; D – Armigatoidei; E – Paraclupeoidei, F – Clupeoidei.

APPENDIX 6-1. Extant taxa included in this study, with information about their distribution, water salinity preference, accession numbers of examined museum specimens or literature sources.

Information about species ranges and environment is from Froese and Pauly (2019), Whitehead (1985), and Whitehead et al. (1988). Superscripts indicate type of material examined: cs – cleared and stained; ct – μ CT-scan; f – fresh or fluid-preserved; s – skeletal material.

Species	Range	Environment	Specimens/ Sources examined
Denticipitioidei			
Denticipitidae			
1. <i>Denticeps clupeioides</i> Clausen, 1959	Africa: rivers of Cameroon, Nigeria, Benin	Freshwater	CMN FI 1963-0117.1 ^f ; AMNH 235843 ^f ; MNHN-IC-1960-0391 ^{f, cs} ; 12 specimens
Clupeoidei			
Engraulidae			
1. <i>Anchoa hepsetus</i> (Linnaeus, 1758)	Western Atlantic	Marine, brackish	JFBM 47609 ^{ct} ; VIMS 35549 ^{ct} ; 2 specimens
2. <i>Anchoa mitchilli</i> (Valenciennes, 1848)	Western North and central Atlantic	Marine; freshwater; brackish; amphidromous	MNHN-IC-A-7930 ^f (lectotype), B-3097 ^f (paralectotype), 0000-3723 ^f (syntype); TCWC 16401.01 ^{ct} ; JFBM 47610 ^{ct} ; 5 specimens
3. <i>Anchoviella lepidentostole</i> (Fowler, 1911)	Western Atlantic	Marine, freshwater, brackish	CMN 69-3 ^f ; 3 specimens
4. <i>Cetengraulis edentulus</i> (Cuvier, 1829)	Western Atlantic: Caribbean to Brazil	Marine, brackish	MNHN-IC-0000-0899 ^f (neotype); 1 specimen

5. <i>Coilia lindmani</i> Bleeker, 1857	Asia: rivers of Vietnam, Thailand, Sumatra, Malaysia, Kalimantan	Freshwater, brackish	MNHN-IC-1923-0205 ^f , 1923-0203 ^f (syntypes) 2 specimens
6. <i>Engraulis japonicus</i> Temminck & Schlegel, 1846	Western Pacific	Marine	JFBM 48201 ^{ct} 1 specimen
7. <i>Lycengraulis grossidens</i> (Spix and Agassiz, 1829)	Western Atlantic	Marine, freshwater, brackish, anadromous	MNHN-IC-0000-3720 ^f (lectotype), 0000-1102 ^f (paralectotype) 2 specimens
8. <i>Setipinna tenuifilis</i> (Valenciennes, 1848)	Eastern Indian Ocean	Marine, freshwater, brackish, amphidromous	MNHN-IC-0000-3731 ^f (lectotype), B-3102 ^f (paralectotype); JFBM 48806 ^{ct} ; CMN 77-0025 ^f 5 specimens
9. <i>Stolephorus indicus</i> (Van Hasselt, 1823)	Indo-Pacific	Marine, brackish, oceanodromous	JFBM 48609 ^{ct} 1 specimen

Pristigasteridae

1. <i>Chirocentrodon bleekermanus</i> (Poey, 1867)	Western Atlantic	Marine, brackish	Whitehead, 1985; Sato, 1994
2. <i>Ilisha megaloptera</i> (Swainson, 1839)	Indo-Pacific	Marine, freshwater, brackish, anadromous	MNHN-IC-0000-3708 ^f (lectotype), 0000-3709 ^f , 0000-3707 ^f (paralectotypes) 3 specimens
3. <i>Ilisha melastoma</i> (Bloch and Schneider, 1801)	Indo-Pacific	Marine, brackish, amphidromous	MNHN-IC-0000-3711 ^f (lectotype), 0000-3936 ^f (paralectotype); JFBM 48661 ^{ct} ; VIMS 7532 ^{ct} . 4 specimens
4. <i>Odontognathus panamensis</i> (Steindachner, 1876)	Eastern Central Pacific:	Marine, brackish	Whitehead, 1985; Whitehead and Rodriguez-

	Honduras to Panama Bay		Sanchez, 1995; Sato, 1994
5. <i>Opisthopterus tardoore</i> (Cuvier, 1829)	Indo-West Pacific	Marine, brackish, amphidromous	MNHN-IC-0000- 1688 ^f (neotype); JFBM 48658 ^{ct} 2 specimens
6. <i>Pellona ditchela</i> Valenciennes, 1847	Indo-West Pacific: East Africa to Papua New Guinea	Marine, freshwater, brackish, anadromous	UAMZ F6353.3 ^{ct} ; CMN 81-277 ^f 2 specimens
<hr/>			
Chirocentridae			
1. <i>Chirocentrus dorab</i> (Forsskål, 1775)	Indo-Pacific	Marine, brackish, amphidromous	UWFC 021129 ^{ct} ; CMN 77-0025 ^f 2 specimens
2. <i>Chirocentrus nudus</i> Swainson, 1839	Indo-West Pacific	Marine	ROM R8500 ^s ; CMN 81-0277 ^f 2 specimens
<hr/>			
Dussumieriidae			
1. <i>Dussumieria acuta</i> Valenciennes, 1847	Indo-Pacific	Marine, brackish, freshwater	MNHN-IC-0000- 3697 ^f (lectotype); JFBM 47630 ^{ct} ; CMN 77-0023 ^f 3 specimens
2. <i>Dussumieria elopsoides</i> Bleeker, 1849	Indo-Pacific	Marine	MNHN-IC-1966- 0258 ^f , 1966-0259 ^f , 1966-0260 ^f , 0000- 3997 ^f , 1966-0264 ^f (syntypes); UAMZ F4499 ^{cs} 7 specimens
3. <i>Etrumeus micropus</i> (Temminck and Schlegel, 1846)	Western Pacific: from south coast of Japan to South China Sea	Marine	UWFC 012792 ^{ct} ; AMNH 2670 ^f 2 specimens
4. <i>Etrumeus sadina</i> (DeKay, 1842)	Northwestern Atlantic: from the Bay of Fundy to	Marine, oceanodromous	VIMS 35550 ^{ct} ; CMN 54-629 ^f ; CMN 71-0664 ^f ; AMNH 228260 ^f

	the Gulf of Mexico		7 specimens
<hr/> Clupeidae <hr/>			
1. <i>Alosa aestivalis</i> (Mitchill, 1814)	Western Atlantic: Florida to Nova Scotia	Marine, freshwater, brackish, anadromous	UWFC 020612 ^{ct} ; CMN 79-0998 ^s 2 specimens
2. <i>Alosa alosa</i> (Linnaeus, 1758)	Eastern Atlantic: Norway to Mauritania	Marine, freshwater, brackish, anadromous	MNHN-IC-0000- 3676 ^f (lectotype), 0000-5434 ^f , 0000- 3398 ^f , 0000-3133 ^f (paralectotypes); AMNH 230430 ^f 5 specimens
3. <i>Alosa caspia</i> (Eichwald, 1838)	Black Sea, Sea of Azov, Caspian Sea	Marine, freshwater, brackish, anadromous	MNHN-IC-1987- 0708 ^f (paratype); CMN 59-332 ^f , uncat. material ^{f, s} 10 specimens
4. <i>Alosa fallax</i> (Lacepede, 1803)	Eastern Atlantic: Norway to Morocco	Marine, freshwater, brackish, anadromous	MNHN-IC-0000- 3750 ^f , 0000-3752 ^f , 0000-6263 ^f (syntypes); AMNH 1150 ^f 4 specimens
5. <i>Alosa immaculata</i> Bennett, 1835	Black Sea and Sea of Azov	Marine, freshwater, brackish, anadromous	Uncat. material ^{f, s} 30 specimens
6. <i>Alosa mediocris</i> Mitchill, 1814	Western Atlantic: Florida to New Brunswick	Marine, freshwater, brackish, anadromous	CMN 65559 ^f , 224878 ^f , Z-470 ^s . 5 specimens
7. <i>Alosa pseudoharengus</i> (Wilson, 1811)	Western Atlantic: South Carolina to Labrador	Marine, freshwater, brackish, anadromous	MNHN-IC-A- 7641 ^f (lectotype), B-3077 ^f (paralectotype); UWFC 004166 ^{ct} ; AMNH 41835 ^f , 70599 ^f , 75977 ^f ; CMN 73-493 ^f , 59- 332 ^f , 65-130-5 ^s , Z- 4276 ^s , 66-449 ^s

			15 specimens
8. <i>Alosa sappidissima</i> (Wilson, 1811)	Western Atlantic: Florida to Labrador	Marine, freshwater, brackish, anadromous	VIMS 35551 ^{ct} ; AMNH 49345 ^f , 221607 ^f , 38917 ^f ; CMN 86-3 ^f , Z-365 ^s , 363 ^s ; AMNH F 4510 ^{cs} , 4510 ^f 12 specimens
9. <i>Amblygaster sirm</i> (Walbaum, 1792)	Indo-West Pacific	Marine	UWFC 006291 ^{ct} 1 specimen
10. <i>Amblygaster leiogaster</i> (Valenciennes, 1847)	Indo-West Pacific	Marine	AMNH 8164 ^f 2 specimens
11. <i>Anodontostoma chacunda</i> (Hamilton-Buchanan, 1822)	Indo-West Pacific	Marine, freshwater, brackish, anadromous	JFBM 48955 ^{ct} ; UWFC 015142 ^{ct} 2 specimens
12. <i>Brevoortia patronus</i> Goode, 1878	Western Central Atlantic: Gulf of Mexico	Marine	TCWC 15424.02 ^{ct} 1 specimen
13. <i>Brevoortia tyrannus</i> (Latrobe, 1802)	Western Atlantic: Florida to Nova Scotia	Marine, brackish	UWFC 006269 ^{ct} ; VIMS 35552 ^{ct} ; CMN Z-286 ^s 5 specimens
14. <i>Clupea harengus</i> Linnaeus, 1758	North Atlantic	Marine, brackish, oceanodromous	CMN 75-114 ^s , 75-117 ^s ; MNHN-IC-2004-1489 ^f , 2004-1490 ^f ; uncat. material 10 specimens
15. <i>Clupea pallasii</i> Valenciennes, 1847	Northern Pacific	Marine, freshwater, brackish	UWFC 004669 ^{ct} , CMN 87-0021 ^s 2 specimens
16. <i>Clupeichthys aesarnensis</i> Wongratana, 1983	Asia: Laos, Thailand, Cambodia	Freshwater, potadromous	JFBM 48896 ^{ct} 1 specimen
17. <i>Dorosoma cepedianum</i> (Lesueur, 1818)	Northwest Atlantic: Gulf of Mexico to Great Lakes	Marine, freshwater, brackish, anadromous	MNHN-IC-0000-0004 ^f , 0000-3630 ^f , 0000-3631 ^f , 0000-3633 ^f (syntypes); TCWC 16900.14 ^{ct} ,

			VIMS 18573 ^{ct} ; AMNH 243379 ^f , 75-102 ^s ; UAMZ F2455 ^{cs} 10 specimens
18. <i>Escualosa thoracata</i> (Valenciennes, 1847)	Indo-West Pacific	Marine, freshwater, brackish, amphidromous	MNHN-IC-0000- 3172 ^f , 0000-3173 ^f (lectotypes), B- 3076 ^f , B-3075 ^f , 0000-3229 ^f (paralectotypes); JFBM 48651 ^{ct} ; AMNH 32502 ^f 19 specimens
19. <i>Harengula jaguana</i> Poey, 1865	Western Atlantic	Marine, brackish	VIMS 35602 ^{ct} ; TCWC 16904.09 ^{ct} , 16904.28 ^{ct} ; AMNH 49463 ^f ; CMN 69- 3 ^f ; MNHN-IC- 2002-0937 ^f , 2002- 0967 ^f , 2002-0984 ^f 12 specimens
20. <i>Harengula thrissina</i> (Jordan and Gilbert, 1882)	Eastern Pacific	Marine, brackish	CMN 68-1078 ^f 1 specimen
21. <i>Herklotsichthys punctatus</i> (Rüppell, 1837)	Red Sea and Gulf of Aden, Eastern Mediterranean	Marine	CMN 81-1149 ^f ; MNHN-IC-1977- 1078 ^f . 4 specimens
22. <i>Herklotsichthys quadrimaculatus</i> (Rüppell, 1837)	Indo-Pacific	Marine, freshwater, brackish	MNHN-IC-0000- 0895 ^f (holotype), 0000-0900 ^f , 0000- 0666 ^f (lectotypes), 0000-3106 ^f , B- 3073 ^f (paralectotypes); USNM 327940 ^{ct} 6 specimens
23. <i>Hyperlophus vittatus</i> (Castelnau, 1875)	Western Pacific	Marine, brackish, amphidromous	MNHN-IC-1897- 0391 ^f , 0000-3755 ^f (syntypes); AMNH 55288 ^f 10 specimens

24. <i>Jenkinsia lamprotaenia</i> (Gosse, 1851)	Western Central Atlantic	Marine	UAMZ F4865 ^{ct,f} 2 specimens
25. <i>Konosirus punctatus</i> (Temminck and Schlegel, 1846)	Western Pacific	Marine, brackish, oceanodromous	Whitehead, 1985; Sato, 1994
26. <i>Lile stolifera</i> Jordan and Gilbert, 1882	Eastern Pacific	Marine	MNHN-IC-1887-0380 ^f ; CMN 68-1686 ^f ; UWFC 022631 ^{ct} 9 specimens
27. <i>Limnothrissa. miodon</i> (Boulenger, 1906)	Africa	Freshwater	Whitehead, 1985; Sato, 1994
28. <i>Nematalosa come</i> (Richardson, 1846a)	Western Pacific: coasts of Australia	Marine	JFBM 48151 ^{ct} 1 specimen
29. <i>Nematalosa erebi</i> (Günther, 1868)	Oceania and Asia	Freshwater, brackish, potamodromous	Whitehead, 1985; Sato, 1994
30. <i>Nematalosa japonica</i> Regan, 1917	Western Pacific: Japan to Taiwan Island	Marine	AMNH 28124 ^f 1 specimen
31. <i>Odaxothrissa losera</i> Boulenger, 1899	Africa	Freshwater	Grande, 1985
32. <i>Opisthonema libertate</i> (Günther, 1867)	Eastern Pacific: California (USA) to Peru	Marine	UWFC 017512 ^{ct} 1 specimen
33. <i>Opisthonema oglinum</i> (Lesueur, 1818)	Western Atlantic: Gulf of Maine (USA) to Brazil	Marine	VIMS 35555 ^{ct} ; TCWC 4226.01 ^{ct} ; MNHN-IC-0000-3735 ^f (lectotype), B-3088 ^f , 0000-3938 ^f (paralectotypes), 2002-0945 ^f 6 specimens
34. <i>Pellonula vorax</i> Günther, 1868	Africa	Marine, freshwater,	OS TNC17-325 ^{ct} 1 specimen

		brackish, anadromous	
35. <i>Potamalosa richmondia</i> (Macleay, 1879)	Oceania: Southeastern Australia	Marine, freshwater, brackish; catadromous	Whitehead, 1985; Sato, 1994
36. <i>Sardina pilchardus</i> (Walbaum, 1792)	Northeast Atlantic	Marine, freshwater, brackish, oceanodromous	MNHN-IC-2004- 0582 ^f , 2004- 0583 ^f ; uncat. material. 5 specimens
37. <i>Sardinella albella</i> (Valenciennes, 1847)	Indo-West Pacific	Marine	JFBM 48875 ^{ct} ; AMNH 251484 ^f ; MNHN-IC-0000- 0665 ^f (lectotype), 0000-3231 ^f (paralectotype) 7 specimens
38. <i>Sardinella gibbosa</i> (Bleeker, 1849)	Indo-West Pacific	Marine	MNHN-IC-1966- 0343 ^f , 1966-0344 ^f (syntypes); JFBM 47442 ^{ct} 3 specimens
39. <i>Sardinella lemuru</i> Bleeker, 1853a	Indo-West Pacific	Marine, oceanodromous	JFBM 47629 ^{ct} 1 specimen
40. <i>Sardinella zunasi</i> (Bleeker, 1854)	Western Pacific	Marine, oceanodromous	Whitehead, 1985; Sato, 1994
41. <i>Sardinops sagax</i> (Jenyns, 1842)	Indo-Pacific	Marine, oceanodromous	UWFC 047455 ^{ct} , 004853 ^{ct} ; AMNH 7406 ^f , 33482 ^f
42. <i>Spratelloides delicatulus</i> (Bennett, 1831)	Indo-Pacific	Marine	UAMZ F4505 ^{ct,f} 4 specimens
43. <i>Spratelloides gracilis</i> (Temminck and Schlegel, 1846)	Indo-Pacific	Marine	Whitehead, 1985; Sato, 1994
44. <i>Sprattus fuegensis</i> (Jenyns, 1842)	Southwest Atlantic	Marine, oceanodromous	CMN 86-0690 ^f 5 specimens
45. <i>Sprattus muelleri</i> (Klunzinger, 1880)	New Zealand.	Marine	Whitehead, 1985; Sato, 1994

46. <i>Stolothrissa tanganicae</i> Regan, 1917	Africa	Freshwater	Whitehead, 1985; Sato, 1994
47. <i>Tenualosa thibaudeaui</i> (Durand, 1940)	Asia: Laos, Thailand, Vietnam, Cambodia	Freshwater, potadromous	JFBM 48961 ^{ct} 1 specimen
<hr/>			
Outgroups			
<hr/>			
Ostariophysii: Gonorynchiformes: Chanidae			
1. <i>Chanos chanos</i> (Forsskål, 1775)	Indo-Pacific	Marine, freshwater; amphidromous	UAMZ F8463 ^s 1 specimen
<hr/>			
Ostariophysii: Cypriniformes: Cyprinidae			
2. <i>Cyprinus carpio</i> Linnaeus, 1758	Eurasia	Freshwater, brackish	UAMZ ICH- F8557 ^s , 8811 ^s 2 specimens
<hr/>			
Euteleostei: Protacanthopterygii: Esociformes: Esocidae			
3. <i>Esox lucius</i> Linnaeus, 1758	Circumpolar	Freshwater, brackish	UAMZ ICH- F305 ^s 1 specimen
<hr/>			

APPENDIX 6-2. Fossil taxa included in this study, with information about their temporal and geographic occurrence, palaeoenvironmental settings at site of occurrence, and age ranges used for time calibration analysis.

Species	Occurrence	Environment	Geological Age	Calibration range (mya)	Specimens/Sources examined
<i>Armigatus alticorpus</i> Forey et al., 2003	Lebanon	Marine	Late Cretaceous	93.9 – 100.5	NHMUK P.63134 (holotype), TMP 1998.65.11 5 specimens
<i>Armigatus brevissimus</i> (Blainville, 1818)	Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	UALVP 5087, 17620, 47258; MNHN.F.HAK481 4 specimens
<i>Armigatus dalmaticus</i> Murray et al., 2016	Dalmatia, Croatia	Marine, reef platform	Late Cretaceous, Senonian (Campanian)	72.1 – 83.6	CNHM 9423 (holotype), 9259, 9277, 9287 4 specimens
<i>Armigatus namourensis</i> Forey et al., 2003	Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	Forey et al., 2003 1 specimen
<i>Armigatus oligodentatus</i> Vernygora and Murray, 2016	Morocco	Marine	Late Cretaceous (Cenomanian/Turonian)	90 – 95	UALVP 51679 (holotype), 47146, 47155, 51602, 51622, 51623, 51680, 51681 8 specimens
<i>Chasmoclupea aegyptica</i> Murray et al., 2005	Egypt	Freshwater, fluvial deposits	Oligocene	28.4 – 33.9	DPC6202 (cast) 1 specimen
<i>Codoichthys carnivalii</i> Silva Santos, 1994	São Luís-Grajaú Basin, NE Brazil	Lacustrine-estuarine	Early Cretaceous (late Aptian)	113 – 115	De Figueiredo and Ribeiro, 2016
<i>Diplomystus birdi</i> Woodward, 1895	Hakel, Mount Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	CMN FV 30564; NHMUK P.83, 96a 3 specimens
<i>Diplomystus dentatus</i> Cope, 1877	Wyoming, USA	Freshwater	Early Eocene	48.5 – 53.5	UALVP 17830, TMP 86.224.81, 86.224.85, 86.224.89; CMN 8881, 2107, 41639 7 specimens

<i>Diplomystus dubertreti</i> Signeux, 1951	Sahel Alma, Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	MNHN F.SHA2053 (holotype), 2052 (paratype) 2 specimens
<i>Diplomystus shengliensis</i> Zhang et al., 1985	Shandong Province, China	Freshwater	Middle Eocene	38 – 48	Chang and Maisey, 2003
' <i>Diplomystus</i> ' <i>solignaci</i> Gaudant and Gaudant, 1971	Tunisia	Marine	Late Cretaceous (Senonian)	65.5 – 89.9	Gaudant and Gaudant, 1971
<i>Ellimma branneri</i> (Jordan, 1910)	Sergipe Basin, Brazil	Brackish and marine	Early Cretaceous (Aptian – Albian)	100.5 – 125	Chang and Maisey, 2003 NHMUK P.13375, 13376, 13377, 13378, 13382, 13392, 13396. 7 specimens
<i>Ellimmichthyes goodi</i> Eastman, 1912	Equatorial Guinea, Central Africa	Freshwater	Early Cretaceous (Aptian - Albian)	100.5 – 125	
<i>Ellimmichthyes maceioensis</i> Malabarba et al., 2004	Maceio Formation, Brazil	Lagoon with marine and freshwater incursion	Early Cretaceous (Aptian – early Albian)	110 – 120	Malabarba et al., 2004
<i>Ellimmichthys longicostatus</i> (Cope, 1886)	Bahia, Brazil	Estuarine	Early Cretaceous (Hauterivian– Barremian)	125 – 132.9	NHMUK P.7109, 10350, 8254, 8256, 9611 5 specimens
<i>EOellimmichthys surpestes</i> Marammà et al., 2019	Bolca, Italy		late early Eocene, late Ypresian	49 – 51	Marammà et al., 2018
<i>Ezcutuberezi carmenae</i> Poyato-Ariza et al., 2000	Basque Country, Spain	Deltaic and lacustrine	Early Cretaceous (Valanginian - Barremian)	125 – 139.8	Poyato-Ariza et al., 2000
<i>Foreyclupea loonensis</i> Vernygora et al., 2016	NWT, Canada	Marine	Early Cretaceous (Albian)	100.5 – 113	UALVP 17535 (holotype) 1 specimen
<i>Gasteroclupea branisai</i> Signeux, 1964	Cyara, Bolivia	Freshwater (or brackish)	Late Cretaceous – Paleocene (Maastrichtian-Danian)	62 – 72.1	Marammà and Carnevale, 2017
<i>Horseshoeichthys armiserratus</i> Newbrey et al., 2010	Alberta, Canada	Freshwater	Late Cretaceous (Santonian — Maastrichtian)	66 – 86.3	TMP 2001.045.0093 (holotype) 1 specimen
<i>Knightia eocaena</i> Jordan, 1907	Green River Formation, Wyoming, USA	Freshwater	Eocene	48.5 – 53.5	Grande, 1982

<i>Kwangoclupea dartevellei</i> (Casier, 1965)	Congo, Africa	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	Taverne, 1997
<i>Ornategulum sardinioides</i> (Pictet, 1850)	Hakel, Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	Forey, 1973
<i>Palaeodenticeps tanganyikae</i> Greenwood, 1960	Tanganyika Territory, Tanzania, East Africa	Lacustrine shales	middle – late Tertiary	45.66 – 46.0	WM 352/92, 301/96, 432/96, 128/96, 100/96
<i>Paraclupea chetungensis</i> Sun, 1956	Chawan Formation, China	Freshwater	Early Cretaceous	112.4 – 120	Chang and Grande, 1997
<i>Paraclupea seilacheri</i> Alvarado-Ortega and Melgarejo-Damián, 2017	Puebla, Mexico	Marine lagoon with freshwater incursions	Early Cretaceous (Albian)	100.5 – 113	Alvarado-Ortega and Melgarejo-Damián, 2017
<i>Pseudoellimma gallae</i> De Figueiredo, 2009	Alagoas, Brazil	Brackish lake with irregular inputs of marine water flow	Early Cretaceous (Barremian)	125 – 129.4	De Figueiredo, 2009
<i>Ranulfoichthys dorsonudum</i> Alvarado-Ortega, 2014	Puebla, Mexico	Marine	Early Cretaceous (Albian)	100.5 – 113	Alvarado-Ortega, 2014
<i>Rhombichthys intoccabillis</i> Khalloufi et al., 2010	Ein Yabrud, Palestine	Shallow, low energy marine platform	Late Cretaceous (early – middle Cenomanian)	95 – 100.5	Khalloufi et al., 2010
<i>Santanaclupea silvasantosi</i> Maisey, 1993	Santana Formation, Brazil	Estuarine	Early Cretaceous (Albian)	100.5 – 106.5	Maisey, 1993
<i>Scutatoclupea applegatei</i> (Alvarado-Ortega and Ovalles-Damián, 2008)	Chiapas, Mexico	Estuarine	Early Cretaceous (Aptian/Albian)	100.5 – 125	Alvarado-Ortega and Ovalles-Damián, 2008
<i>Scutatuspinosus itapagipensis</i> Silva Santos and Silva Corrêa, 1985	Bahia, Brazil	Lacustrine	Early Cretaceous (Neocomian)	130 – 145	NHMUK P.10072, 10073. 2 specimens

<i>Sorbinichthys africanus</i> Murray and Wilson, 2011	Akrabou Formation, Morocco	Marine	Late Cretaceous (Cenomanian/Turonian)	90 – 95	UALVP 51640 (holotype), 47186, 51641 3 specimens
<i>Sorbinichthys elusivo</i> Bannikov and Bacchia, 2000	Namoura, Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	Bannikov and Bacchia, 2000
<i>Thorectichthys marocensis</i> Murray and Wilson, 2013	Akrabou Formation, Morocco	Marine	Late Cretaceous (Cenomanian/Turonian)	90 – 95	UALVP 47178 (holotype), 47134, 51647, 51649 4 specimens
<i>Thorectichthys rhadinus</i> Murray and Wilson, 2013	Akrabou Formation, Morocco	Marine	Late Cretaceous (Cenomanian/Turonian)	90 – 95	UALVP 51653 (holotype), 51664, 51715 3 specimens
<i>Triplomystus noorae</i> Forey et al., 2003	Namoura, Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	NHMUK P.62517, 62519, 63325 3 specimens
<i>Triplomystus oligoscutatus</i> Forey et al., 2003	Namoura, Lebanon	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	NHMUK P.63154, 63155. 2 specimens
<i>Trollichthys bolcensis</i> Marramà and Carnevale, 2015	Monte Bolca, Italy	Marine	late Ypresian, Early Eocene	49 – 51	Marramà and Carnevale, 2015
<i>Tycheoichthys dunveganensis</i> Hay et al., 2007	Alberta, Canada	Marine	Late Cretaceous (Cenomanian)	93.9 – 100.5	CMN 52730 (holotype) 1 specimen

APPENDIX 6-3. List of morphological characters used in phylogenetic analysis

I NEUROCRANIUM

1. Lateral processes of mesethmoid for articulation of palatine: [0] absent; [1] present (Sato, 1994 [character 1]).
2. Lateral process of mesethmoid for attachment of maxillary ligament: [0] absent; [1] present (Sato, 1994 [character 2]).
3. Shape of lateral process of mesethmoid for attachment of maxillary ligament: [0] spinelike; [1] knoblike (Sato, 1994 [character 2]).

I modified original character coding into contingent scheme making new character 3 (shape of the process) contingent on the character 2 (presence or absence of the process).

4. Position of dorsal process of mesethmoid: [0] dorsal process far back from the tip of vomer; [1] dorsal process same level as the tip of vomer; [2] dorsal process projected anteriorly beyond the tip of vomer (Sato, 1994 [character 3]).
5. Premaxilla articulates with dorsal process of mesethmoid: [0] absent; [1] present (Sato, 1994 [character 3]).

Original character 3 in Sato (1994) is a compound character assessing the condition of two independent features (position of the dorsal process and its articulation with the premaxillae); I split into two characters – new character 4: Position of the dorsal process of the mesethmoid and new character 5: Articulation between premaxilla and the dorsal process of the mesethmoid.
6. Contact between lateral ethmoid wings and anterior portion of mesethmoid: [0] absent; [1] present (Sato, 1994 [character 4]).

Original character 4 (Sato, 1994) was converted to a binary character to remove subjective character state (original character state [1]). Original description was modified for clarity.

7. Number of lateral processes of lateral ethmoid: [0] One lateral process; [1] Two lateral processes (Sato, 1994 [character 5]).

8. Olfactory nerve foramen of lateral ethmoid: [0] absent; [1] foramen formed only by lateral ethmoid; [2] foramen surrounded by lateral ethmoid and ethmoid cartilage (Sato, 1994 [character 6]).

I modified original character to remove character state attribute – foramen size. Given an already small size of the feature it is nearly impossible to quantify size difference between the “small” and “large” character states. In the modified character, character states [3] and [4] from the original character statement are combined with the character states [1] and [2] of the modified character respectively.

9. Anterior projection of the articular base of lateral ethmoid: [0] absent; [1] present (Sato, 1994 [character 7]).

10. Shape of vomer: [0] vomer short (length is less than twice the width); [1] vomer long and slender (length is at least two times greater than the width) (Sato, 1994 [character 8]).

I added descriptions to quantify the original character states to reduce subjectivity in scoring this character.

11. Ossification of anterodorsal portion of vomer in adult specimens: [0] anterodorsal portion not ossified; [1] anterodorsal portion ossified (Sato, 1994 [character 9]).

12. Vomerine teeth: [0] absent; [1] present (Sato, 1994 [character 10]).

13. Parasphenoid keel under internal carotid artery foramina (under anterior margin of the prootic): [0] parasphenoid keel not expanded; [1] parasphenoid keel well expanded (Sato, 1994 [character 11]).
14. Posterior ramus of parasphenoid extending posteriorly from the anterior margin of basioccipital: [0] absent; [1] present (Sato, 1994 [character 12]).
15. Posterior opening of myodome: [0] absent; [1] present (Sato, 1994 [character 13]).
16. Anterior arm of orbitosphenoid: [0] anterior arm short and not reaching to lateral ethmoid; [1] anterior arm long reaching to lateral ethmoid; [2] anterior arm absent (Sato, 1994 [character 14]).

I added character state [2] to accommodate condition in *Denticeps* and an outgroup taxon *Chanos chanos*.

17. Optic fenestra in interorbital septum: [0] Optic fenestra formed by orbitosphenoid, pterosphenoid and basisphenoid; [1] Optic fenestra only formed by orbitosphenoid; [2] Optic fenestra surrounded by pterosphenoid and basisphenoid; [3] Optic fenestra surrounded by pterosphenoid and prootic (Sato, 1994 [character 15]).

I added character state [2] and [3] to accommodate conditions observed in *Denticeps* and an outgroup taxon *Cyprinus carpio* respectively.

18. Oculomotor foramen: [0] Discrete foramen for oculomotor nerve absent; [1] Oculomotor foramen formed by basisphenoid and prootic; [2] Oculomotor foramen formed by basisphenoid, orbitosphenoid, pterosphenoid and prootic (Sato, 1994 [character 16]).

I modified original character to remove subjective attribute of the foramen size.

19. Basisphenoid foramen: [0] Basisphenoid foramen absent; [1] Median vascular foramen in basisphenoid present; [2] Lateral vascular foramen in basisphenoid present (Sato, 1994 [character 17]).
20. Basisphenoid vertical process: [0] absent; [1] present at anterior extent (Sato, 1994 [character 17]).
- I split original character 17, a compound character, into two characters to describe conditions of the two independent features – basisphenoid foramen and vertical process (new characters 19 and 20).
21. Hypophysial fenestra: [0] hypophysial fenestra only surrounded by basisphenoid; [1] hypophysial fenestra surrounded by basisphenoid and prootics; [2] hypophysial fenestra surrounded by orbitosphenoid and prootics (Sato, 1994 [character 18]).
22. Recessus lateralis: [0] absent; [1] present.
23. Number of recessus lateralis foramina: [0] four; [1] three; [2] two; [3] one (Sato, 1994 [character 19]).
24. Arrangement of sensory canal openings in recessus lateralis: [0] all four discrete; [1] discrete infraorbital and preopercular openings, merged cephalic and accessory temporal canal openings; [2] discrete infraorbital and cephalic canal openings, merged preopercular and accessory temporal openings; [3] discrete preopercular and cephalic canal openings, merged infraorbital and accessory temporal openings; [4] discrete infraorbital canal opening, merged openings for preopercular, cephalic, and accessory temporal canals; [5] discrete preopercular canal opening, merged infraorbital, accessory temporal and cephalic canal openings; [6] all merged (Sato, 1994 [character 19]).
25. Pterotic bulla: [0] absent; [1] present (Sato, 1994 [character 20]).

26. Posterior process of pterotic: [0] located posterior to hyomandibular facet; [1] located medially to hyomandibular facet (Sato, 1994 [character 21]).

27. Posterior process of prootic: [0] absent; [1] present, but does not contact process of pterotic; [2] present, contacts process of the pterotic (Sato, 1994 [character 21]).

I split original character 21 into two characters (new characters 26 and 27) to specify states for the posterior processes of the pterotic and prootic bones independently.

28. Pre-epiotic fossa: [0] Pre-epiotic fossa absent; [1] Pre-epiotic fossa present and not penetrating posteriorly into sub-epiotic fossa; [2] Pre-epiotic fossa present and penetrating posteriorly into sub-epiotic fossa (Sato, 1994 [character 22]).

29. Auditory fenestra: [0] Auditory fenestra is formed by the exoccipital and prootic; basioccipital is excluded; [1] Auditory fenestra formed by the prootic, exoccipital, and basioccipital [2] auditory fenestra only surrounded by exoccipital (Sato, 1994 [character 23]).

I added new character state [2] to accommodate the condition in outgroup taxa *Cyprinus carpio* and *Chanos chanos*.

30. Dorsal profile of frontals: [0] dorsal profile of frontals nearly straight in lateral view; [1] dorsal profile of frontals gently concave in lateral view (Sato, 1994 [character 26]).

31. Ridge between orbital wall and levator arcus palatini fossa on frontal and sphenotic: [0] ridge absent or poorly developed; [1] ridge well developed (Sato, 1994 [character 27]).

32. Anterior frontal fontanelle: [0] Anterior frontal fontanelle absent; [1] Anterior frontal fontanelle present (Sato, 1994 [character 32]).

33. Posterior frontal fontanelles: [0] posterior frontal fontanelles absent; [1] A single median posterior frontal fontanelle present; [2] two lateral posterior fontanelles present (anterior

- process of supraoccipital divides the posterior frontal fontanelle into two) (Sato, 1994 [character 33]).
34. Fronto-parietal striations: [0] absent; [1] present (Sato, 1994 [character 34]).
35. Supratemporal canal on parietals: [0] groove or tube for supratemporal canal absent; [1] groove for supratemporal canal present; [2] tube for supratemporal canal present (Sato, 1994 [character 36]).
36. Parietal-epioccipital ridge: [0] Parietal-epioccipital ridge absent; [1] Parietal-epioccipital ridge present (Sato, 1994 [character 37]).
37. Participation of pterotic in sub-epiotic fossa: [0] Pterotic not participating in sub-epiotic fossa; [1] Pterotic participating in sub-epiotic fossa (Sato, 1994 [character 38]).
38. Position of intercalar: [0] Intercalar covers over junction of prootic, pterotic and exoccipital; [1] Intercalar only covers over exoccipital (Sato, 1994 [character 39]).
39. Pre-neural arch above foramen magnum: [0] absent; [1] present (Sato, 1994 [character 41]).
40. Skull roof condition: [0] parietal bones separated by supraoccipital (lateroparietal condition); [1] parietal bones contacting each other in the midline (medioparietal condition) (Zaragueta-Bagils, 2004 [character 1]; Murray and Wilson, 2013 [characters 2]).

II NASAL, CIRCUMORBITALS AND SCLEROTICS

41. Supraorbital canal of nasals: [0] Nasal bears bony tube for supraorbital canal; [1] Nasal bears bony flange to form a narrow groove for supraorbital canal, the groove opens posteriorly; [2] Nasal is folded anteriorly to form a broad groove along its longitudinal

axis; [3] nasal greatly expanded without a groove or bony tube to hold sensory canal (Sato, 1994 [character 42]).

The original character definition was modified for clarity.

I added character state [3] to accommodate the condition in *Denticeps*.

42. Medial process of supraorbital: [0] absent; [1] present (Sato, 1994 [character 43]).

43. Sensory canal of antorbital: [0] no bony constituent for sensory canal; [1] bony flange for sensory canal present; [2] bony tube for sensory canal present; [3] Antorbital complex shaped to accommodate the olfactory organ (Sato, 1994 [character 44]).

The original character definition was modified for clarity.

In *Cyprinus*, the antorbital bone is fused with the lachrymal.

44. Contact between first and third infraorbital bones: [0] absent – first infraorbital does not reach third infraorbital posteriorly; [1] present – first infraorbital reaches third infraorbital posteriorly (Sato, 1994 [character 45]).

The original character definition was modified for clarity.

45. Infraorbital 3 expanded, covering large part of the anterior arm of the preoperculum: [0] absent; [1] present.

46. Sixth infraorbital: [0] absent; [1] present (Sato, 1994 [character 46]).

47. Sclerotics: [0] absent; [1] one small sclerotic present; [2] two small sclerotics present (covering < 50% of the periphery of the eyeball); [3] two large sclerotics present covering >50% of the periphery of the eyeball (Sato, 1994 [character 47]).

I added descriptions to quantify the original character states and facilitate scoring of this character.

III SUSPENSORIUM AND OPERCULAR BONES

- 48. Palatine teeth: [0] absent; [1] present (Sato, 1994 [character 48]).
- 49. Ectopterygoid teeth: [0] absent; [1] present (Sato, 1994 [character 50]).
- 50. Endopterygoid teeth: [0] absent; [1] present (Sato, 1994 [character 52]).
- 51. Metapterygoid teeth: [0] absent; [1] present (Sato, 1994 [character 53]).
- 52. Fusion between metapterygoid and ectopterygoid: [0] absent; [1] present (Sato, 1994 [character 54]).

The original character definition was modified for clarity.

- 53. Metapterygoid foramen: [0] absent; [1] present (Sato, 1994 [character 55]).
- 54. Metapterygoid-quadrata fenestra: [0] metapterygoid-quadrata fenestra absent; [1] metapterygoid-quadrata fenestra present (Sato, 1994 [character 56]).
- 55. Anterior margin of metapterygoid located anterior to quadrata: [0] absent; [1] present (Di Dario, 2009 [character 1])
- 56. Interhyal articulation: [0] Interhyal articulates with both hyomandibula and symplectic; [1] Interhyal articulates with symplectic, the articular facet is located near the posterodorsal end of symplectic; [2] Interhyal articulates with symplectic, the articular facet is located near the posteroventral end of symplectic (Sato, 1994 [character 57]).
- 57. Inclination of ventral process of hyomandibula: [0] ventral process inclined anteriorly; [1] ventral process almost vertical; [2] ventral process inclined posteriorly (Sato, 1994 [character 58]).
- 58. Cranial articulation of hyomandibular: [0] hyomandibula articulates with sphenotic and pterotic; [1] hyomandibula articulates with sphenotic only (Sato, 1994 [character 59]).

59. Articular head of hyomandibula: [0] Single head articulating with both sphenotic and pterotic; [1] Hyomandibular head divided in two parts, anterior and posterior, of almost equal size and articulates with both sphenotic and pterotic; [2] Hyomandibula head divided into two parts, anteriorly and posteriorly, the posterior much larger than the anterior, and articulates on both sphenotic and pterotic; [3] Single, undivided hyomandibular head, articulating on sphenotic only (Sato, 1994 [character 59]).

The original character 59 (Sato, 1994) referred to two independent features – bones which head of the hyomandibula articulates with (new character 62) and shape of the articulation head of the hyomandibula (single head, two heads of equal size, and two heads of unequal size).

60. Ventral limb of hyomandibula and quadrate separated by metapterygoid: [0] present, metapterygoid separates hyomandibula and quadrate; [1] absent, hyomandibular meets quadrate (Di Dario, 2009 [character 2]).

61. Bony tube for sensory canal on preopercle: [0] bony tube unbranched at angle of preopercle; [1] bony tube branched at angle of preopercle (Sato, 1994 [character 60]).

62. Radial striations on opercle: [0] absent; [1] present (Sato, 1994 [character 61]).

IV JAWS

Upper jaw

63. Relative length of premaxilla: [0] premaxilla short (less than 50% of the maxillary length); [1] premaxilla long ($\geq 50\%$ of the maxillary length) (Sato, 1994 [character 62]).

I added descriptions to quantify the original character states and removed intermediate character state (original character state [1]) to reduce subjectivity in scoring this character.

64. Premaxillary teeth: [0] absent; [1] present, small (tooth height is ≤ 5 times tooth base diameter); [2] present, large, fang-like (tooth height is > 5 times tooth base diameter; fang-like teeth usually visible even when mouth is closed) (Sato, 1994 [character 63]).
65. Articular head of maxilla: [0] small, width is equal to or less than the width of the central shaft; [1] large, wider than the central shaft (looks like an expansion at the articular end of the maxilla) (Sato, 1994 [character 64]).
66. Central shaft of maxilla: [0] long, greater than the length of the articular head; [1] very short (shorter than the length of the articular head) or essentially absent (Sato, 1994 [character 64]).
67. Distal blade of maxilla: [0] strongly expanded, width greater than that of the central shaft; [1] slender, width is roughly the same as that of the central shaft (Sato, 1994 [character 64]).
68. Shape of distal blade of maxilla: [0] almost straight; [1] sickle-shaped with an upturned distal end; [2] bent, distal end pointing ventrally (Sato, 1994 [character 64]).
- The original character 64 (Sato, 1994) is split into four characters (new characters 64—67) to separate distinct parts of the maxilla.
69. Maxillary teeth: [0] absent; [1] present (Sato, 1994 [character 65]).
70. Anterior supramaxilla: [0] absent; [1] present (Sato, 1994 [character 66]).
71. Position of posterior supramaxilla: [0] on posterior portion of distal blade of maxilla; [1] on middle or anterior portions of distal blade of maxilla; [2] posterior supramaxilla absent (Sato, 1994 [character 67]).

I added character state [2] to account for condition observed in *Denticeps* as well as in some outgroup taxa (*Chanos chanos* and *Cyprinus carpio*)

72. Hypomaxilla (a paired bone positioned between distal end of premaxilla and middle of maxilla): [0] absent; [1] present (Sato, 1994 [character 68]).

73. Median notch in the ventral margin of the upper jaw: [0] absent; [1] present.

Lower jaw

74. Shape of dentary: [0] triangular with high coronoid process (depth of dentary greatly increases posteriorly); [1] triangular with low coronoid process (depth of dentary is roughly uniform throughout its length) (Sato, 1994 [character 69]).

75. Anterior end of dentary: [0] dentary is straight at symphysis; [1] dentary is curved ventrally at symphysis (Sato, 1994 [character 69]).

I split original character 69 from Sato (1994) into two (new characters 73 and 74) to describe conditions of the coronoid process and symphysis separately.

76. Dentary teeth: [0] absent; [1] present, minute or small (tooth height is ≤ 5 times tooth base diameter); [2] present, large, fanglike (tooth height is > 5 times tooth base diameter; fang-like teeth usually visible even when mouth is closed), extending well back along the border of coronoid process (Sato, 1994 [character 70]).

77. Separation between coronoid process of dentary and angulo-articular forming gap between ventral margin of coronoid process and dorsal margin of angulo-articular: [0] absent; [1] present (Sato, 1994 [character 71]).

The original character definition was modified for clarity.

78. Lateral wall of the facet of angulo-articular for quadrate articulation: [0] absent; [1] present (Sato, 1994 [character 72]).

The original character definition was modified for clarity.

79. Fusion of retroarticular to angulo-articular: [0] bones not fused; [1] bones fused together (Sato, 1994 [character 73]).

V HYAL APPARATUS

80. Length of the anterior ceratohyal: [0] length is ≤ 3 times its depth; [1] length is > 3 times the depth (Sato, 1994 [character 74]).

I modified original character 74 from Sato (1994) to describe length of the anterior ceratohyal only instead of length of the entire hyal arch which is a composite of the hypohyals, plus anterior and posterior ceratohyals.

81. Tooth patch on dorsal hypohyal: [0] absent; [1] present (Sato, 1994 [character 75]).

82. Anterior extensions of anterior ceratohyal: [0] absent; [1] present (Sato, 1994 [character 76]).

83. Foramen of hyoid arterial duct in anterior ceratohyal: [0] absent; [1] present (Sato, 1994 [character 77]).

84. Dorsal process of posterior ceratohyal: [0] The dorsal process absent or small; [1] The dorsal process large, distinctly upturned (Sato, 1994 [character 78]).

85. Hyoid arterial duct in posterior ceratohyal: [0] The hyoid arterial duct completely grooved; [1] The hyoid arterial duct contained at least partially in a tube formed by bony wall (Sato, 1994 [character 79]).

86. Foramen of hyoid arterial duct in posterior ceratohyal: [0] absent; [1] present (Sato, 1994 [character 80]).

87. Number of branchiostegal rays: [0] 4 – 10; [1] 13 – 20 (Sato, 1994 [character 81]).

88. Anterior processes of urohyal: [0] A single pair of anterior processes present; [1] Two pairs of the anterior processes present (Sato, 1994 [character 82]).
89. Dorsal blade or urohyal: [0] well-developed, maximum height of blade is > 25% of urohyal length; [1] weakly developed, maximum height of blade is < 25% of urohyal length; [2] absent (Sato, 1994 [character 83]).
90. Ventral blade of urohyal: [0] well-developed, maximum height of blade is > 25% of urohyal length; [1] weakly developed, maximum height of blade is < 25% of urohyal length; [2] absent (Sato, 1994 [character 83]).
91. Lateral wings of urohyal: [0] well-developed, forming distinct triangular profile of urohyal in dorsal and ventral views; [1] weakly developed, lateral expansions are present but do not form a distinct triangle in dorsal and ventral views of urohyal, instead lateral expansions are uniformly narrow; [2] absent (Sato, 1994 [character 83]).
92. Filamentous lateral wings of urohyal: [0] absent; [1] present (Sato, 1994 [character 83]).
- I split original character 83 from Sato (1994) into four (new characters 88 - 91) to describe conditions of the different element of the urohyal independently.

VI BRANCHIAL ARCHES

93. Mediopharyngobranchial cartilage: [0] absent; [1] present (Sato, 1994 [character 84]).
94. Shape of second infrapharyngobranchial: [0] triangular; [1] rectangular (Sato, 1994 [character 85]).
95. Length of anterior processes of second infrapharyngobranchial: [0] long, contributing over 30% of total length of infrapharyngobranchial; [1] short, less than 30% of total length of second infrapharyngobranchial (Sato, 1994 [character 85]).

I modified original character 85 (Sato, 1994) to describe shape of the second infrapharyngobranchial independently from the length of its anterior process (new characters 93, 94). I also excluded infrapharyngobranchial three from the character description because it does not vary in the length of the anterior process and it is isometric (and therefore appears to be correlated) with the second infrapharyngobranchial.

96. Foramen of fourth epibranchial: [0] absent; [1] present (Sato, 1994 [character 95]).

97. Anterior facet of fourth epibranchial: [0] The anterior facet of the fourth epibranchial not projected; [1] The anterior facet of the fourth epibranchial projected anteriorly (Sato, 1994 [character 96]).

98. Size of basihyal: [0] absent; [1] small, smaller than first basibranchial; [2] large, equal or greater in size than first basibranchial (Sato, 1994 [character 100]).

I added quantitative descriptors to the character states to make scoring of this character more objective.

99. Basihyal teeth: [0] absent; [1] present (Sato, 1994 [character 101]).

100. Number of basihyal cartilages: [0] One; [1] Two; [2] Three (Sato, 1994 [character 102]).

101. Anterior process of second basibranchial: [0] The anterior process not reaching to 1st basibranchial; [1] The anterior process reaching to 1st basibranchial (Sato, 1994 [character 105]).

102. Posterior process of second basibranchial: [0] posterior process not reaching to third basibranchial; [1] posterior process reaching to third basibranchial (Sato, 1994 [character 107]).

103. Shape of fifth ceratobranchial: [0] narrow, length is >3 times width; [1] wide, length is < 2 times width, ceratobranchial is spoon-shaped with proximal expansion (Sato, 1994 [character 120]).

VII PECTORAL GIRDLE

104. Dorsal process of posttemporal: slender and sharp [0]; sub-rectangular [1]; broad, wider at distal tip than at midpoint of bone [2] (Zaragueta-Bagils, 2004 [character 22]; Murray and Wilson, 2013 [character 23])

105. Ventral end of supracleithrum projecting posteriorly beyond cleithrum: [0] absent, supracleithrum does not project posteriorly beyond cleithrum; [1] present, ventral end of supracleithrum projects posteriorly well beyond cleithrum leaving a gap between supracleithrum and cleithrum (Sato, 1994 [character 123]).

106. Number of postcleithra: [0] Three; [1] Two; [2] One [3] None (Sato, 1994 [character 124]).

I modified this character to combine two character states (original character states [2] and [3]) which described the same number of postcleithra (one), but differentiated between small (original character state [2]) and very small, vestigial (original character state [3]) size of the postcleithrum. Combining these two character states into a single state (new character state [2]) removes subjectivity of discriminating between two similar size categories.

107. Medial process of cleithrum: [0] absent; [1] present (Sato, 1994 [character 125]).

108. Size of median process of cleithrum: [0] small, not reaching dorsal arm of cleithrum; [1] large, reaching anterior margin of dorsal arm of cleithrum (Sato, 1994 [character 125]).

I modified coding scheme of the original character 125 (Sato, 1994) to contingent coding.
109. Lateral flange of cleithrum: [0] absent; [1] present (Sato, 1994 [character 126]).

110. Size of coracoid: [0] moderate size, height of coracoid is less than height of dorsal limb of cleithrum; [1] very large, height of coracoid equal or greater than height of dorsal limb of cleithrum (Sato, 1994 [character 127]).

I added quantifying descriptions to the original character states to reduce subjectivity in scoring this character.

111. Mesocoracoid: [0] absent; [1] present (Sato, 1994 [character 128]).

112. Number of proximal radials of pectoral fin: [0] four; [1] three; [2] two (Sato, 1994 [character 129]).

I added character state [2] to accommodate condition in *Denticeps*.

113. Number of distal radials of pectoral fin: [0] Six; [1] Five; [2] Four; [3] three; [4] absent (Sato, 1994 [character 130]).

I added character state [3] to accommodate condition in *Denticeps*.

114. Accessory radials of pectoral fin: [0] present; [1] absent (Sato, 1994 [character 131]).

VIII PELVIC GIRDLE

115. Basipterygium: [0] absent; [1] present (Sato, 1994 [character 132]).

116. Notch in the anterior end of pelvic plate: [0] absent; [1] present (Sato, 1994 [character 132]).

117. Posteriorly directed “ischiac” processes: [0] absent; [1] present (Sato, 1994 [character 132]).

I split original character 132 in Sato (1994) into three characters (new characters 114 – 116) to describe independent conditions of the pelvic girdle.

118. Number and disposition of pelvic radials: [0] Four radials; [1] Three radials present; lacking first radial; [2] Three radial present, lacking fourth radial; [3] Two radials present; [4] Radials absent (Sato, 1994 [character 133]).

I added new character state [3] to accommodate condition present in *Denticeps clupeioides* (two radials).

IX DORSAL FIN AND ANAL FIN SKELETONS

119. Number of supraneurals: [0] One; [1] Four to twenty-five; [2] Twenty-eight or more (Sato, 1994 [character 134]).

120. Gap between last supraneural and first pterygiophore of dorsal fin: [0] absent; [1] present (Sato, 1994 [character 135]).

121. Inclination of supraneurals: [0] inclined posteriorly; [1] inclined anteriorly or vertical (Sato, 1994 [character 136]).

122. Supraneurals forming a fan-shaped structure with at least one anteriormost predorsal bone inclined anterior-posteriorly or oriented dorsoventrally to meet proximal end of the next supraneural: [0] absent; [1] present (Vernygora, 2015 [character 59]).

123. Middle pterygiophores of dorsal fin: [0] present; [1] absent (Sato, 1994 [character 137]).

124. Stay of dorsal fin support: [0] absent; [1] present (Sato, 1994 [character 138]).

I changed the original character 138 (Sato, 1994) into a binary character to remove vague descriptors of the length of the element.

125. Last dorsal fin ray: [0] short, shorter than preceding rays; [1] very long, exceeding length of the longest anterior fin rays (Sato, 1994 [character 139]).

126. Anal fin stay: [0] absent; [1] present, posterior process of anal stay not bifurcated; [2] present, posterior process of anal stay bifurcated (Sato, 1994 [character 140]).

X CAUDAL FIN SKELETON

127. Relation between parhypural and first preural centrum: [0] parhypural autogenous; not fused with first preural centrum; [1] parhypural fused with first preural centrum (Sato, 1994 [character 141]).

128. Direction of hypurapophysis: [0] free end oriented posteroventrally; [1] free end oriented anteroventrally (Sato, 1994 [character 142]).

129. Relation between first preural centrum and first ural centrum: [0] First preural centrum not fused with first ural centrum; [1] First preural centrum fused with first ural centrum (Sato, 1994 [character 143]).

130. Length of the neural spine on first preural centrum: [0] short, shorter than the neural spine of the second preural centrum; [1] long; equal to or longer than the neural spine of the second preural centrum (Sato, 1994 [character 144]).

131. Notch between neural spine on first preural centrum and pleurostyler rod: [0] present; [1] absent (Sato, 1994 [character 145]).

132. Number of ural centra: [0] Two; first ural centrum not fused with second ural centrum; [1] One; first ural centrum fused with second ural centrum (Sato, 1994 [character 146]).

133.Length of second hypural: [0] short, not reaching the level of the hypural 1 distally; [1] long, reaching posteriorly to the level of the distal ends of other hypurals (Sato, 1994 [character 147]).

I added descriptions to the original character states to reduce subjectivity in scoring this character.

134.Notch of third hypural: [0] absent; [1] present (Sato, 1994 [character 148]).

135.Relative size of third hypural: [0] smaller than the 4th hypural; [1] equal in size to the 4th hypural; [2] larger than other hypurals (Sato, 1994 [character 149]).

136.Elongation of upper hypurals: [0] not elongated, posterior margins of hypurals form an arch; [1] elongated, distal ends of the upper hypurals project beyond posterior margins of the lower hypurals (Sato, 1994 [character 150]).

137.Number of epurals: [0] One; [1] Two; [2] Three (Sato, 1994 [character 152]).

138.Pegs on basal segments of two middle principal caudal-fin rays: [0] Pegs on both middle rays absent; [1] Peg on upper middle ray present, but absent on the lower middle ray; [2] Pegs on both middle rays present (Sato, 1994 [character 153]).

139.Shape of basal segments of two middle caudal-fin rays: [0] The basal segments without spatula blade; [1] The basal segments with spatula blade (Sato, 1994 [character 154]).

I modified description of the original character 154 (Sato, 1994) to clarify that it is focused on the shape rather than the length of the basal segments of the middle caudal-fin rays.

140.Length of hypural 1: [0] long, reaching ural centrum 1; [1] short, not reaching ural centrum 1 (Zaragueta-Bagils, 2004 [character 25]; Murray and Wilson, 2013 [character 28]).

141. Diastema between second and third hypural: [0] third hypural expanded posteriorly, leaving no gap or notch between second and third hypural; [1] small triangular notch between the second and third hypural; [2] deep triangular notch between second and third hypural; [3] third hypural has a concave ventral edge forming a large concavity between second and third hypurals (Zaragueta-Bagils, 2004 [character 28]; Murray and Wilson, 2013 [characters 31, 32]).

142. Number of uroneurals: three [0]; two [1]; one [2] (Zaragueta-Bagils, 2004 [character 29]; Murray and Wilson, 2013 [characters 34]).

XI VERTEBRAL COLUMN

143. Anterior expansion on first rib: [0] absent; [1] present (Sato, 1994 [character 159]).

144. Basal pockets of ribs for swimbladder: [0] absent; [1] present (Sato, 1994 [character 160]).

145. Halves of the neural arches of most abdominal vertebrae: [0] separate; [1] fused medially (Zaragueta-Bagils, 2004 [character 17]; Murray and Wilson, 2013 [character 18]).

146. Rib articulation: [0] all ribs articulate with parapophyses along the abdominal region; [1] anteriormost ribs articulate with deep pits on the lateral side of all abdominal centra and those located posteriorly articulate with well-developed parapophyses; [2] all ribs articulate with deep pits on the lateral side of all abdominal centra (Zaragueta-Bagils, 2004 [character 18]; Murray and Wilson, 2013 [character 19]).

147. Epineurals and epipleurals in the caudal region: [0] absent; [1] present (Zaragueta-Bagils, 2004 [character 19]; Murray and Wilson, 2013 [character 20]).

148. Epicentrals: [0] absent; [1] present (Zaragueta-Bagils, 2004 [character 20]; Murray and Wilson, 2013 [character 21]).
149. Fusion between epicentrals and anterior ribs: [0] absent; [1] present (Patterson and Johnson, 1995; DiDario, 2002).
150. Cartilage chevrons at tips of epicentrals: [0] absent; [1] present (Patterson and Johnson, 1995; DiDario, 2002).
151. Posteriorly directed parapophyses of the second vertebra: [0] absent; [1] present (DiDario, 2002).
152. Interzygapophysal articulation: [0] absent; [1] present (DiDario, 2002).
153. Ornamentation of precaudal vertebrae: [0] two lateral pits; [1] three lateral pits; [2] multiple pits.

XII SCALES AND SCUTES

154. Dorsal scutes: [0] absent; [1] present (Sato, 1994 [character 155]).
155. Dorsal scute series: [0] one dorsal scute present just behind the occiput; [1] one dorsal scute present just in front of the dorsal fin; [2] complete series of scutes from behind the occiput to the origin of dorsal fin; [3] incomplete series of dorsal scutes with unscuted gap behind occiput (Sato, 1994 [character 155]).
- I changed the original character coding of the character 155 (Sato, 1994) to a contingent scheme (new characters 146 – 147). I added a new character state to the character 147 ([3] incomplete series of scutes with unscuted gap behind the occiput) describing condition present in fossil *ellimmichtyiform* species of the genus *Armigatus*.

156. Prepelvic scutes (excluding pelvic scute): [0] absent; [1] present (Sato, 1994 [character 156]).
157. Postpelvic scutes: [0] absent; [1] present (Sato, 1994 [character 157]).
158. Shape of pelvic scute: [0] Pelvic scute bears long, nearly straight ascending processes on each side; [1] Pelvic scute W-shaped; [2] Pelvic scute U-shaped, divided in the midline (Sato, 1994 [character 158]).
159. Shape of scutes in anterior part of predorsal series: [0] subrectangular; [1] heart-shaped or ovoid; [2] rhomboid (Zaragueta-Bagils, 2004 [character 47]; Murray and Wilson, 2013 [characters 48]).
160. Shape of scutes in posterior part of predorsal series: [0] subrectangular; [1] heart-shaped or ovoid; [2] rhomboid (Zaragueta-Bagils, 2004 [character 47]; Murray and Wilson, 2013 [characters 49]).
161. Series of spines on the posterior margin of the lateral wings of the predorsal scutes: [0] absent; [1] present. (Zaragueta-Bagils, 2004 [character 48]; Murray and Wilson, 2013 [characters 50]).
162. Prominent median strong spine on posteriormost predorsal scutes: [0] absent; [1] present (Zaragueta-Bagils, 2004 [character 49]; Murray and Wilson, 2013 [characters 51]).
163. Size of scutes of predorsal series: [0] all scutes of same size; [1] irregular in size, size of scutes increasing posteriorly (Zaragueta-Bagils, 2004 [character 50]; Murray and Wilson, 2013 [characters 52]).
164. Surface of predorsal scutes: [0] smooth; [1] ornamented with radiating grooves (Zaragueta-Bagils, 2004 [character 51]; Murray and Wilson, 2013 [characters 53]).

165. Number of predorsal scutes: [0] six to fourteen; [1] sixteen to nineteen; [2] twenty to forty-one (Zaragueta-Bagils, 2004 [character 52]; Murray and Wilson, 2013 [characters 54]).
166. Postpelvic abdominal scutes bearing very prominent and strong ventral spine: [0] spine absent; [1] spine present (Zaragueta-Bagils, 2004 [character 54]; Murray and Wilson, 2013 [characters 57]).
167. Size of lateral wings of abdominal scutes: [0] small; [1] large, extended upward and covering the abdominal cavity laterally for at least one quarter of the distance from ventral body edge to vertebral column. (Zaragueta-Bagils, 2004 [character 55]; Murray and Wilson, 2013 [characters 58]).
168. Shape of lateral wing of abdominal series scutes: [0] spine-like, with large spaces between wings of scutes; [1] wide or spatula-like, with wings of adjacent scutes touching for most of their length (Zaragueta-Bagils, 2004 [character 55]; Murray and Wilson, 2013 [characters 59]).
169. Postdorsal scute series: [0] absent; [1] present (Murray and Wilson, 2013 [characters 60]).
170. Lateral projections of most posterior predorsal scutes inclined antero-dorsally: [0] absent; [1] present.
171. Pelvic axillary scale: [0] absent; [1] present.
172. Pectoral axillary scale: [0] absent; [1] present.
173. Denticles (odontodes) covering dermal bones of skull: [0] absent; [1] present (Clausen, 1959; Greenwood, 1960, 1968a; Grande, 1985).

174. Complete series of lateral line scales: [0] absent; [1] present (Greenwood, 1960, 1968; Grande, 1985)

XIII OTHER CHARACTERS

175. Postcoelomic diverticula of swimbladder: [0] absent; [1] Postcoelomic diverticula passing through space between haemal spines and pterygiophores of anal fin; [2] Postcoelomic diverticula passing through on either or both sides of anal pterygiophores (Sato, 1994 [character 160]).

APPENDIX 6-4. GenBank accession numbers for molecular sequence data used in this study.

Protein coding sequences for 13 mitochondrial markers (ATP6, ATP8, COI, COII, COIII, CYTB, ND1, ND2, ND3, ND4, ND4L, ND5, ND6) were obtained from complete mitochondrial genomes (mtDNA) for which accession numbers are provided. Asterisk (*) indicates that sequence for congeneric species was used.

Species	mtDNA	RAG1	RAG2
Denticipitioidei			
Denticipitidae			
1. <i>Denticeps clupeioides</i> Clausen, 1959	NC_007889.1	DQ912100.1	XM028957876.1
Clupeioidi			
Engraulidae			
1. <i>Anchoa hepsetus</i> Linnaeus, 1785		DQ912112.1	DQ912145.1
2. <i>Anchoa mitchilli</i> (Valenciennes, 1848)		JQ012553.1	JQ012698.1
3. <i>Anchoviella lepidentostole</i> (Fowler, 1911)	NC_014269.1*	JQ012597.1	JQ012635.1
4. <i>Cetengraulis edentulus</i> (Cuvier, 1829)		JQ012578.1	JQ012693.1
5. <i>Coilia lindmani</i> Bleeker, 1857	NC_014271.1	DQ912123.1	DQ912157.1
6. <i>Engraulis japonicus</i> Temminck and Schlegel, 1846	NC_003097.1	AY430205.1	MG958260.1
7. <i>Lycengraulis grossidens</i> (Spix and Agassiz, 1829)	NC_014279.1	JQ012624.1	JQ012641.1
8. <i>Setipinna tenuifilis</i> (Valenciennes, 1848)	NC_020468.1*	MG958386.1	MG958245.1
9. <i>Stolephorus indicus</i> (van Hasselt, 1803)	NC_042729.1*	MG958396.1	JQ012671.1
Pristigasteridae			
1. <i>Ilisha megaloptera</i> (Swainson, 1839)	NC_009584.1*	KJ158151.1	KJ158099.1
2. <i>Ilisha melastoma</i> (Bloch and Schneider, 1801)	NC_009585.1*	MH325514.1	MH325485.1
3. <i>Opisthopterus tardoore</i> (Cuvier, 1829)		MG958390.1	MG958292.1
4. <i>Pellona ditchela</i> Valenciennes, 1847	NC_014268.1*	DQ912101.1	DQ912134.1
Chirocentridae			
1. <i>Chirocentrus dorab</i> (Forsskål, 1775)	NC_006913.1	DQ912127.1	DQ912163.1
Dussumieriidae			
1. <i>Dussumieria acuta</i> Valenciennes, 1847		MG958379.1	MG958267.1
2. <i>Dussumieria elopsoides</i> Bleeker, 1849	NC_035063.1		
3. <i>Etrumeus micropus</i> (Temminck and Schlegel, 1846)	AP009139.1	MG958394.1	MG958291.1
4. <i>Etrumeus sadina</i> (De Kay, 1842)	NC_009583.1	DQ912110.1	DQ912143.1
Clupeidae			
1. <i>Alosa aestivalis</i> (Mitchill, 1814)	NC_037017.1		DQ912146.1
2. <i>Alosa alosa</i> (Linnaeus, 1758)	NC_009575.1	MG958431.1	MG958301.1

3.	<i>Alosa fallax</i> (Lacepede, 1803)		MG958381.1	MG958302.1
4.	<i>Alosa mediocris</i> Mitchell, 1814	NC_037016.1	KJ158146.1	
5.	<i>Alosa pseudoharengus</i> (Wilson, 1811)	NC_009576.1	DQ912115.1	DQ912149.1
6.	<i>Alosa sappidissima</i> (Wilson, 1811)	NC_014690.1	DQ912116.1	DQ912150.1
7.	<i>Amblygaster sirm</i> (Walbaum, 1792)	NC_035064.1		
8.	<i>Anodontostoma chacunda</i> (Hamilton-Buchanan, 1822)	NC_021446.1	MG958380.1	MG958280.1
9.	<i>Brevoortia patronus</i> Goode, 1879		DQ912105.1	DQ912138.1
10.	<i>Brevoortia tyrannus</i> (Latrobe, 1802)	NC_014266.1	DQ912106.1	DQ912139.1
11.	<i>Clupea harengus</i> Linnaeus, 1758	NC_009577.1	DQ912114.1	XM031565155.1
12.	<i>Clupea pallasii</i> Valenciennes, 1847	NC_009578.1	DQ912118.1	DQ912152.1
13.	<i>Clupeichthys aesarnensis</i> Wongratana, 1983	NC_016719.1	MG958402.1	MG958305.1
14.	<i>Dorosoma cepedianum</i> (LeSueur, 1818)	NC_008107.1	DQ912099.1	DQ912132.1
15.	<i>Escualosa thoracata</i> (Valenciennes, 1847)	NC_016706.1		MG958268.1
16.	<i>Harengula jaguana</i> Poey, 1865	NC_016667.1	DQ912122.1	DQ912156.1
17.	<i>Herklotsichthys punctatus</i> (Rüppell, 1837)		MG958404.1*	MG958263.1*
18.	<i>Herklotsichthys quadrimaculatus</i> (Rüppell, 1837)		MG958423.1	MG958269.1
19.	<i>Hyperlophus vittatus</i> (Castelnau, 1875)	NC_016671.1		
20.	<i>Jenkinsia lamprotaenia</i> (Gosse, 1851)	NC_006917.1	DQ912107.1	DQ912140.1
21.	<i>Konosirus punctatus</i> (Temminck and Schlegel, 1846)	NC_016694.1		
22.	<i>Lile stolifera</i> Jordan and Gilbert, 1881		KJ158137.1	KJ158100.1
23.	<i>Nematalosa come</i> (Richardson, 1846)	NC_021447.1	MG958440.1	MG958281.1
24.	<i>Nematalosa erebi</i> (Günther, 1868)	NC_043853.1	MG958441.1	MG958283.1
25.	<i>Nematalosa japonica</i> Regan, 1917	NC_009586.1	MG958392.1	MG958282.1
26.	<i>Odaxothrissa losera</i> Boulenger, 1899	NC_009590.1	DQ912131.1	DQ912167.1
27.	<i>Opisthonema libertate</i> (Günther, 1867)			KJ158101.1
28.	<i>Opisthonema oglinum</i> (LeSueur, 1818)		DQ912111.1	DQ912144.1
29.	<i>Pellonula vorax</i> Gunther, 1868	AP009231.1	DQ912130.1	DQ912166.1
30.	<i>Potamalosa richmondia</i> (Macleay, 1879)	NC_016674.1	MG958417.1	MG958289.1
31.	<i>Sardina pilchardus</i> (Walbaum, 1792)	NC_009592.1	MG958429.1	DQ912158.1
32.	<i>Sardinella albella</i> (Valenciennes, 1847)	NC_016726.1	MG958415.1	KP325113.1
33.	<i>Sardinella gibbosa</i> (Bleeker, 1849)	NC_037131.1	MG958421.1	KP325115.1
34.	<i>Sardinella lemuru</i> Bleeker, 1853	NC_044472.1*	KJ158136.1	
35.	<i>Sardinella zunasi</i> (Bleeker, 1854)	NC_039553.1		
36.	<i>Sardinops sagax</i> (Jenyns, 1842)	NC_002616.1*		MG958295.1
37.	<i>Spratelloides delicatulus</i> (Bennett, 1831)	NC_009588.1	DQ912128.1	DQ912164.1
38.	<i>Spratelloides gracilis</i> (Temminck and Schlegel, 1846)	NC_009589.1	DQ912129.1	DQ912165.1
39.	<i>Sprattus muelleri</i> (Klunzinger, 1880)	NC_016669.1	MG958418.1	
40.	<i>Sundasalanx</i> sp. Roberts, 1981	NC_016663.1		
41.	<i>Tenuialosa thibaudeaui</i> (Durand, 1940)	NC_016719.1	MG958385.1	MG958279.1

Outgroups

Ostariophysii: Gonorynchiformes: Chanidae

4.	<i>Chanos chanos</i> (Forsskål, 1775)	NC_004693.1	AY430207.1	XM030766370.1
----	---------------------------------------	-------------	------------	---------------

Ostariophysii: Cypriniformes: Cyprinidae

5.	<i>Cyprinus carpio</i> Linnaeus, 1758	NC_001606.1	MG806244.1	AY787041.1
----	---------------------------------------	-------------	------------	------------

Euteleostei: Protacanthopterygii: Esociformes: Esocidae

6.	<i>Esox lucius</i> Linnaeus, 1758	NC_004593.1	AY380542.1	XM010883747.2
----	-----------------------------------	-------------	------------	---------------

APPENDIX 6-5. Configuration file used in the Partition Finder 2 analysis

```
## ALIGNMENT FILE ##
alignment = infile.phy;

## BRANCHLENGTHS: linked | unlinked ##
branchlengths = unlinked;

## MODELS OF EVOLUTION: all | allx | mrbayes | beast | gamma | gammai | <list> ##
models = mrbayes;

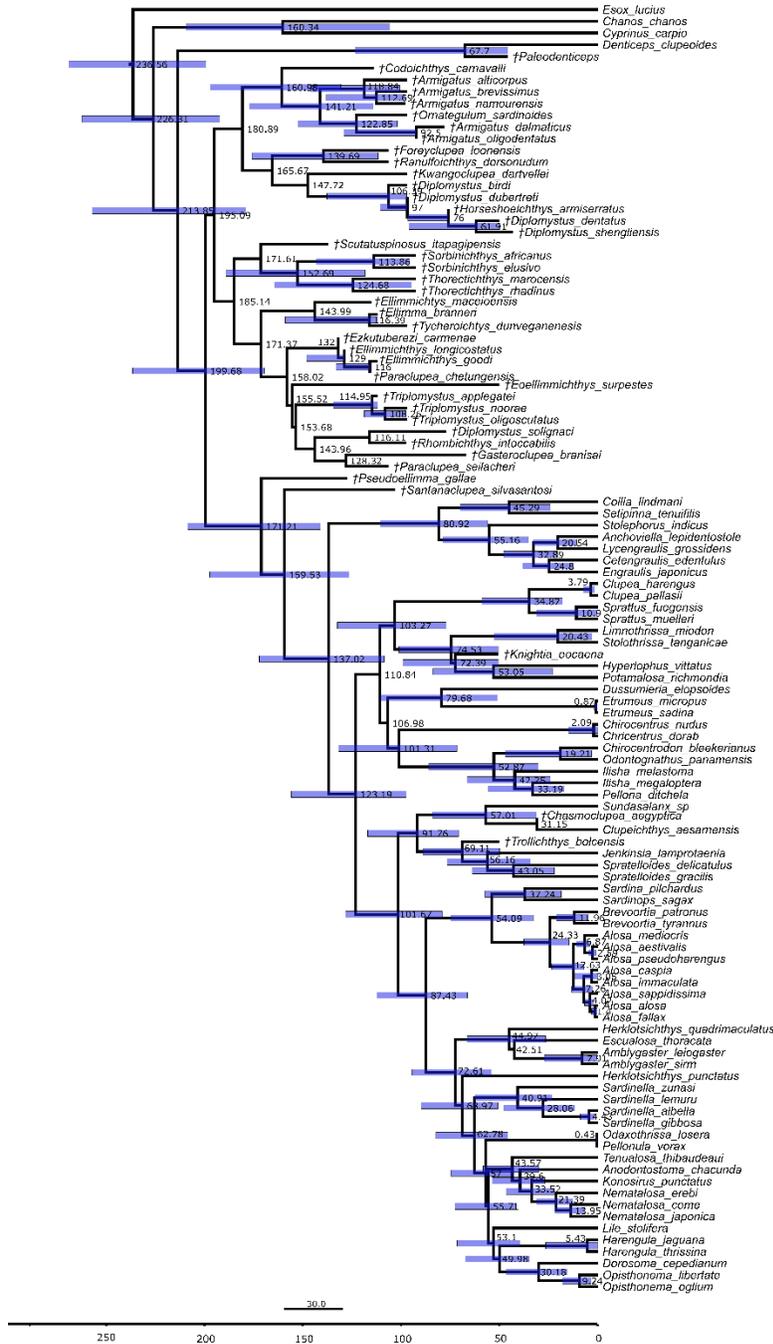
# MODEL SELECCTION: AIC | AICc | BIC #
model_selection = BIC;

## DATA BLOCKS ##
charset ATP6_pos1 = 1-681\3;
charset ATP6_pos2 = 2-681\3;
charset ATP6_pos3 = 3-681\3;
charset ATP8_pos1 = 682-849\3;
charset ATP8_pos2 = 683-849\3;
charset ATP8_pos3 = 684-849\3;
charset COI_pos1 = 850-2391\3;
charset COI_pos2 = 851-2391\3;
charset COI_pos3 = 852-2391\3;
charset COII_pos1 = 2392-3082\3;
charset COII_pos2 = 2393-3082\3;
charset COII_pos3 = 2394-3082\3;
charset COIII_pos1 = 3083-3865\3;
charset COIII_pos2 = 3084-3865\3;
charset COIII_pos3 = 3085-3865\3;
charset CYTB_pos1 = 3866-4996\3;
charset CYTB_pos2 = 3867-4996\3;
charset CYTB_pos3 = 3868-4996\3;
charset ND1_pos1 = 4997-5971\3;
charset ND1_pos2 = 4998-5971\3;
charset ND1_pos3 = 4999-5971\3;
charset ND2_pos1 = 5972-6994\3;
charset ND2_pos2 = 5973-6994\3;
charset ND2_pos3 = 5974-6994\3;
charset ND3_pos1 = 6995-7342\3;
charset ND3_pos2 = 6996-7342\3;
charset ND3_pos3 = 6997-7342\3;
charset ND4_pos1 = 7343-8722\3;
charset ND4_pos2 = 7344-8722\3;
charset ND4_pos3 = 7345-8722\3;
charset ND4L_pos1 = 8723-9020\3;
charset ND4L_pos2 = 8724-9020\3;
```

```
charset ND4L_pos3 = 8725-9020\3;  
charset ND5_pos1 = 9021-10769\3;  
charset ND5_pos2 = 9022-10769\3;  
charset ND5_pos3 = 9023-10769\3;  
charset ND6_pos1 = 10770-11285\3;  
charset ND6_pos2 = 10771-11285\3;  
charset ND6_pos3 = 10772-11285\3;  
charset RAG1_pos1 = 11286-12749\3;  
charset RAG1_pos2 = 11287-12749\3;  
charset RAG1_pos3 = 11288-12749\3;  
charset RAG2_pos1 = 12750-13967\3;  
charset RAG2_pos2 = 12751-13967\3;  
charset RAG2_pos3 = 12752-13967\3;
```

```
## SCHEMES, search: all | user | greedy | rcluster | rclusterf | kmeans ##  
search = greedy;
```

APPENDIX 6-6. Time-calibrated phylogram of Clupeomorpha from Bayesian inference analysis of combined data. Bars indicate 95% credibility intervals for age estimates. Median age values are shown at nodes. Branch lengths are proportional to time. Scale bar indicates number of substitutions per site.



CHAPTER 7

Conclusions and future research

7.1 General conclusions

The primary aim of my thesis was to provide an updated classification of Clupeomorpha (herrings and allies) by conducting a comprehensive phylogenetic analysis of the group. Despite the abundance and economic importance of clupeomorph fishes, their evolutionary interrelationships remain surprisingly poorly understood. The most extensive morphological studies of Clupeomorpha that included both living and fossil taxa were conducted over three decades ago (Grande, 1982, 1985) and did not fully resolve evolutionary relationships, nor were they performed using current, more objective, methods of evolutionary analyses. Later morphological revisions focused on extant clupeiform lineages only, and reconstructed phylogenetic relationships using either distance methods (Sato, 1994) or maximum parsimony (Di Dario, 2009). There is, however, a growing body of research suggesting that probabilistic methods of phylogenetic inference (maximum likelihood and Bayesian inference) provide more accurate estimations of evolutionary relationships compared to traditional parsimony (Wright and Hillis 2014, O'Reilly et al. 2016, Puttick et al. 2017, O'Reilly et al. 2018, Puttick et al. 2019), although, it is not clear which probabilistic method performs best when applied to morphological data and whether performance of these methods varies across different software implementations. In Chapter 2, I assessed the performance of four major probabilistic software packages— the Bayesian inference-based MrBayes and RevBayes, and the maximum likelihood-based IQ-TREE and RAxML— under variable taxonomic sampling and levels of missing data conditions. The results of the simulation study indicate that increased taxonomic sampling

improves accuracy, precision, and resolution of reconstructed topologies. Bayesian inference applications were the most consistent, accurate, and robust to variation in taxonomic sampling under all tested conditions according to the Robinson-Foulds tree distance metric. If recovery of incorrectly inferred clades is to be avoided in systematics, Bayesian inference should be the preferred method for the analysis of morphological data.

Probabilistic methods, however, are far from optimal in terms of evolutionary models for morphological data. The explicit time-reversible Mk model of morphological data evolution (Lewis, 2001) specifies an oversimplified evolutionary scenario with a single rate of character state substitutions and equal character state frequencies; these assumptions are almost never corroborated by empirical data sets. Optimization of model-based approaches for morphological data analysis is necessary for a more accurate estimate of evolutionary relationships. In Chapter 3, I developed and presented a new model of morphological trait evolution, FreqMorph (Frequency model for Morphological data). The model implements empirical character state frequencies for each individual character in a data set. The major findings indicate that using empirical character state frequencies can improve accuracy of phylogenetic reconstructions in some cases, especially when the number of taxa is large and branch lengths are of particular importance. However, FreqMorph in its current implementation is sensitive to sampling and is prone to bias when empirical character state frequencies deviate substantially from the true underlying frequencies which should be considered when selecting an appropriate model for a phylogenetic analysis of morphological data.

Delimiting taxonomic units that can be used in phylogenetic analysis is another pertinent issue in systematic studies. My major goal in this thesis was to perform a species level analysis of superorder Clupeomorpha, therefore delimiting species boundaries was an essential topic that

I addressed in Chapters 4 and 5 using neontological and palaeontological case studies. In the case study of extant members of genus *Alosa* (Chapter 4), I used an integrative approach to assess differentiation among endemic Ponto-Caspian shads in the Sea of Azov using both morphometric and genetic data (published as Vernygora et al., 2018). Morphological species assignments based on gill raker number were not congruent with genetic lineages determined with mitochondrial DNA and SNPs, instead genetic lineages were associated with sampling location and several other morphometric traits (caudal peduncle depth, preanal length, and head length). Overall, gill raker number was unreliable for delimiting species of *Alosa* and the integrative approach combining genetic and morphological data provided a better framework for species delimitation in this study system.

In palaeontological studies, however, species delimitation relies heavily, if not solely, on morphological data. In Chapter 5, I examined morphological variation among members of extinct genus †*Armigatus*. This group comprises five species of small marine fishes that were abundant in the Late Cretaceous Tethys Sea (Forey et al., 2003; Murray and Wilson, 2013; Vernygora and Murray, 2016; Murray et al., 2016); however, very little is known about one of the species, †*A. alticorpus*, which resulted in recent cladistic analyses excluding this species due to a lack of sufficient information for the taxon. Comparative examination of multiple specimens of each species enabled me to establish a set of morphological traits that allow more reliable species delimitation within genus †*Armigatus*; this includes number of abdominal scutes, vertebrae, and predorsal bones, as well as position of the dorsal fin.

In the last chapter of the thesis (Chapter 6), I bring together morphological and molecular data available for 107 clupeomorph species, including fossil and living taxa, to perform a combined evidence phylogenetic analysis. Results of multiple phylogenetic analyses I performed

indicated a substantial degree of agreement in phylogenetic signal between morphological and molecular data when the data sets were analyzed separately. Synthesis of multiple phylogenetic results suggests that Clupeomorpha are a monophyletic group that can be subdivided into three orders, Denticipitiformes, †Ellimmichthyiformes, and Clupeiformes, with Denticipitidae being the basalmost clupeomorph lineage. Interrelationships within †Ellimmichthyiformes and Clupeiformes are overall consistent with current classification. Because family Clupeidae in its current usage does not represent a monophyletic group, I assign family rank to previously recognized clupeid subfamilies that have been consistently recovered as monophyletic groups: Clupeidae (new usage), Dorosomatidae, and Alosidae.

Overall, this research is the first comprehensive study of Clupeomorpha that combines morphological and molecular data to investigate the evolutionary history of the group. Results of my research will serve to determine major trends in clupeomorph evolution and adaptations to their environment. Additionally, methods used and newly developed in this study are directly applicable across different groups of organisms where morphological data are used to infer phylogenetic relationships.

7.2 Future research

In this thesis, I cover three major topics that are of particular interest to me that I hope to continue developing in the future. The first and most broadly applicable line of future research includes methodological advances in the field of phylogenetic inference. I started on this topic in Chapter 3 by introducing a new model of morphological evolution that can be used in a probabilistic framework of phylogenetic inference. Full implementation of the new model, however, requires further work that involves creating a template for BEAST2 software to fully

incorporate the new model into the program's interface and make it easily available to users. Additionally, the model should be further improved to allow character state frequencies to be estimated during the Markov Chain Monte Carlo sampling procedure. This would, at least in part, compensate for the current susceptibility of the model to error when character state frequencies observed in a data set deviate significantly from the true character state frequencies. The next step in advancing this line of research would be to relax the morphological model of evolution beyond the current FreqMorph implementation and to allow asymmetrical character state transition probabilities reflecting empirical observations that some character state changes are less frequent than others. This approach roughly approximates the use of step matrices in Sankoff parsimony (1975), but the 'weight' of character state changes is defined as a probability value estimated as one of the model parameters during the Markov Chain sampling process.

Another avenue for future research is the species delimitation problem. I touched on this topic in Chapters 4 and 5 by looking at interspecific variation in closely related extant and fossil species. Extant clupeomorph fishes provide a great study system for applying the integrative taxonomy approach (Yeates et al., 2011). Unfortunately, very few species complexes within Clupeomorpha are sufficiently studied to provide the morphological and molecular data necessary for an integrative analysis. Genus *Alosa* is one of the few examples with well-studied members. Some European and North American species of *Alosa* have been the subject of thorough multi-disciplinary research involving morphological, molecular, and ecological analyses; however, other members of the genus that have a smaller distribution and mostly local commercial value (e.g., *A. maeotica*, *A. tanaica*, *A. immaculata*) are considerably less studied and limited genetic resources are available to investigate genetic structure of these endemic species complexes. In Chapter 4, I provided the first genome-wide SNP data for endemic Ponto-

Caspian shads of genus *Alosa*, and I am interested in continuing to collect morphological and molecular data to address not only their species boundaries but also potential drivers of diversification among species in the Ponto-Caspian basin, whether it is spatial, temporal, or some other form of ecological divergence.

Finally, large scale phylogenetic studies of Clupeomorpha need to continue in future to address remaining questions in clupeomorph systematics. Major unresolved issues include the interrelationships among major clupeomorph lineages, primarily Dussumieriidae, Pristigasteridae, and Engraulidae, reliable diagnosis and taxonomic composition of Clupeidae and Dorosomatidae, and the phylogenetic position of fossil taxa. These questions should be addressed by expanding taxonomic sampling of both extant and fossil clupeomorphs. The morphological data set should also be expanded with soft tissue characters such as characteristics of the digestive tract and swim bladder.

References

- Di Dario, F. 2009. Chirocentrids as engrauloids: evidence from suspensorium, branchial arches, and infraorbital bones (Clupeomorpha, Teleostei). *Zoological Journal of the Linnean Society* 156:363–383.
- Forey, P. L., L. Yi, C. Patterson, and C. E. Davies. 2003. Fossil fishes from the Cenomanian (Upper Cretaceous) of Namoura, Lebanon. *Journal of Systematic Palaeontology* 1:227–330.
- Grande, L. 1982. A revision of the fossil genus *Diplomystus*, with comments on the interrelationships of clupeomorph fishes. *American Museum Novitates* 2728:1–34.
- Grande, L. 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. *Bulletin of the American Museum of Natural History* 181:231–372.
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Murray, A. M., and M. V. H. Wilson. 2013. Two new paraclupeid fishes (Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco; pp. 267–290 in G. Arratia, H.-P. Schultze, and M. V. H. Wilson (eds.), *Mesozoic Fishes 5 – Global Diversity and Evolution*. Verlag Dr. Friederich Pfeil, Munich, Germany, 560 pp.
- Murray, A. M., O. Vernygora, S. Japundžić, J. Radovčić, M. V. H. Wilson, D. Bardack, and T. Grande. 2016. Relationships of the species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) and the description of a new species from the Cretaceous of Dalmatia, Croatia. *Journal of Vertebrate Paleontology* 36:e1226851.

- O'Reilly, J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* 61:105–118.
- O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biological Letters* 12:20160081.
- Puttick, M. N., J. E. O'Reilly, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, and P. C. J. Donoghue. 2017. Uncertainty: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B* 284:p.20162290.
- Puttick, M. N., J. E. O'Reilly, D. Pisani, and P. C. J. Donoghue. 2019. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology* 62:1–17.
- Sankoff, D. 1975. Minimal mutation trees of sequences. *SIAM Journal of Applied Mathematics* 28:35–42.
- Vernygora, O. and A. M. Murray. 2016. A new species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous of Morocco, and its phylogenetic relationships. *Journal of Vertebrate Paleontology* 36:e1031342.
- Vernygora, O.V., C. S. Davis, A. M. Murray, and F. A. H. Sperling. 2018. Delimitation of *Alosa* species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches. *Journal of Fish Biology* 93:1216–1228.

Wright, A. M. and D. M. Hillis. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS ONE 9:e109210.

Yeates D. K., A. Seago, L. Nelson, S. L. Cameron, L. Joseph, and J. W. H. Trueman. 2011. Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* 36:209–217.

LITERATURE CITED

- Abadi, S., D. Azouri, T. Pupko, and I. Mayrose. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications* 10: 10.1038/s41467-019-08822-w.
- Abdo, Z., V. N. Minin, P. Joyce, and J. Sullivan. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. *Molecular Biology and Evolution* 22:691–703.
- Adachi, J. and M. Hasegawa. 1995. Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. *Journal of Molecular Evolution* 40: 622–628.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on automatic control* 19:716–723.
- Alexandrino, P., R. Faria, D. Linhares, F. Castro, M. Le Corre, R. Sabatié, J. L. Bagliniere, and S. Weiss. 2006. Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. *Journal of Fish Biology* 69:242–259.
- Alvarado-Ortega, J. 2014. Ancient herring from the Tlayúa Quarry (Cretaceous, Albian) near Tepexi de Rodríguez, Puebla State, central Mexico, closing the gap in the early diversification of Clupeomorpha. *Cretaceous Research* 50:171–186.
- Alvarado-Ortega, J. and E. Ovalles-Damián. 2008. *Triplomystus applegatei*, sp.nov. (Teleostei: Ellimmichthyiformes), a rare “triple armored herring” from El Espinal Quarry (Early Cretaceous), Chiapas, southeastern Mexico. *Journal of Vertebrate Paleontology* 28:53–60.

- Alvarado-Ortega, J. and M. Melgarejo-Damián. 2017. *Parachupea seilacheri* sp. nov., a double armored herring (Clupeomorpha, Ellimmichthyiformes) from the Albian limestones of Tlayúa quarry, Puebla, Mexico. *Revista Mexicana de Ciencias Geológicas* 34:234–249.
- Alvarado-Ortega, J., E. Ovalles-Damián, and G. Arratia. 2008. A review of the interrelationships of the order Ellimmichthyiformes (Teleostei: Clupeomorpha); pp. 257–278 in G. Arratia, H.-P. Schultze, and M.V.H. Wilson (eds.), *Mesozoic Fishes 4 – Homology and Phylogeny*. Verlag Dr. Friederich Pfeil, Munich, Germany, 502 pp.
- Antipa, G. 1904. Die Clupeinen des westlichen Teiles des Schwarzen Meeres und der Donaumündungen. *Anzeiger der Kaiserlichen Akademie der Wissenschaften, Wien, Mathematisch-Naturwissenschaftliche Classe* 41:299–303.
- Arratia, G. 1997. Basal teleosts and teleostean phylogeny. *Palaeo Ichthyologica* 7:1–168.
- Bagley, J. C., F. Alda, M. F. Breitman, E. Bermingham, E. P. van den Berghe, and J. B. Johnson. 2015. Assessing species boundaries using multilocus species delimitation in a morphologically conserved group of neotropical freshwater fishes, the *Poecilia sphenops* species complex (Poeciliidae). *PloS ONE* 10:e0121139.
- Bakhshalizadeh, S. and A. Bani. 2017. Use of geometric morphometrics to identify ecophenotypic variation of juvenile Persian sturgeon *Acipenser persicus*. *Scientia Marina* 81:187–193.
- Bannikov, A. F. 2015. A new genus for the Cenomanian ellimmichthyiform fishes from Lebanon and Mexico. *Bollettino della Societa Paleontologica Italiana* 54:211–218.
- Bannikov, A. F. and F. Bacchia. 2000. A remarkable clupeomorph fish (Pisces, Teleostei) from a new Upper Cretaceous marine locality in Lebanon. *Senckenbergiana Lethaea* 80:3–11.

- Beacham, T. D. 1990. A genetic analysis of meristic and morphometric variation in chum salmon (*Onchorynchus keta*) at three different temperatures. *Canadian Journal of Zoology* 68:225–229.
- Bennett, E. T. 1832. Observations on a collection of fishes from the Mauritius, presented by Mr. Telfair, with characters of new genera and species. *Proceedings of the Committee of Science and Correspondence of the Zoological Society of London* 31:165–169.
- Bennett, E. T. 1835. A letter to the Secretary by Keith E. Abbott, Esq., a collection of skins of mammals and birds, and of preserved reptiles, fishes, and insects. *Proceedings of the Zoological Society of London* 3:89–92.
- Bentzen, P., W. C. Leggett, and G. G. Brown. 1993. Genetic relationships among the shads (*Alosa*) revealed by mitochondrial DNA analysis. *Journal of Fish Biology* 43:909–917.
- Berg, L. S. 1940. Sistema ryboobraznykh i ryb, nye zhivushchikh i iskopaemykh: Classification of fishes, both recent and fossil. *Izd-vo Akademii nauk SSSR*, 517 pp.
- Berg, L. S. 1948. Freshwater fishes of the USSR and adjacent countries. vol. 1. *Izdatelstvo Akademii Nauk SSSR, Moskva and Leningrad*, 504 pp.
- Besag, J. and P. J. Green. 1993. Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society B* 55:25–37.
- Beukers-Stewart, B. D. and G. P. Jones. 2004. The influence of prey abundance on the feeding ecology of two piscivorous species of coral reef fish. *Journal of Experimental Marine Biology and Ecology* 299:155–184.
- Bianchi, C. N., C. Morri, M. Chiantore, M. Montefalcone, V. Parravicini, and A. Rovere. 2012. Mediterranean Sea biodiversity between the legacy from the past and a future of change;

- pp1–55 N. Stambler (ed.), *Life in the Mediterranean Sea: a look at habitat changes*. New York, NY: Nova Science Publishers, 749 pp.
- Blainville, H. M. de. 1818. Sur les ichthyolites ou les poissons fossiles. Paris, *Nouvelles Dictées d'Histoire naturelle* 37:310–395.
- Blaxter, J. H. S. and J. R. Hunter. 1982. The biology of the clupeoid fishes. *Advances in Marine Biology* 20:1–223.
- Bleeker, P. 1849a. A contribution to the knowledge of the ichthyological fauna of Celebes. *Journal of the Indian Archipelago and Eastern Asia (Singapore)* 31:65–74.
- Bleeker, P. 1849b. Bijdrage tot de kennis der ichthyologische fauna van het eiland Madura, met beschrijving van eenige nieuwe soorten. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*. 22:1–16.
- Bleeker, P. 1851. Nieuwe bijdrage tot de kennis der ichthyologische fauna van Celebes. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 2:209–224.
- Bleeker, P. 1853a. Diagnostische beschrijvingen van nieuwe of weinig bekende vischsoorten van Batavia. Tiental I–VI. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 4:451–516.
- Bleeker, P. 1853b. Nalezingen op de ichthyologie van Japan. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen* 25:1–56.
- Bleeker, P. 1854. Faunae ichthyologicae japonicae species novae. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 6:395–426.
- Bleeker, P. 1855. Nalezingen op de vischfauna van Sumatra. *Visschen van Lahat en Sibogha*. *Natuurkundig Tijdschrift voor Nederlandsch Indië* 9:257–280.

- Bleeker, P. 1857. Zesde bijdrage tot de kennis der vischfauna van Sumatra. Visschen van Padang, Troessan, Priaman, Sibogha en Palembang. *Acta Societatis Regiae Scientiarum Indo-Neêrlandicae* 3:1–50.
- Bleeker, P. 1859. Enumeratio specierum piscium hucusque in Archipelago indico observatarum adjectis habitationibus citationibusque, ubi descriptiones earum recentiores reperiuntur, nec non speciebus Musei Bleekeriani Bengalensibus, Japonicis, Capensibus Tasmanicisque. *Acta Societatis Regiae Scientiarum Indo-Neerlandicae* 6:1–276.
- Bleeker, P. 1972. Atlas ichthyologique des Indes Orientales Neerlandaises 6:79–143.
- Bloch, M. E. and J. G. Schneider. 1801. *Systema Ichthyologiae Iconibus. Illustratum. Post obitum auctoris opus inchoatum absolvit, correxit, interpolavit Jo. Gottlob Schneider, Saxo. Berolini. Sumtibus Auctoris Impressum et Bibliopolio Sanderiano Commissum.* 110 pp.
- Bloom, D. D. and J. P. Egan. 2018. Systematics of Clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. *Neotropical Ichthyology* 16:p.e180095.
- Bloom, D. D. and N. R. Lovejoy. 2012 Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (Family Engraulidae). *Journal of Evolutionary Biology* 25:701–715.
- Bloom, D. D. and N. R. Lovejoy. 2014. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proceedings of the Royal Society B* 281:10.1098/rspb.2013.2081.
- Bloom, D. D., M. D. Burns, and T. A. Schriever. 2018. Evolution of body size and trophic position in migratory fishes: a phylogenetic comparative analysis of Clupeiformes

- (anchovies, herring, shad and allies). *Biological Journal of the Linnean Society* 125:302–314.
- Bogdanowicz, D. and K. Giaro 2012. Matching split distance for unrooted binary phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9:150–160.
- Bogdanowicz, D., K. Giaro, and B. Wróbel. 2012. TreeCmp: comparison of trees in polynomial time. *Evolutionary Bioinformatics* 8:EBO.S9657.
- Boisneau, P., C. Mennesson-Boisneau and R. Guyomard. 1992. Electrophoretic identity between allis shad *Alosa alosa* L. and twaite shad, *Alosa fallax* (Lacepede). *Journal of Fish Biology* 40:731–738.
- Borsa, P. 2002. Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of the Linnean Society* 75:261–269.
- Bouckaert, R. R., and A. J. Drummond. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology* 17: 10.1186/s12862-017-0890-6.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10:e1003537.
- Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, and M. Matschiner. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15:p.e1006650.

- Boulenger, G. A. 1899. Matériaux pour la faune du Congo. Zoologie.--Série I. Poissons nouveaux du Congo. Quatrième Partie. Polyptères, Clupes, Mormyres, Characins. Annales du Musée du Congo (Ser. Zoology) 1:59–96.
- Boulenger, G. A. 1906. Fourth contribution to the ichthyology of Lake Tanganyika.--Report on the collection of fishes made by Dr. W. A. Cunnington during the Third Tanganyika Expedition, 1904-1905. Transactions of the Zoological Society of London 17(1):537–601.
- Bowen, B. R., B. R. Kreiser, P. F. Mickle, J. F. Schaefer, and S. B. Adams. 2008. Phylogenetic relationships among North American *Alosa* species (Clupeidae). Journal of Fish Biology 72:1188–1201.
- Boyadzhieva-Doychinova, D., F. Tserkova, M. Gevezova, D. Klisarova, and I. Denev. 2012. Application of variability in its sequences for studying biodiversity of Black Sea clupeid fishes (Pisces: Clupeidae). Acta Zoologica Bulgarica 4:95–102.
- Brown, J. W., C. Parins-Fukuchi, G. W. Stull, O. M. Vargas, and S. A. Smith. 2017. Bayesian and likelihood phylogenetic reconstructions of morphological traits are not discordant when taking uncertainty into consideration: a comment on Puttick et al. Proceedings of the Royal Society B 284:p.20170986.
- Brusatte, S. L., G. T. Lloyd, S. C. Wang, and M. A. Norell. 2014. Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. Current Biology 24:2386–2392.
- Bulgakov, G. P. 1926. To the knowledge of Caspian herrings. Description of a new species of the genus *Caspialosa*. Byulleten' Sredneaziatskii Gosudarstvennyi Universitet, Taschkent 13:27–39.

- Cantatore, P., M. Roberti, G. Pesole, A. Ludovico, F. Milella, M. N. Gadaleta, and C. Saccone. 1994. Evolutionary analysis of cytochrome b sequences in some Perciformes: evidence for a slower rate of evolution than in mammals. *Journal of Molecular Evolution* 39:589–597.
- Capobianco, A., H. T. Beckett, E. Steurbaut, P. D. Gingerich, G. Carnevale, and M. Friedman. 2020. Large-bodied sabre-toothed anchovies reveal unanticipated ecological diversity in early Palaeogene teleosts. *Royal Society Open Science* 7:192260.
- Casier, E. 1946. La faune ichthyologique de l'Yprésien de la Belgique. *Mémoires du Musée Royal d'Histoire Naturelle de Belgique* 104:1–267.
- Casier, E. 1965. Poissons fossils de la Série du Kwanga (Congo). *Sciences Geologiques. Memoire* 50:1–64.
- Castelnaud, F. L. 1875. Researches on the fishes of Australia. Philadelphia Centennial Expedition of 1876. *Intercolonial Exhibition Essays* 2:1–52.
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* 1:171–182.
- Chang, M. M. and C. C. Chou. 1977. On late Mesozoic fossil fishes from Zhejiang Province, China. *Memoirs, Institute of Vertebrate Palaeontology and Palaeoanthropology. Academia Sinica* 12:1–60.
- Chang, M. M. and J. G. Maisey. 2003. Redescription of *yEllimma branneri* and *yDiplomystus shengliensis*, and relationships of some basal clupeomorphs. *American Museum Novitates* 3404:1–35.
- Chapman, R. W., J. C. Patton, and B. Eleby. 1994. Comparisons of mitochondrial DNA variation in four alosid species as revealed by the total genome, the NADH dehydrogenase I and

- cytochrome b regions; pp. 249–262 in A. Beaumont (ed.), *Genetics and Evolution of Aquatic Organisms*. Chapman and Hall, London. 539 pp.
- Chavarie, L., K. L. Howland, and W. M. Tonn. 2013. Sympatric polymorphism in Lake Trout: the coexistence of multiple shallow-water morphotypes in Great Bear Lake. *Transactions of the American Fisheries Society* 142:814–823.
- Clausen, H. S. 1959. Denticipitidae, a new family of primitive Isospondylous teleosts from West African fresh-water. *Videnskabelige Meddelelser Naturhistorisk Forening, Kjøbenhavn* 121:141–151.
- Conrad, J. L. 2008. Phylogeny and systematics of Squamata (Reptilia) based on morphology. *Bulletin of the American Museum of Natural History* 310:1–182.
- Cope, E. D. 1877. A contribution to the knowledge of the ichthyological fauna of the Green River Shales. *Bulletin of United States Geological and Geographical Survey* 3:807–819.
- Cope, E. D. 1886. A contribution to the vertebrate paleontology of Brazil. *Proceedings of the American Philosophical Society* 23:3–4.
- Coscia, I., V. Rountree, J. J. King, W. K. Roche, and S. Mariani. 2010. A highly permeable species boundary between two anadromous fishes. *Journal of fish biology* 77:1137–1149.
- Critchlow, D. E., D. K. Pearl, and C. Qian. 1996. The triples distance for rooted bifurcating phylogenetic trees. *Systematic Biology* 45:323–334.
- Cunningham, C. W., H. Zhu, and D. M. Hillis. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. *Evolution* 52:978–987.
- Cuvier, G. 1816. *Le Règne Animal distribué d'après son organisation pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée*. Les reptiles, les poissons, les mollusques et les annelids, 532 pp.

- Cuvier, G. 1829. Le Règne Animal, distribué d'après son organisation, pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Edition 2. 406 pp.
- Czesny, S., J. Epifanio, and P. Michalak. 2012. Genetic divergence between freshwater and marine morphs of alewife (*Alosa pseudoharengus*): a 'next-generation' sequencing analysis. PloS ONE 7:e31803.
- Daget, J. 1984. Denticipitidae; p. 10 in J. Daget, J. P. Gosse, D. F. E. Thys van den Audenaerde (eds.), Check-list of the freshwater fishes of Africa. Cloffa 1. Paris, Orstom/Tervuren, MRAC, 410 pp.
- Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:10.1038/nmeth.2109.
- Dayrat, B. 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society 85:407–415.
- De Figueiredo, F. J. 2009. A new Clupeiform fish from the Lower Cretaceous (Barremian) of Sergipe-Alagoas Basin, northeastern Brazil. Journal of Vertebrate Paleontology 29:993–1005.
- De Kay, J. E. 1842. Zoology of New-York, or the New-York fauna; comprising detailed descriptions of all the animals hitherto observed within the state of New-York, with brief notices of those occasionally found near its borders, and accompanied by appropriate illustrations. Part IV. Fishes. W. & A. White & J. Visscher, Albany, 415 pp.
- Department of Fisheries and Ocean. Canada. 2020. Seafisheries landed quantity by province, 2018. Preliminary statistics. Electronic version 10/02/2020 accessed 20 May 2020.
- Deraniyagala, P. E. P. 1929. Ceylon sardines. Spolia Zeylanica (The Ceylon Journal of Science, Section B. Zoology and Geology) 15:31–47.

- Di Dario, F. 2002. Evidence supporting a sister-group relationship between Clupeoidea and Engrauloidea (Clupeomorpha). *Copeia* 2002:496–503.
- Di Dario, F. 2004. Homology between the recessus lateralis and cephalic sensory canals, with the proposition of additional synapomorphies for the Clupeiformes and the Clupeoidei. *Zoological Journal of the Linnean Society* 141:257–270.
- Di Dario, F. 2009. Chirocentrids as engrauloids: evidence from suspensorium, branchial arches, and infraorbital bones (Clupeomorpha, Teleostei). *Zoological Journal of the Linnean Society* 156:363–383.
- Di Dario, F. and M. C. De Pinna. 2006. The supratemporal system and the pattern of ramification of cephalic sensory canals in *Denticeps clupeoides* (Denticipitoidei, Teleostei): additional evidence for monophyly of Clupeiformes and Clupeoidei. *Papéis Avulsos de Zoologia* 46:107–123.
- Dodson, J. J., N. Aubin-Horth, V. Thériault, and D. J. Páez. 2013. The evolutionary ecology of alternative migratory tactics in salmonid fishes. *Biological Reviews* 88:602–625.
- Dornburg, A., Z. Su, and J. P. Townsend. 2018. Optimal rates for phylogenetic inference and experimental design in the era of genome-scale data sets. *Systematic Biology* 68(1):145–156.
- Doyle, J. A. and P. K. Endress. 2014. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: ANITA lines and relatives of Chloranthaceae. *International Journal of Plant Sciences* 175:555–600.
- Drooger, C. W. 1954. Remarks on the species concept in paleontology. *The Micropaleontologist* 8:23–26.

- Dubertret, L. 1959. Contribution à la stratigraphie et à la paléontologie du Crétacé et du Nummulitique de la marge N. W. De la péninsule arabique. Notes et Mémoires sur le Moyen Orient, Paris 7:193–220.
- Dubertret, L. and H. Vautrin. 1937. Révision de la stratigraphie du Crétacé du Liban. Notes et Mémoires, Haut-Commissariat de la République Française en Syrie et au Liban 2:43–73.
- Dubertret, L. 1966. Liban, Syrie et bordure de pays voisins. 1^é partie: tableau stratigraphique. Notes et Mémoires sur le Moyen Orient, Paris 8:249–358.
- Dupuis, J. R., A. D. Roe, and F. A. H. Sperling. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology* 21:4422–4436.
- Durand, J. 1940. Notes sur quelques poissons d'espèces nouvelles ou peu connues des eaux douces cambodgiennes. Institut Océanographique de l'Indochine Note 36:1–40.
- Eastman, C. R. 1912. Tertiary fish-remains from Spanish Guinea in West Africa. *Annals of the Carnegie Museum* 8:370–378.
- Eclipse Foundation. 2016. Eclipse Neon v. 4.6. Available at <https://eclipse.org/downloads/index-developer.php>. Accessed on September 28, 2018.
- Egan, J. P., D. D. Bloom, C.-H. Kuo, M. P. Hammer, P. Tongnunui, S. P. Iglesias, M. Sheaves, C. Grudpan, and A. M. Simons. 2018. Phylogenetic analysis of trophic niche evolution reveals a latitudinal herbivory gradient in Clupeoidei (herrings, anchovies, and allies). *Molecular Phylogenetics and Evolution* 124:151–161.
- Eichwald, C. E. von. 1838. Faunae Caspii maris primitiae. *Bulletin de la Société Impériale des Naturalistes de Moscou* 11:125–174.
- Eyo, J. E. 2003. Congeneric discrimination of morphometric characters among members of the Pisces Genus *Clarias* (Clariidae) in Anambra River, Nigeria. *The Zoologist* 2:1–17.

- FAO Fisheries department. 2019. FAO yearbook. Fishery and Aquaculture Statistics 2017/FAO annuaire. Statistiques des pêches et de l'aquaculture 2017/ FAO anuario. Estadísticas de pesca y acuicultura 2017. Rome/Roma, 109 pp.
- FAO. 2016. FAO Yearbook. Fishery and Aquaculture Statistics. 2014. Rome/Roma, FAO. 105 pp.
- Faria, R., A. Pinheiro, T. Gabaldón, S. Weiss and P. Alexandrino. 2011. Molecular tools for species discrimination and detection of hybridization between two closely related Clupeid fishes *Alosa alosa* and *A. fallax*. *Journal of Applied Ichthyology* 27:16–20.
- Faria, R., B., Wallner, S. Weiss, and P. Alexandrino. 2004. Isolation and characterization of eight dinucleotide microsatellite loci from two closely related clupeid species (*Alosa alosa* and *A. fallax*). *Molecular Ecology Notes* 4:586–588.
- Faria, R., S. Weiss, and P. Alexandrino. 2006. A molecular phylogenetic perspective on the evolutionary history of *Alosa* spp. (Clupeidae). *Molecular Phylogenetics and Evolution* 40:298–304.
- Faria, R., S. Weiss, and P. Alexandrino. 2012. Comparative phylogeography and demographic history of European shads (*Alosa alosa* and *A. fallax*) inferred from mitochondrial DNA. *BMC Evolutionary Biology* 12:10.1186/1471-2148-12-194.
- Felsenstein, J. 1978. The number of evolutionary trees. *Systematic Zoology* 27:27–33.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368–376.
- Felsenstein, J. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer associates, 580 pp.

- Feng, C., M. Pettersson, S. Lamichhaney, C.-J. Rubin, N. Rafati, M. Casini, A. Folkvord, and L. Andersson. 2017. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *eLife*:6:e23907.
- Fennessy, S. T., P. Pradervand, and P. A. de Bruyn. 2010. Influence of the sardine run on selected nearshore predatory teleosts in KwaZuluNatal. *African Journal of Marine Science* 32:375–382.
- Figueiredo, F. J. and D. R. Ribeiro. 2016. Relationships of †*Codoichthys carnavaalii* Santos, 1994 (Teleostei, Clupeomorpha, † Ellimmichthyiformes) from the Late Aptian of São Luís-Grajaú Basin, NE Brazil. *Anais da Academia Brasileira de Ciências* 88:1277–1307.
- Figueiredo, F. J. and D. R. Ribeiro. 2017. Reassessment and Relationships of †*Scutatuspinosus itapagipensis* (Teleostei, Clupeomorpha, † Ellimmichthyiformes) from the Neocomian of Recôncavo Basin, Northeastern Brazil. *Anais da Academia Brasileira de Ciências* 89:799–823.
- FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3:1084–1092.
- Forey, P. L. 1973. A primitive clupeomorph fish from the middle Cenomanian of Hakel, Lebanon. *Canadian Journal of Earth Sciences* 10:1302–1318.
- Forey, P. L. 1975. A fossil clupeomorph fish from the Albian of the Northwest Territories of Canada, with notes on cladistic relationships of clupeomorphs. *Journal of Zoology* 175:151–177.
- Forey, P. L. 2004. A three-dimensional skull of a primitive clupeomorph from the Cenomanian English Chalk and implications for the evolution of the clupeomorph *acustico-lateralis*

- system; pp. 405–427 in G. Arratia, and A. Tintori (eds.), *Mesozoic Fishes 3—Systematics, Paleoenvironments and Biodiversity*. Verlag Dr. Friedrich Pfeil, Munich, 649 pp.
- Forey, P. L., L. Yi, C. Patterson, and C. E. Davies. 2003. Fossil fishes from the Cenomanian (Upper Cretaceous) of Namoura, Lebanon. *Journal of Systematic Palaeontology* 1:227–330.
- Forsskål, P. S. 1775. *Descriptiones animalium, avium, amphibiorum, piscium, insectorum, vermium; quae in itinere orientali observavit. Post mortem auctoris edidit Carsten Niebuhr*. Hauniae. 164 pp.
- Fowler, H. W. 1911. Notes on clupeoid fishes. *Proceedings of the Academy of Natural Sciences of Philadelphia* 63:204–221.
- Fowler, H. W. 1934. Descriptions of new fishes obtained 1907 to 1910, chiefly in the Philippine Islands and adjacent seas. *Proceedings of the Academy of Natural Sciences of Philadelphia* 85:233–367.
- Fricke, R., W. N. Eschmeyer, and J. D. Fong. 2020. Eschmeyer’s catalog of fishes: species by family/subfamily. <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>. Electronic version accessed 28 May 2020.
- Froese, R. and D. Pauly. 2019. FishBase. World Wide Web electronic publication. www.fishbase.org. Electronic version 12/2019 accessed 28 May 2020.
- Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology and evolution* 27:480–488.

- Gaudant, M., and J. Gaudant. 1971. Une nouvelle espèce de *Diplomystus* (Poisson téléostéen) dans le Crétacésupérieur du Sudtunisien. *Bulletin de la Société géologique de France* 13:156–159.
- Gill, T. 1872. Arrangement of the families of fishes, or Classes Pisces, Marsipobranchii, and Leptocardi. *Smithsonian Miscellaneous Collections* 11:1–49.
- Gill, T. N. 1861. Synopsis of the subfamily of Clupeinae, with descriptions of new genera. *Proceedings of the Academy of Natural Sciences of Philadelphia* 13:33–38.
- Girgensohn, O. G. L. 1846. Anatomie und Physiologie des Fisch-Nervensystems. *Mémoires de l'Académie Impériale des Sciences de Saint Pétersbourg* 5:275–589.
- Goloboff, P. A. 1999. Analyzing large datasets in reasonable times: solutions for composite optima. *Cladistics* 15:415–428.
- Goloboff, P. A. and J. S. Arias. 2019. Likelihood approximations of implied weights parsimony can be selected over the Mk model by the Akaike information criterion. *Cladistics* 35:695–716.
- Goloboff, P. A. and S. A. Catalano. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* 32:221–238.
- Goloboff, P. A., A. Torres, and J. S. Arias. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* 34:407–437.
- Goloboff, P. A., J. S. Farris, and K. C. Nixon. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774–786.

- Goode, G. B. 1878. A revision of the American species of the genus *Brevoortia*, with a description of a new species from the Gulf of Mexico. *Proceedings of the United States National Museum* 1:30–42.
- Goodrich, E. S. 1909. Vertebrata Craniata (First Fascicle: Cyclostomes and Fishes). In *A Treatise on Zoology*; in R. Lankester (ed.) Part IX. Black, London, 518 pp.
- Gosline, W. A. 1960. Contributions toward a classification of modern isospondylous fishes. *Bulletin of the British Museum of Natural History* 6:325–365.
- Gosline, W. A. 1980. The evolution of some structural systems with reference to the interrelationships of modern lower teleostean fish groups. *Japanese Journal of Ichthyology* 27:1–28.
- Gosse, P. H. 1851. *A naturalist's sojourn in Jamaica*. Longman, Brown, Green and Longmans, London, 508 pp.
- Grande, L. 1982. A revision of the fossil genus *Diplomystus*, with comments on the interrelationships of clupeomorph fishes. *American Museum Novitates* 2728:1–34.
- Grande, L. 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. *Bulletin of the American Museum of Natural History* 181:231–372.
- Grande, L. and G. J. Nelson. 1985. Interrelationships of fossil and Recent anchovies (Teleostei: Engrauloidea) and description of a new species from the Miocene of *Cyprus*. *American Museum Novitates* 2826:1–16.
- Grande, T., W. C. Borden, and W. L. Smith. 2013. Limits and relationships of Paracanthopterygii: a molecular framework for evaluating past morphological hypotheses. *Mesozoic fishes* 5:385–418.

- Grant, W.S., F. Lecomte and B.W. Bowen. 2010. Biogeographical contingency and the evolution of tropical anchovies (genus *Cetengraulis*) from temperate anchovies (genus *Engraulis*). *Journal of Biogeography* 37:1352–1362.
- Gray, J. E. 1830. Illustrations of Indian zoology; chiefly selected from the collection of Major-General Hardwicke, F.R.S. 1:1–202.
- Greenwood, P. H. 1960. Fossil denticipitid fishes from east Africa. *Bulletin of British Museum of Natural History* 5:1–11.
- Greenwood, P. H. 1968. The osteology and relationships of the Denticipitidae, a family of clupeomorph fishes. *Bulletin of the British Museum of Natural History (Zoology)* 16:215–273.
- Greenwood, P. H., D. E. Rosen, S. H. Weitzman, and G. S. Myers. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History* 131:339–455.
- Grimm, O. von. 1901. The herrings of the Sea of Azov. *Vestnik Rybopromyshlennosti* 16:57–70.
- Guillerme T. and N. Cooper. 2016. Effects of missing data on topological inference using a Total Evidence approach. *Molecular Phylogenetics and Evolution* 94:146–158.
- Guindon, S. and O. Gascuel. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52:696–704.
- Guindon, S., J. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59:307–321.
- Günther, A. 1867. On the fishes of the states of Central America, founded upon specimens collected in fresh and marine waters of various parts of that country by Messrs. Salvin and

- Godman and Capt. J. M. Dow. Proceedings of the Zoological Society of London 1866:600–604.
- Günther, A. 1868. Catalogue of the fishes in the British Museum. Catalogue of the Physostomi, containing the families Heteropygii, Cyprinidae, Gonorhynchidae, Hyodontidae, Osteoglossidae, Clupeidae,... Halosauridae, in the collection of the British Museum. 7:1–512.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuclear Acids Symposium Series 41:95–98.
- Halliday, T. J., M. dos Reis, A. U. Tamuri, H. Ferguson-Gow, Z. Yang, and A. Goswami. 2019. Rapid morphological evolution in placental mammals post-dates the origin of the crown group. Proceedings of the Royal Society B 286:p.20182418.
- Hamilton-Buchanan, F. 1822. An account of the fishes found in the river Ganges and its branches. Edinburgh and London, 405 pp.
- Hasselman, D. J., D. Ricard, and P. Bentzen. 2013. Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. Molecular ecology 22:1558–1573.
- Hata, H., S. Lavoué, and H. Motomura. 2020. *Stolephorus babarani*, a new species of anchovy (Teleostei: Clupeiformes: Engraulidae) from Panay Island, central Philippines. Zootaxa 4718:509–520.
- Hay, M. J., S. L. Cumbaa, A. M. Murray, and A. G. Plint. 2007. A new paraclupeid fish (Clupeomorpha, Ellimmichthyiformes) from a muddy marine prodelta environment: middle Cenomanian Dunvegan Formation, Alberta, Canada. Canadian Journal of Earth Sciences 44:775–790.

- Heled, J. and R. R. Bouckaert. 2013. Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology* 13:10.1186/1471-2148-13-221.
- Hemleben, C. 1977. Rote Tiden und die oberkretazischen Plattenkalke im Libanon. *Neues Jahrbuch für Geologie und Paläontologie, Monatshefte* 1977:239–255.
- Hermus, C. R. and M. V. H. Wilson. 2004. A new species of *Erichalcis* from Arctic Canada, with a revised diagnosis of the genus; pp. 405–427 in G. Arratia, and A. Tintori (eds.), *Mesozoic Fishes 3—Systematics, Paleoenvironments and Biodiversity*. Verlag Dr. Friedrich Pfeil, Munich, 649 pp.
- Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* 383:130.
- Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47:3–8.
- Hillis, D. M., D. D. Pollock, J. A. McGuire, and D. J. Zwickl. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Systematic Biology* 52:124–126.
- Hoestlandt, H. 1991. *The freshwater fishes of Europe*. Vol. 2. Clupeidae, Anguillidae. Aula, Wiesbaden. 447 pp.
- Hoff, M., S. Orf, B. Riehm, D. Darriba, and A. Stamatakis. 2016. Does the choice of nucleotide substitution models matter topologically? *BMC Bioinformatics* 17:10.1186/s12859-016-0985-x.
- Höhna, S. and A. J. Drummond. 2012. Guided tree topology proposals for Bayesian phylogenetic inference. *Systematic Biology* 61:1–11.
- Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65:726–736.

- Holder, M. T., J. Sukumaran, and P. O. Lewis. 2008. A justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. *Systematic Biology* 57:814–821.
- Hubbs, C. L. 1929. The generic relationships and nomenclature of the California sardine. *Proceedings of the California Academy of Sciences* 18:261–265.
- Hückel, U. 1970. Die Fischeschiefer von Hakel und Hjoula in der Oberkreide des Libanon. *Neues Jahrbuch für Geologie und Paläontologie, Abhandlungen, Stuttgart* 135:113–149.
- Hückel, U. 1974. Geochemischer Vergleich der Plattenkalke Solnhofens und des Libanon mit anderer Kalken. *Neues Jahrbuch für Geologie und Paläontologie, Abhandlungen, Stuttgart* 145:279–305.
- Huelsenbeck, J. P. 1995. Performance of phylogenetic methods in simulation. *Systematic Biology* 44:17–48.
- Hughes, L. C., G. Ortí, Y. Huang, Y. Sun, C. C. Baldwin, A. W. Thompson, D. Arcila, R. Betancur-R, C. Li, L. Becker, and N. Bellora. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences* 115:6249–6254.
- Jenyns, L. 1842. Fish; Part IV in C. Darwin (ed.), *The zoology of the voyage of H. M. S. Beagle, under the command of Captain Fitzroy, R. N., during the years 1832 to 1836*. Smith, Elder, and Co, London, 172 pp.
- Jolly, M. T., P. S. Maitland, and M. J. Genner. 2011. Genetic monitoring of two decades of hybridization between allis shad (*Alosa alosa*) and twaite shad (*Alosa fallax*). *Conservation Genetics* 12:1087–1100.

- Jolly, M. T., M. W. Aprahamian, S. J. Hawkins, P. A. Henderson, R. Hillman, N. O'Maoiléidigh, and M. J. Genner. 2012. Population genetic structure of protected allis shad (*Alosa alosa*) and twaite shad (*Alosa fallax*). *Marine biology* 159:675–687.
- Jones, A. W., E. P. Palkovacs, and D. M. Post. 2013. Recent parallel divergence in body shape and diet source of alewife life history forms. *Evolutionary Ecology* 27:1175–1187.
- Jordan, D. S. 1907. The fossil fishes of California: with supplementary notes on other species of extinct fishes 5:95–145.
- Jordan, D. S. 1910. Description of a collection of fossil fishes from the bituminous shales at Riaco Doce, state of Alagoas, Brazil. *Annals of the Carnegie Museum* 7:23–24.
- Jordan, D. S. 1913. *Ellimma*, a genus of fossil herrings. *Proceedings of the Biological Society of Washington* 26:1–79.
- Jordan, D. S. 1923. A Classification of fishes including families and genera as far as known. Stanford University Publications, University Series, Biological Sciences 3:77–243.
- Jordan, D. S. and B. W. Evermann. 1896. Description of a new species of shad (*Alosa alabamae*) from Alabama; pp. 203–205 in B. W. Evermann, Report to the United States Commission on Fisheries. G.P.O., Washington, 610 pp.
- Jordan, D. S. and B. W. Evermann. 1896. Description of a new species of shad (*Alosa alabamae*) from Alabama; pp. 203–205 in B. W. Evermann, Report to the United States Commission on Fisheries. G.P.O., Washington, 610 pp.
- Jordan, D. S. and C. H. Gilbert. 1882a. Catalogue of the fishes collected by Mr. John Xantus at Cape San Lucas, which are now in the United States National Museum, with descriptions of eight new species. *Proceedings of the United States National Museum* 5:353–371.

- Jordan, D. S. and C. H. Gilbert. 1882b. Descriptions of thirty-three new species of fishes from Mazatlan, Mexico. *Proceedings of the United States National Museum* 4:338–365.
- Jordan, D. S. and J. O. Snyder. 1900. A list of fishes collected in Japan by Keinosuke Otaki, and by the United States steamer Albatross, with descriptions of fourteen new species. *Proceedings of the United States National Museum* 23:335–380.
- Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules; pp. 21–132 in H. N. Munro(ed.), *Mammalian protein metabolism*. Academic Press, New York, 590 pp.
- Kahilainen, K. K., A. Siwertsson, K. Ø. Gjelland, R. Knudsen, T. Bøhn, and P. A. Amundsen. 2011. The role of gill raker number variability in adaptive radiation of coregonid fish. *Evolutionary Ecology* 25:573–588.
- Kaiser, H. F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20:141–151.
- Kass, R. E. and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–795.
- Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30:772–780.
- Katoh, K., K. Misawa, K. I. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30:3059–3066.
- Kelchner, S. A. and M. A. Thomas. 2007. Model use in phylogenetics: nine key questions. *Trends in Ecology and Evolution* 22:87–94.
- Keskin, E. and H. H. Atar. 2013. DNA barcoding commercially important fish species of Turkey. *Molecular Ecology Resources* 13:788–797.

- Khalloufi, B., R. Zaragüeta-Bagils, and H. Lelièvre. 2010. *Rhombichthys intoccabilis*, gen. et sp. nov. (Ellimmichthyiformes, Clupeomorpha, Teleostei), from the Cenomanian (Upper Cretaceous) of Ein Yabrud, Middle East: anatomical description and phylogenetic implications. *Journal of Vertebrate Paleontology* 30:57–67.
- Kim, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Systematic Biology* 47:43–60.
- Kitching, I. J., P. L. Forey, C. J. Humphries, and D. D. William. 1998. *Cladistics: the theory and practise of parsimony analysis* (2nd edition). Oxford University Press, USA, 228 pp.
- Klopfstein, S., L. Vilhelmsen, and F. Ronquist. 2015. A nonstationary Markov model detects directional evolution in hymenopteran morphology. *Systematic Biology* 64:1089–1103.
- Klunzinger, C. B. 1879. Die v. Müller'sche Sammlung australischer Fische. *Anzeiger der Kaiserlichen Akademie der Wissenschaften, Wien, Mathematisch-Naturwissenschaftliche Classe* 16:254–261.
- Kottelat, M. 1997. European freshwater fishes. An heuristic checklist of the freshwater fishes of Europe (exclusive of former USSR), with an introduction for non-systematists and comments on nomenclature and conservation. *Biologia, Bratislava, Section Zoology* 52:1–271.
- Kottelat, M. and J. Freyhof. 2007. *Handbook of European freshwater fishes*. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany, 646 pp.
- Krijgsman, W., M. Stoica, I. Vasiliev, and V. V. Popov. 2010. Rise and fall of the Paratethys Sea during the Messinian Salinity Crisis. *Earth and Planetary Science Letters* 290:183–191.
- Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* 11:459–468.

- Kuhner, M. K. and J. Yamato. 2014. Practical performance of tree comparison metrics. *Systematic Biology* 64:205–214.
- Lacepède, B. G. E. 1800. Histoire naturelle des poissons. 2:1–632.
- Lacepède, B. G. E. 1803. Histoire naturelle des poissons. Tome cinquième. Chez Plassan, Paris, 803 pp.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular biology and evolution* 34:772–773.
- Latrobe, B. H. 1802. A drawing and description of the *Clupea Tyrannus* and *Oniscus Prægustator*. *Transactions of the American Philosophical Society* 5:77–81.
- Latrobe, B. H. 1802. A drawing and description of the *Clupea tyrannus* and *Oniscus praegustator*. *Transactions of the American Philosophical Society* 5:77–81.
- Lavoué, S., M. Miya, A. Kawaguchi, T. Yoshino, and M. Nishida. 2008. The phylogenetic position of an undescribed paedomorphic clupeiform taxon: mitogenomic evidence. *Ichthyological research* 55:328–334.
- Lavoué, S., M. Miya, and M. Nishida. 2010. Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced miniaturization in the order Clupeiformes. *Molecular Phylogenetics and Evolution* 56:480–485.
- Lavoué, S., M. Miya, K. Saitoh, N. B. Ishiguro. 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Molecular Phylogenetics and Evolution* 43:1096–1105.

- Lavoué, S., M. Miya, P. Musikasinthorn, W. J. Chen, and M. Nishida. 2013. Mitogenomic evidence for an Indo-West Pacific origin of the clupeioidi (Teleostei: Clupeiformes). *PLoS ONE* 8:e56485.
- Lavoué, S., P. Konstantinidis, and W. J. Chen. 2014. Progress in clupeiform systematics; pp. 3–42 in K. Ganas (ed.), *Biology and ecology of sardines and anchovies*, CRC Press, 394 pp.
- Leache, A. D., B. L. Banbury, J. Felsenstein, A. N. M. de Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64:1032–1047.
- Lecointre, G., H. Philippe, H. L. V. Lê, and H. Le Guyader. 1993. Species sampling has a major impact on phylogenetic inference. *Molecular Phylogenetics and Evolution* 2:205–224.
- Lee, M. S. Y., J. Soubrier, and G. D. Edgecombe. 2013. Rates of phenotypic and genomic evolution during the Cambrian explosion. *Current Biology* 23:1889–1895.
- Lemmon, A. R. and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology* 53:265–277
- Lesueur, C. A. 1818. Descriptions of several new species of North American fishes. *Journal of the Academy of Natural Sciences, Philadelphia* 1:222–235.
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Li, C. and G. Orti. 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 44:386–398.

- Li, W. L. S. and A. J. Drummond. 2011. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* 29:751–761.
- Linck, H. F. 1790. Versuch einer Eintheilung der Fische nach den Zähnen. *Magazin für das Neueste aus der Physik und Naturgeschichte*, Gotha 6:28–38.
- Linck, H. F. 1790. Versuch einer Eintheilung der Fische nach den Zähnen. *Magazin für das Neueste aus der Physik und Naturgeschichte*, Gotha 6:28–38.
- Lindsey, C. C. 1981. Stocks are chameleons: plasticity in gill rakers of coregonid fishes. *Canadian Journal of Fisheries and Aquatic Sciences* 38:1497–1506.
- Linnaeus, C. 1735. *Systema Naturae, sive Regna Tria Naturae Systematice Proposita per Classes, Ordines, Genera, and Species*. Theodorum Haak, Lugduni Batavorum, Leiden, 13 pp.
- Linnaeus, C. 1758. *Systema Naturae per Regna Tria Naturae, secundum Classes, Ordines, Genera, Species, cum Characteribus, Differentiis, Synonymis, Locis*. Tonus I. Editio Decima, Reformata. Laurentii Salvii, Holmiae, 824 pp.
- Linnaeus, C. 1761. *Fauna Suecicasitens animalia Suecicaeregni: quadrupedia, aves, amphibian, pisces, insect, vermes*. Impensis Direct. Laurentii Salvii, Stockholm, 579 pp.
- Luo, A., H. Qiao, Y. Zhang, W. Shi, S. Y. Ho, W. Xu, A. Zhang, and C. Zhu. 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evolutionary Biology* 10:10.1186/1471-2148-10-242.
- Macleay, W. 1879. On the Clupeidae of Australia. *Proceedings of the Linnean Society of New South Wales* 4:363–385.

- Maddison, W. P. and D. R. Maddison. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.6 <http://mesquiteproject.org>
- Maisey, J. G. 1993. A new clupeomorph fish from the Santana Formation (Albian) of NE Brazil. *American Museum Novitates* 3076:1–15.
- Malabarba, M. C., D. A. do Carmo, I. Gómez-Pérez, and J. V. de Queiroz-Neto. 2004. A new clupeomorph fish from the Cretaceous Maceió Formation, Alagoas Basin, NE Brazil. *Neues Jahrbuch für Geologie und Paläontologie – Abhandlungen* 233:255–274.
- Marramà, G. and G. Carnevale. 2015. Eocene round herring from Monte Bolca, Italy. *Acta Palaeontologica Polonica* 60:701–710.
- Marramà, G. and G. Carnevale. 2017. The relationships of *Gasteroclupea branisai* Signeux, 1964, a freshwater double-armored herring (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous-Paleocene of South America. *Historical Biology* 29:904–917.
- Marramà, G., A. F. Bannikov, J. Kriwet, and G. Carnevale. 2019. An Eocene paraclupeid fish (Teleostei, Ellimmichthyiformes) from Bolca, Italy: the youngest marine record of double-armoured herrings. *Papers in Palaeontology* 5:83–98.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17:10–12.
- Martin, A. P., G. J. P. Naylor, and S. R. Palumbi. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* 357:153–155.
- McBride, M. C., T. V. Willis, R. G. Bradford, and P. Bentzen. 2014. Genetic diversity and structure of two hybridizing anadromous fishes (*Alosa pseudoharengus*, *Alosa aestivalis*) across the northern portion of their ranges. *Conservation Genetics* 15:1281–1298.

- McCart, P. and B. Andersen. 1967. Plasticity of gillraker number and length in *Oncorhynchus nerka*. *Journal of the Fisheries Research Board of Canada* 24:1999–2002.
- McDowall, R. M. 2001. Diadromy, diversity and divergence: implications for speciation process. *Fish and Fisheries* 2:278–286.
- McDowall, R. M. 2003. Shads and anadromy: implications for ecology, evolution, and biogeography; pp. 11–23 in K. E. Limburg and J. R. Waldman (eds.), *Biodiversity, Status, and Conservation of the World's Shads*. Baltimore, MD: American Fisheries Society, 370 pp.
- Melvin, G. D., M. I. Dadswell, and A. McKennie. 1992. Usefulness of meristic and morphometric characters in discriminating populations of American shad (*Alosa sapidissima*) (Ostreichthyes:Clupeidae) inhabiting a marine environment. *Canadian Journal of Fisheries and Aquaculture Sciences* 49:266–280.
- Mezhzherin, S. V. and L. V. Fedorenko. 2005. Genetic structure of the spawning population of Danube shad, *Alosa pontica*, Eichwaldt 1838 (Clupeiformes, Alosiinae). *Tsitologiya i genetika* 39:35–40.
- Mezhzherin, S. V., L. V. Fedorenko, and D. B. Verlatyi. 2009. Differentiation and allozyme variability of shads genus *Alosa* (Clupeiformes, Alosiinae) from Azov-Black Sea basin. *Cytology and genetics* 43:118–122.
- Michel, C., B. J. Hicks, K. N. Stolting, A. C. Clarke, M. I. Stevens, R. Tana, A. Meyer, and M. R. van den Heuvel. 2008. Distinct migratory and non-migratory ecotypes of an endemic New Zealand eleotrid (*Gobiomorphus cotidianus*)—implications for incipient speciation in island freshwater fish species. *BMC Evolutionary Biology* 8:49–63.

- Miller, R. R. 1960. Systematics and biology of the gizzard shad (*Dorosoma cepedianum*) and related fishes. *Fisheries* 1:371–392.
- Minh, B. Q., M. A. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30:1188–1195.
- Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology* 52:674–683.
- Mitchill, S. L. 1814. Report, in part, of Samuel L. Mitchill, M. D., Professor of Natural History, on the fishes of New-York. D. Carlisle, New York, 28 pp.
- Miyashita, T. 2010. Unique occipital articulation with the first vertebra found in pristigasterids, chirocentrids, and clupeids (Teleostei: Clupeiformes: Clupeoidei). *Ichthyological Research* 57:121–132.
- Moritz, C. and C. Cicero. 2004. DNA barcoding: promises and pitfalls. *Public Library of Science Biology* 2:1529–1531.
- Müller, J. 1845. Über den Bau und die Grenzen der Ganoiden und über das natürlichen System der Fische. *Abhandlungen Akademie der Wissenschaften*, Berlin 1844:117–216.
- Murray, A. M. and M. V. H. Wilson. 2011. A new species of *Sorbinichthys* (Teleostei: Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco. *Canadian Journal of Earth Sciences* 48:1–9.
- Murray, A. M., and M. V. H. Wilson. 2013. Two new paraclupeid fishes (Clupeomorpha: Ellimmichthyiformes) from the Late Cretaceous of Morocco; pp. 267–290 in G. Arratia, H.-P. Schultze, and M. V. H. Wilson (eds.), *Mesozoic Fishes 5 – Global Diversity and Evolution*. Verlag Dr. Friederich Pfeil, Munich, Germany, 560 pp.

- Murray, A. M., E. L. Simons, and Y. S. Attia. 2005. A new clupeid fish (Clupeomorpha) from the Oligocene of Fayum, Egypt, with notes on some other fossil clupeomorphs. *Journal of Vertebrate Paleontology* 25:300–308.
- Murray, A. M., M. G. Newbrey, A. G. Neuman, and D. B. Brinkman. 2016. New articulated osteoglossomorph from Late Cretaceous freshwater deposits (Maastrichtian, Scollard Formation) of Alberta, Canada. *Journal of Vertebrate Paleontology* 36:e1120737.
- Murray, A. M., O. Vernygora, S. Japundžić, J. Radovčić, M. V. H. Wilson, D. Bardack, and T. Grande. 2016. Relationships of the species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) and the description of a new species from the Cretaceous of Dalmatia, Croatia. *Journal of Vertebrate Paleontology* 36:e1226851.
- Narum, S. R., C. Contor, A. Talbot, and M. S. Powell. 2004. Genetic divergence of sympatric resident and anadromous forms of *Oncorhynchus mykiss* in the Walla Walla River, USA. *Journal of Fish Biology* 65:471–488.
- Nelson, G. J. 1970. The hyobranchial apparatus of teleostean fishes of the families Engraulidae and Chirocentridae. *American Museum Novitates* 2410:1–30.
- Nelson, G. J. 1973. Relationships of clupeomorphs, with remarks on the structure of the lower jaw in fishes; pp.333–349 in P. H. Greenwood, R. S. Miles, and C. Patterson (eds.), *Interrelationships of fishes*. Academic Press, London, 536 pp.
- Nelson, G. J. 1983. *Anchoa argentivittata*, with notes on other eastern Pacific anchovies and the Indo-Pacific genus *Encrasicholina*. *Copeia* 1983:48–54.
- Nelson, G. J. 1984. Notes on the rostral organ of anchovies (Family Engraulidae). *Japanese Journal of Ichthyology* 31:86–87.

- Nelson, G. J. 1986. Identity of the Anchovy *Engraulis clarki* with notes on the species-groups of *Anchoa*. *Copeia* 1986:891–902.
- Nelson, G. J. and M. N. Rothman. 1973. The species of gizzard shads (Dorosomatinae) with particular reference to the Indo-Pacific region. *Bulletin of the American Museum of Natural History* 150:131–206.
- Nelson, J. S., T. C. Grande, and M. V. H. Wilson. 2016. *Fishes of the World*. Fifth edition. John Wiley and Sons, Hoboken, New Jersey, 707 pp.
- Newbrey, M. G., A. M. Murray, D. B. Brinkman, M. V. H. Wilson, and A. G. Neuman. 2010. A new articulated freshwater fish (Clupeomorpha, Ellimmichthyiformes) from the Horseshoe Canyon Formation, Maastrichtian, of Alberta, Canada. *Canadian Journal of Earth Sciences* 47:1183–1196.
- Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274.
- Nolan, K. A., J. R. Waldman, and I. Wirgin. 2003. Intraspecific and interspecific molecular differentiation of American shad and Alabama shad: a synthesis; pp. 295–302 in K. E. Limburg and J. R. Waldman (eds.), *Biodiversity, Status, and Conservation of the World's Shads*. American Fisheries Society, Symposium 35, Maryland, 370 pp.
- Nøttestad, L. and B. E. Axelsen. 1999. Herring schooling manoeuvres in response to killer whale attacks. *Canadian Journal of Zoology* 77:1540–1546.
- Novacek, M. and Q. Wheeler. 1992. Extinct taxa: Accounting for 99.999% of the earth's biota; pp. 1–16 in M. Novacek and Q. Wheeler (eds.), *Extinction and Phylogeny*. New York, Columbia University Press, 253 pp.

- Nye, T. M. W., P. Lio, and W. R. Gilks. 2006. A novel algorithm and web based tool for comparing two alternative phylogenetic trees. *Bioinformatics* 22:117–119.
- O'Donoghue, S. H., L. Drapeau, and V. M. Peddemors. 2010. Broad-scale distribution patterns of sardine and their predators in relation to remotely sensed environmental conditions during the KwaZuluNatal sardine run. *African Journal of Marine Science* 32:279–291.
- O'Donoghue, S. H., L. Drapeau, S. F. J. Dudley, and V. M. Peddemors. 2010. The KwaZulu-Natal sardine run: shoal distribution in relation to nearshore environmental conditions, 1997–2007. *African Journal of Marine Science* 32:293–307.
- O'Donoghue, S. H., P. A. Whittington, B. M. Dyer, and V. M. Peddemors. 2010. Abundance and distribution of avian and marine mammal predators of sardine observed during the 2005 KwaZuluNatal sardine run survey. *African Journal of Marine Science* 32:361–374.
- Ogilby, J. D. 1892. On some undescribed reptiles and fishes from Australia. *Records of the Australian Museum* 2:23–26.
- Ogilby, J. D. 1897. Rough-backed herrings, in "Notes and Exhibits" section. *Proceedings of the Linnean Society of New South Wales* 21:504–505.
- O'Reilly J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018a. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* 61:105–118.
- O'Reilly J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018b. Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff et al. *Palaeontology* 61:631–635.
- O'Reilly J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of

- precision in the estimation of phylogeny from discrete morphological data. *Biological Letters* 12:20160081.
- O'Reilly, J. E. and P. C. J. Donoghue. 2018. The efficacy of consensus tree methods for summarizing phylogenetic relationships from a posterior sample of trees estimated from morphological data. *Systematic Biology* 67:354–362.
- O'Reilly, J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* 61:105–118.
- O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biological Letters* 12:20160081.
- Østbye, K., T. Bernatchez, T. F. Næsje, K. – J. M. Himberg, and K. Hindar. 2005. Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology* 14:4371–4387.
- Palkovacs, E. P., K. B. Dion, D. M. Post, and A. Caccone. 2008. Independent evolutionary origins of landlocked alewife populations and rapid parallel evolution of phenotypic traits. *Molecular Ecology* 17:582–597.
- Paradis, E. and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.
- Parrish, J. K. 1993. Comparison of the hunting behavior of four piscine predators attacking schooling prey. *Ethology* 95:233–246.

- Patterson, C. 1967. Are the teleosts a polyphyletic group. *Colloques Internationaux du Centre National de la Recherche Scientifique* 163:93–109.
- Patterson, C. and D. E. Rosen. 1977. Review of ichthyodectiform and other Mesozoic teleost fishes, and the theory and practice of classifying fossils. *Bulletin of the American Museum of Natural History* 158:81–172.
- Pavlov, P. I. 1959. Shads of genus *Alosa* in the North-western Part of Black Sea. Kyiv: Akademia Nauk. UkrSSR, 252 pp.
- Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135.
- Pictet, F. J. 1850. Description de quelques poissons fossiles du Mont Liban. Jules-Guillaume Fick, 59 pp.
- Poe, S., and J. J. Wiens. 2000. Character selection and the methodology of morphological phylogenetics; pp. 20–36 in J. J. Wiens (ed.), *Phylogenetic analysis of morphological data*. Washington, D.C., Smithsonian Institution Press, 232 pp.
- Poey, F. 1865. Peces nuevos de la Isla de Cuba. *Repertorio Fisico-Natural de la Isla de Cuba* 1:181–192.
- Poey, F. 1867. Peces Cubanos especies nuevas. *Repertorio Fisico-Natural de la Isla de Cuba* 2:229–245.
- Pollock, D. D., D. J. Zwickl, J. A. McGuire, D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51:664–671.
- Popov, S. V., F. Rögl, A. Y. Rozanov, F. F. Steininger, I. G. Shcherba, and M. Kovac. 2004. Lithological-Paleogeographic maps of Paratethys-10 maps Late Eocene to Pliocene, 46 pp.

- Posada, D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25:1253–1256.
- Posada, D. and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50:580–601.
- Posada, D. and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53:793–808.
- Poyato-Ariza, F. J., M. A. López-Horgue, M. A. Garcia-Garmilla, and F. Garcia-Garmilla. 2000. A new Early Cretaceous clupeomorph fish from the Arratia Valley, Basque Country, Spain. *Cretaceous Research* 21:571–585.
- Pravotorov, B. I. 2000. On population structure of the Black Sea – Azov Sea fish *Alosa caspia tanaica* (Grimm), (Clupeidae). *Hydrobiological Journal* 36:101–105.
- Puttick, M. N., J. E. O'Reilly, D. Oakley, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, and D. Pisani. 2017. Parsimony and maximum-likelihood phylogenetic analyses of morphology do not generally integrate uncertainty in inferring evolutionary history: a response to Brown et al. *Proceedings of the Royal Society B: Biological Sciences* 284:20171636.
- Puttick, M. N., J. E. O'Reilly, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, and P. C. J. Donoghue. 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B: Biological Sciences* 284:p.20162290.

- Puttick, M. N., J. E. O'Reilly, D. Pisani, and P. C. Donoghue. 2019. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology* 62:1–17.
- Pyron, R. A. 2017. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Systematic Biology* 66:38–56.
- Pyron, R. A., F. T. Burbrink, and J. J. Wiens. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology* 13:10.1186/1471-2148-1113-1193.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <https://www.R-project.org/>
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rafinesque, C. S. 1820. Fishes of the Ohio River. *Western Review and Miscellaneous Magazine* 2:169–177.
- Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology* 67:901–904.
- Rambaut, A., Suchard, M. A., Xie, D. and Drummond, A. J. 2014. Tracer v1.6. Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Ramler, D., A. Palandačić, G. B. Delmastro, J. Wanzenböck, and H. Ahnelt. 2017. Morphological divergence of lake and stream Phoxinus of Northern Italy and the Danube basin based on geometric morphometric analysis. *Ecology and evolution* 7:572–584.

- Rannala, B., J. P. Huelsenbeck, Z. Yang, and R. Nielsen. 1998. Taxon sampling and the accuracy of large phylogenies. *Systematic Biology* 47:702–710.
- Regan, C. T. 1917a. A revision of the clupeid fishes of the genus *Pellonula* and of related genera in the rivers of Africa. *Annals and Magazine of Natural History* 19:198–207.
- Regan, C. T. 1917b. A revision of the clupeoid fishes of the genera *Pomolobus*, *Brevoortia* and *Dorosoma* and their allies. *Annals and Magazine of Natural History* 19:297–316.
- Regan, C. T. 1929. Fishes. *Encyclopaedia Britannica* 9:305–329.
- Reist, J., E. Gyselman, J. Babaluk, J. Johnson, and R. Wissink. 1995. Evidence for two morphotypes of Arctic char (*Salvelinus alpinus* (L.)) from Lake Hazen, Ellesmere Island, Northwest Territories, Canada. *Nordic Journal of Freshwater Research* 71:396–410.
- Revell, L. J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223.
- Richardson, J. 1846a. Ichthyology of the voyage of H. M. S. Erebus & Terror, under the command of Captain Sir James Clark Ross, R. N., F. R. S.; pp. 1–139 in: J. Richardson & J. E. Gray (eds.): *The zoology of the voyage of H. M. S. Erebus & Terror, under the command of Captain Sir J. C. Ross, R. N., F. R. S., during the years 1839 to 1843*. E. W. Janson, London, 139 pp.
- Richardson, J. 1846b. Report on the ichthyology of the seas of China and Japan. Report of the British Association for the Advancement of Science 15th meeting 1845:187–320.
- Roberts, T. R. 1981. Sundasalangidae, a new family of minute freshwater salmoniform fishes from southeast Asia. *Proceedings of the California Academy of Sciences* 42:295–302.
- Robertson, D. R., O. Domínguez-Domínguez, Y. M. L. Aroyo, R. M. Mendoza, and N. Simões. 2019. Reef-associated fishes from the offshore reefs of western Campeche Bank, Mexico,

- with a discussion of mangroves and seagrass beds as nursery habitats. *ZooKeys* 843:71–115.
- Robinson, B. W. and K. J. Parsons. 2002. Changing times, spaces, and faces: tests and implications of adaptive morphological plasticity in the fishes of northern postglacial lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 59:1819–1833.
- Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53:131–147.
- Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53:131–147.
- Roesch, C., B. Lundsgaard-Hansen, P. Vonlanthen, A. Taverna, and O. Seehausen. 2013. Experimental evidence for trait utility of gill raker number in adaptive radiation of a north temperate fish. *Journal of Evolutionary Biology* 26:1578–1587.
- Ronquist F., N. Lartillot, and M. J. Phillips. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B* 371:20150136.
- Ronquist, F., J. P. Huelsenbeck, and M. Teslenko. 2019. MrBayes version 3.2 manual: tutorials and model summaries. Available online:
https://github.com/NBISweden/MrBayes/blob/develop/doc/manual/Manual_MrBayes_v3.2.pdf
- 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* 61:539–542.

- Ronquist, F., N. Lartillot, and M. J. Phillips. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society. B* 371:20150136.
- Rüppell, W. P. E. S. 1837. Neue Wirbelthiere zu der Fauna von Abyssinien gehörig. Fische des Rothen Meeres. Siegmund Schmerber, Frankfurt am Main, 148 pp.
- Saint Marc, P. 1974. Étude stratigraphique et micropaléontologique de l'Albien, du Cenomanien et du Turonien du Liban. *Notes et Mémoires sur le Moyen Orient, Paris* 13:8–42.
- Sanderson, M. J. and M. J. Donoghue. 1989. Patterns of Variation in Levels of Homoplasy. *Evolution* 43:1781–1795.
- Sankoff, D. 1975. Minimal mutation trees of sequences. *SIAM Journal of Applied Mathematics* 28:35–42.
- Sato, Y. 1994. Cladistic-phenetic relationships and the evolutionary classification of the suborder Clupeoidei (Pisces, Teleostei). Unpublished PhD thesis. Tokyo Metropolitan University, Tokyo, 186 pp.
- Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of entomology* 55:421–438.
- Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27:592–593.
- Schrago, C. G., B. O. Aguiar, and B. Mello. 2018. Comparative evaluation of maximum parsimony and Bayesian phylogenetic reconstruction using empirical morphological data. *Journal of Evolutionary Biology* 31:1477–1484.

- Schram, F. R., C. H. J. Hof, and F.A. Steeman. 1999. Thylacocephala (Arthropoda: Crustacea?) from the Cretaceous of Lebanon and implications for thylacocephalan systematics. *Palaeontology* 42:769–797.
- Schwartz, R. S. and R. L. Mueller. 2010. Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evolutionary Biology* 10:10.1186/1471-2148-10-5.
- Scotland, R. W., R. G. Olmstead, and J. R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology* 52:539–548.
- Sidlauskas, B. L., J. H. Mol, and R. P. Vari. 2011. Dealing with allometry in linear and geometric morphometrics: a taxonomic case study in the *Leporinus cylindriformis* group (Characiformes: Anostomidae) with description of a new species from Suriname. *Zoological Journal of the Linnean Society* 162:103–130.
- Signeux, J. 1951. Notes paléoichthyologiques. *Diplomystus dubertreti*, une nouvelle espèce du Sénonien du Liban. *Bulletin du Muséum National d'Histoire Naturelle* 23:692–693.
- Signeux, J. 1964. *Gasteroclupea branisai*, clupéidé nouveau du Crétacé supérieur de Bolivie. *Bulletin du Muséum National d'Histoire Naturelle* 1964:291–297.
- Silva Santos, R. 1994. Ictiofauna da Formação Codó, Cretáceo Inferior, com a descrição de um novo taxon – *Codoichthys carnavalii* (Pisces-Teleostei). *Anais da Academia brasileira de Ciências* 66:131–144.
- Silva Santos, R., and V. L. Silva Corrêa. 1985. Contribuição ao conhecimento da paléoictiofauna do Cretáceo no Brasil; pp. 169–174 in A. de Campos, C. S. Ferreira, I. M. Brito, and C. F. Viana (eds.), *Coletânea de Trabalhos Paleontológicos, Série Geologia*

- 27(2). Ministério das Minas e Energia-Departamento Nacional de Produção Mineral, Rio de Janeiro, Brazil, 656 pp.
- Simões, T. R., M. W. Caldwell, A. Palci, and R. L. Nydam. 2018. Giant taxon-character matrices II: a response to Laing et al. (2017). *Cladistics* 34:702–707.
- Simões, T. R., M. W. Caldwell, M. Tałanda, M. Bernardi, A. Palci, O. Vernygora, F. Bernardini, L. Mancini, and R. L. Nydam. 2018. The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature* 557:706–709.
- Smith, W. D., J. J. Bizzarro, V. P. Richards, J. Nielsen, F. Márquez-Flarías, and M. S. Shivji. 2009. Morphometric convergence and molecular divergence: the taxonomic status and evolutionary history of *Gymnura crebripunctata* and *Gymnura marmorata* in the eastern Pacific Ocean. *Journal of Fish Biology* 75:761–783.
- Soubrier, J., M. Steel, M. S. Y. Lee, C. Der Sarkissian, S. Guindon, S. Y. W. Ho, and A. Cooper. 2012. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution*. 29:3345–3358.
- Spix, J. B. von and L. Agassiz. 1829. *Selecta genera et species piscium quos in itinere per Brasiliam annis MDCCCXVII-MDCCCXX jussu et auspiciis Maximiliani Josephi I collegit et pingendos curavit Dr J. B. de Spix*. Monachii, 138 pp.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

- Stamatakis, A., F. Blagojevic, D. S. Nikolopoulos, and C. D. Antonopoulos. 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM Cell. *J VLSI Signal Process* 48:271–286.
- Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.
- Steindachner, F. 1876. Ichthyologische Beiträge (V). *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Classe* 74:49-240.
- Stern, N., B. Rinkevich, and M. Goren. 2016. Integrative approach revises the frequently misidentified species of *Sardinella* (Clupeidae) of the Indo-West Pacific Ocean. *Journal of Fish Biology* 89:2282–2305.
- Stiassny, M. L., L. R. Parenti, and G. D. Johnson. 1996. *Interrelationships of Fishes*. Academic Press, London, 496 pp.
- Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4:vey016.
- Sukumaran, J. and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26:1569–1571.
- Sullivan, J. and P. Joyce. 2005. Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 36:445–466.
- Sun, A.-L. 1956. *Paraclupea* – A genus of double-armored herrings from Chekiang. *Acta Palaeontologica Sinica* 4:413–418.
- Svetovidov, A. N. 1945. On *Clupeonella delicatula* (Nordmann) of the Caspian and that of the Black and Azov seas. *Comptes Rendus de l'Academie des Sciences de l'URSS* 46:207–208.

- Svetovidov, A. N. 1952. Fauna of the USSR, Fishes. Herrings (Clupeidae). Academy of Sciences of USSR publishing, Moscow-Leningrad, 331 pp.
- Svetovidov, A. N. 1964. Systematics of North American anadromous Clupeoid fishes of the genera *Alosa*, *Caspialosa* and *Pomolobus*. *Copeia* 196:118–130.
- Swain, D. P. and C. J. Foote. 1999. Stocks and chameleons: the use of phenotypic variation in stock identification. *Fisheries Research* 43:113–128.
- Swainson, W. 1839. On the natural history and classification of fishes, amphibians, and reptiles, or monocardian animals. Spottiswoode & Co., London, 452 pp.
- Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0a. Sinauer, Sunderland, Massachusetts. Available at <https://paup.phylosolutions.com>.
- Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference; pp. 407–514 in D. M. Hillis, C. Moritz, B. K. Mable (eds.), *Molecular systematics*, 2nd edition. Sinauer Associates, Sunderland, Massachuset, 655 pp.
- Taverne, L. 1997. Les clupéomorphes (pisces, teleostei) du cénomanien (crétace) de kipala (Kwango, Zaire): Ostéologie et phylogénie. *Belgian journal of zoology* 127:75–98.
- Taverne, L. 2004. Les poissons cré'tace's de Nardo`. 18°. *Pugliaclupea nolardi* gen. et sp. nov. (Teleostei, Clupeiformes, Clupeidae). *Bollettino del Museo Civico di Storia Naturale di Verona* 28:17–28.
- Team, R. C. 2018. R: A Language and Environment for Statistical Computing.
- Temminck, C. J. and H. Schlegel. 1846. Pisces; pp. 173–269 in: Siebold, P. F. de (ed.), *Fauna Japonica, sive descriptio animalium, quae in itinere per Japoniam suscepto annis 1823-*

- 1830 collegit, notis, observationibus et adumbrationibus illustravit Ph. Fr. de Siebold.
Lugduni Batavorum, 323 pp.
- Teugels, G. G. 2003. Denticipitidae; pp. 122–124 in D. Paugy, C. Lévêque, and G.G Teugels (eds.), The fresh and brackish water fishes of West Africa Volume 1. Coll. faune et flore tropicales 40. Institut de recherche de développement, Paris, France, Muséum national d'histoire naturelle, Paris, France and Musée royal de l'Afrique Central, Tervuren, Belgium, 457 pp.
- Than-Marchese, B. A., J. Alvarado-Ortega, W. A. Matamoros, and E. Velázquez-Velázquez. 2020. *Scombroclupea javieri* sp. nov., an enigmatic Cenomanian clupeomorph fish (Teleostei, Clupeomorpha) from the marine deposits of the Cintalapa Formation, Ocozocoautla, Chiapas, southeastern Mexico. *Cretaceous Research* 112:10.1016/j.cretres.2020.104448.
- Thomas Jr, R. C., D. A. Willette, K. E. Carpenter, and M. D. Santos. 2014. Hidden diversity in sardines: genetic and morphological evidence for cryptic species in the goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849). *PLoS ONE* 9:p.e84719.
- Trifinopoulos, J., L. T. Nguyen, A. Von Haeseler, and B. Q. Minh. 2016. IQ-TREE web server: fast and accurate phylogenetic trees under maximum likelihood. *Nucleic Acids Research* 44:232–235.
- Turan, C. and N. Basusta. 2001. Comparison of morphometric characters of twaite shad (*Alosa fallax nilotica*, Geoffroy Saint-Hilaire, 1808) among three areas in Turkish seas. *Bulletin Français de la Pêche et de la Pisciculture* 362/363:1027–1035.

- Turan, C., D. Ergüden, M. Gürlek, C. Çevik, and F. Turan. 2015. Molecular systematic analysis of shad Species (*Alosa* spp.) from Turkish marine waters using mtDNA Genes. *Turkish Journal of Fisheries and Aquatic Sciences* 15:149–155.
- Valenciennes, A. 1847. In Cuvier, G. and A. Valenciennes. *Histoire naturelle des poissons*. Tome vingtième. Livre vingt et unième. De la famille des Clupéoïdes. *Histoire Naturelle Des Poissons*, 472 pp.
- Valenciennes, A. 1848. In Cuvier, G. and A. Valenciennes. *Histoire naturelle des poissons*. Tome vingt et unième. Suite du livre vingt et unième et des Clupéoïdes. Livre vingt-deuxième. De la famille des Salmonoïdes. 21:1–536.
- Van Hasselt, J. C. 1823. Uittreksel uit een'brief van Dr. JC van Hasselt, aan den Heer CJ Temminck. *Algemeene Konsten Letter-Bode*, II Deel 35:130–133.
- Velotta, J. P., S. D. McCormick, and E. T. Schultz. 2015. Trade-offs in osmoregulation and parallel shifts in molecular function follow ecological transitions to freshwater in the Alewife. *Evolution* 69:2676–2688.
- Venables, W. N. and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York, 498 pp.
- Vernygora, O. V. 2015. Two new species and a revised phylogeny of the Ellimmichthyiformes (Teleostei: Clupeomorpha). Unpublished MSc thesis. University of Alberta, Edmonton, 181 pp.
- Vernygora, O. and A. M. Murray. 2016. A new species of *Armigatus* (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous of Morocco, and its phylogenetic relationships. *Journal of Vertebrate Paleontology* 36:e1031342.

- Vernygora, O., A. M. Murray, and M. V. Wilson. 2016. A primitive clupeomorph from the Albian Loon River Formation (Northwest Territories, Canada). *Canadian Journal of Earth Sciences* 53:331–342.
- Vernygora, O.V., C. S. Davis, A. M. Murray, and F. A. H. Sperling. 2018. Delimitation of *Alosa* species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches. *Journal of Fish Biology* 93:1216–1228.
- Walbaum, J. J. 1792. Petri Artedi sueci genera piscium. In quibus systema totum ichthyologiae proponitur cum classibus, ordinibus, generum characteribus, specierum differentiis, observationibus plurimis. Redactis speciebus 242 ad genera 52. *Ichthyologiae pars III. Ant. Ferdin. Rose, Grypeswaldiae.* 723 pp.
- Walker, J. A. 1997. Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (Gasterosteidae) body shape. *Biological Journal of the Linnean Society* 61:3–50.
- Whidden, C. and F. A. Matsen IV. 2018. Efficiently inferring pairwise subtree prune-and-regraft adjacencies between phylogenetic trees; pp. 77–91 in 2018 Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO). Society for Industrial and Applied Mathematics, 177 pp.
- Whitehead, P. J. P. 1963. A contribution to the classification of clupeoid fishes. *Annals and Magazine of Natural History* 5:737–750.
- Whitehead, P. J. P. 1972. A synopsis of the clupeoid fishes of India. *Journal of Marine Biology Association of India* 14:160–256.

- Whitehead, P. J. P. 1985a. FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part I. Chirocentridae, Clupeidae and Pristigasteridae. FAO Fisheries Synopsis 125:1–303.
- Whitehead, P. J. P. 1985b. King herring: his place amongst the clupeoids. Canadian Journal of Fisheries and Aquatic Sciences 42:3–20.
- Whitehead, P. J. P., G. J. Nelson, and T. Wongratana. 1988. FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei). Part 2. Engraulididae. FAO Fisheries Synopsis 125:305–579.
- Whitley, G. P. 1940. Illustrations of some Australian fishes. Australian Zoologist 9:397–428.
- Whitley, G. P. 1951. New fish names and records. Proceedings of the Royal Zoological Society of New South Wales v. for 1949-50:61–68.
- Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. Journal of Biomedical Informatics 39:34–42.
- Wilson, A. 1811. *Clupea* Heading. In Reese's New Cyclopedia. American edition. v. 9. No pagination.
- Wilson, A. B., G. G. Teugels, and A. Meyer. 2008. Marine incursion: The freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. PLoS ONE 3:e1979.
- Wongratana, T. 1983. Diagnoses of 24 new species and proposal of a new name for a species of Indo-Pacific clupeoid fishes. Japanese Journal of Ichthyology 29:385–407.
- Woodward, A. S. 1895. Catalogue of the Fossil Fishes in the British Museum (Natural History), Cromwell Road, S. W. Part III. Containing the Actinopterygian Teleostomi of the Orders

- Chondrostei (concluded), Protospondyli, Aetheospondyli, and Isospondyli (in part). British Museum (Natural History), London, 706 pp.
- Wright A. M., and D. M. Hillis. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS ONE 9:e109210.
- Wright, A. M., G. T. Lloyd, and D. M. Hillis. 2015. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Systematic Biology 65:602–611.
- Wund, M. A., S. Valena, S. Wood, and J. Baker. 2012. Ancestral plasticity and allometry in threespine stickleback fish reveal phenotypes associated with derived, freshwater ecotypes. Biological Journal of the Linnean Society 105:573–583.
- Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Systematic Biology 60:150–160.
- Yabumoto, Y. and T. Uyeno. 1982. Osteology of clupeiform fish, genus *Hyperlophus* (II). Bulletin Kitakyushu Museum of Natural History 4:77–102.
- Yang, Z. 1995. A space-time process model for the evolution of DNA sequences. Genetics 139:993–1005.
- Yang, Z. I. H. E. N. G., N. Goldman, and A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Molecular Biology and Evolution 11:316–324.
- Yang, Z., N. Goldman, and A. Friday. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Systematic Biology 44:384–399.

- Yeates D. K., A. Seago, L. Nelson, S. L. Cameron, L. Joseph, and J. W. H. Trueman. 2011. Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* 36:209–217.
- Zaragüeta-Bagils, R. 2004. Basal clupeomorphs and ellimmichthyiform phylogeny; pp. 391–404 in G. Arratia, G. Tintori, and A. Tintori (eds.), *Mesozoic Fishes 3 – Systematics, Paleoenvironment and Biodiversity*. Verlag Dr. Friederich Pfeil, Munich, Germany, 703 pp.
- Zhang, M. M., J. J. Zhou, and D. R. Qing. 1985. Tertiary fish fauna from coastal region of Bohai Sea. *Academia Sinica Institute of Vertebrate Paleontology and Paleoanthropology Memoirs* 17:1–60.
- Zhang C., T. Stadler, S. Klopstein, T. A. Heath, and F. Ronquist. 2016. Total-Evidence dating under the fossilized birth–death process. *Systematic Biology* 65:228–249.
- Zhou, X., X. X. Shen, C. T. Hittinger, and A. Rokas. 2018. Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic datasets. *Molecular Biology and Evolution* 35:486–503.
- Zwickl, D. J. and D. M. Hillis. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51:588–598.