Paleontological Data Reveal Unexpected Biogeographic Histories of Extant Organisms: Bonytongue Fishes (Teleostei: Osteoglossomorpha) as a Case Study

by

Alessio Capobianco

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Earth and Environmental Sciences)
in The University of Michigan
2021

Doctoral Committee:

Associate Professor Matt Friedman, Chair
Associate Professor Hernán López-Fernández
Associate Professor Daniel Rabosky
Associate Professor Stephen A. Smith
Professor Jeffrey Wilson Mantilla

Alessio Capobianco

acapo@umich.edu

ORCID iD: 0000-0002-6096-3875

DEDICATION

This dissertation is dedicated with love and gratitude to Beata and Antonio.

ACKNOWLEDGEMENTS

This dissertation was generously funded by the Department of Earth and Environmental Sciences Scott Turner Student Research Grant Award, by the Rackham Graduate School Predoctoral Fellowship Award, and by the Society of Systematic Biologists Graduate Student Research Award. The research in chapter 2, 3, 4, and 5 was performed in collaboration with Matt Friedman. Additionally, the research in chapter 3 included the collaboration of undergraduate research assistant Ethan Foreman. Part of the μCT datasets analyzed throughout this dissertation were processed by research assistant Danielle Goodvin and undergraduate research assistant Kyle Kramer. Most of this work would not have been possible without the natural history collections that provided access to specimens of invaluable scientific importance. I thank Douglas Nelson and Randy Singer at the University of Michigan Museum of Zoology, William F. Simpson at the Field Museum of Natural History, Chicago, Bo Schultz and René L. Sylvestersen at the Fur Museum, Bent E. K. Lindow at the Natural History Museum of Denmark, Emma Bernard at the Natural History Museum, London, Mariagabriella Fornasiero at the Istituto Geologico dell'Università di Padova, Anna Vaccari and Roberta Salmaso at the Museo Civico di Storia Naturale, Verona, Gaël Clement at the Muséum National d'Histoire Naturelle, Paris and Florias Mees at the Musée Royal de l'Afrique Centrale. I really want to thank William Sanders, chief preparator at the University of Michigan Museum of Paleontology, for amazing preparation on some of the specimens I worked on during my dissertation, and for sharing with me some of his Star Wars memorabilia. I thank Carol Abraczinskas for her help and invaluable perspective on figure drafting and composition. I also thank my committee members for their constructive feedback throughout the length of my doctoral studies. I thank all UMMP faculty and the whole UMMP community for incredibly insightful discussions during my 5 years here. We all know that the real doctorate was the friends we made along the way: thus, I want to wholeheartedly thank all friends at the Earth and EEB departments, all the GLSP friends and all the friends at Paleohouse that accompanied my Michigan adventure for 5 long (or short?) years. A special

thanks to Kelly Matsunaga, Kierstin Rosenbach, James Saulsbury, James Andrews and Rodrigo Figueroa (in strictly chronological order of when I met them) for stealing different parts of my heart and being such unique companions through this journey. Another special thanks to some amazing old friends that made the distance from home more bearable: Alexia, Marica, Luigi, Niccolò, this is for you. Mom, dad: you have always been there for me every single day of these 5 years so far away from you. This dissertation is dedicated to you. Finally, I am extremely grateful to my advisor, Matt Friedman, for his extreme patience with me, for his amazing scientific perspective, for his constant help and support, for guiding me when I needed guidance and for leaving me space when I needed space. Seriously, Matt: thank you.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF APPENDICES

ABSTRACT

Paleontological data are invaluable for reconstructing the biogeographic history of living organisms. Nonetheless, information from present-day species (neontological data) dominates biogeographic studies of extant clades, due to either incompleteness of the fossil record or challenges in integrating it into evolutionary inference. In this dissertation, I explore the paleontological record of the freshwater fish clade Osteoglossomorpha (bonytongues) to derive a deep-time perspective on the biogeographic history of this ancient and iconic group of fishes. The complex geographic distribution of extant bonytongues, coupled with their abundant fossil record when compared to other tropical freshwater fishes, makes this group an ideal target for biogeographic investigation through a paleontological lens.

I first consider the temporal and geographic distribution of the fossil record of seven extant freshwater fish groups – including bonytongues – to derive confidence intervals on their times of origin and test the plausibility of vicariant scenarios in which continental break-ups shaped their modern distributions. I find that, even when fish groups are old enough to have been affected by continental fragmentation during the Mesozoic, successive dispersals and regional extinction tend to erase or confound vicariant patterns and shape the geographic distributions that we see today. The middle portion of my dissertation involves the description of two bonytongue fossil specimens from early Cenozoic marine deposits in Greenland and Morocco. The Greenland specimen extends the geographic range of the group to the Arctic and represents one of their earliest records in marine deposits, few million years after the Cretaceous–Paleogene mass extinction. The Moroccan specimen represents a new genus with cranial adaptations related to feeding ecology previously unknown in these fishes. I show how bonytongues reached a surprising ecomorphological diversity in marine settings during the early Cenozoic, and identify key anatomical features providing evidence for the phylogenetic affinities of fossil marine bonytongues with respect to modern species.

Finally, I combine morphological, molecular, geographic, and environmental data in an integration of paleontological and neontological evidence to reconstruct the biogeographic

history of bonytongue fishes under phylogenetic models of biogeographic evolution. I find strong support for a marine origin of osteoglossid bonytongues and for long-distance dispersals—followed by multiple marine-to-freshwater transitions—as the primary cause for the present-day widespread distribution of these fishes in tropical freshwaters. Moreover, I show how fossil data can completely overthrow biogeographic patterns that are apparent from the examination of extant distributions alone, highlighting the perils of ignoring paleontological evidence when inferring ancestral conditions for living organisms. This dissertation provides new insights into the evolutionary history of one of the major lineages of freshwater fishes, and establishes Osteoglossomorpha as a promising model system within vertebrates to explore the impact of paleontological data in evolutionary inference.

CHAPTER 1

Introduction

"It is true, that the animals found fossil in a country are very generally allied to those which still inhabit it; but this is by no means universally the case."

Alfred Russell Wallace – The Geographical Distribution of Animals (1876)

The geographic distribution of organisms on Earth is the result of the complex interplay of evolutionary processes and long-term changes in abiotic systems, including tectonic evolution and climate change. It is not a coincidence that biogeography—the study of the geographic distribution of species and of its causes—features prominently in the seminal works of the two founders of evolutionary biology, Charles Darwin and Alfred Russell Wallace. While Darwin and Wallace focused mostly on the geographic distribution of extant taxa as evidence of evolution, they both integrated those observations with information coming from the fossil record. In particular, Darwin underlined the occurrence of fossil species in the same geographic areas where most similar extant species live today (e.g., fossil sloths and armadillos in South America) as a key proof of common ancestry of these organisms (Darwin, 1859). Remarkably, Wallace emphasized instead that fossil relatives of living organisms can be found in areas where they are *not present* today, and that these fossils are particularly useful for inferring areas of origin of extant taxa and past dispersal events that might not be apparent by examining current distributions alone (Wallace, 1876). Thus, the impact of fossil data on biogeographic inference has been discussed since the earliest days of evolutionary biology theory. However, the relative importance of fossil compared to extant data towards unveiling biogeographic patterns and processes has continuously changed over the history of biogeographic research.

Fossils and historical biogeography

The rise of cladistic methods applied to biogeography in the 1970s and 1980s sparked a heated debate over the utility of fossil data towards biogeographic inference, with several biogeographers arguing that fossils cannot contradict—but only confirm or complete—patterns obtained from living organisms alone and that the inherent incompleteness of the fossil record greatly reduces the weight that should be given to fossils in the historical biogeography of extant clades (Humphries & Parenti, 1986; Brundin, 1988). Contrary to this point of view, Grande (1985) stressed different ways in which fossils can actually influence the reconstruction of historical biogeography of living organisms. These include establishing a minimum age of origin for a taxon, which can corroborate or disprove biogeographic hypotheses; expanding the known geographic range of a taxon; and displaying biogeographic patterns that were subsequently eroded by regional extinction.

The more recent development of biogeographic models based on cladogenetic and anagenetic processes (Ree & Smith, 2008; Matzke, 2014), coupled with the explosion of molecular phylogenetics and the relative ease of obtaining phylogenetic trees for extant organisms, has led to a proliferation of model-based approaches to infer biogeographic histories of a wide range of groups. In most of these studies, fossil data are only used as minimum age calibration points to obtain time-scaled phylogenies (or timetrees) that can be analyzed through biogeographic models (e.g., Buerki *et al*., 2011; Sanmartín, 2012). This is only one of the potential contributions of the fossil record towards biogeographic inference outlined by Grande (1985). However, it is also the easiest to implement, as it requires only age and taxonomic attribution of a fossil.

In the last few years, additional efforts have gone towards implementing more kinds of information that the fossil record provides in biogeographic inference. For example, in phylogeny-based approaches, fossils can be used to constrain the geographic range of some internal nodes of the phylogeny (if there is an *a priori* hypothesis that those fossils are ancestral to an extant clade), or can be directly used as phylogenetic tips in a biogeographic analysis to provide both temporal and geographic data at the same time (e.g., Meseguer *et al*., 2015; Tavares *et al*., 2018; Oliveros *et al.,* 2020). Other approaches that do not require a phylogenetic input can be used to estimate dispersal and extinction rates through time while accounting for the geographic distribution of fossils and for the incompleteness of the fossil record (Silvestro *et al*., 2016). Nonetheless, integration of the full range of fossil data available for a certain group of

living organisms remains somewhat limited in the biogeographic literature, as it requires a combination of extensive morphological assessment of both extinct and extant taxa, paleogeographic and paleoenvironmental reconstructions, and models that allow for the use of fossil data (or better yet explicitly consider the temporal and geographic biases of the fossil record).

Freshwater fishes: an iconic system in biogeography

Freshwater fishes have featured prominently in biogeographic research over the years, thanks to their diversity, ecological and economical relevance, somewhat limited dispersal ability across different watersheds, and striking biogeographic patterns at a continental scale. It is not a coincidence that a significant portion of the seminal works in historical biogeography has been written by ichthyologists and paleoichthyologists (e.g., Rosen, 1978; 1988; Parenti, 1981; 1991; Patterson, 1981; Grande 1985).

In particular, freshwater fish distribution has been often used as evidence of continental vicariance, by which the progressive break-up of the supercontinent Pangea during the Mesozoic caused the disjunct geographic distribution that we can observe in many groups today (Parenti, 1981; Murphy & Collier, 1997; Chakrabarty, 2004). The most famous examples involve pairs of sister groups that can be found on both sides of the southern Atlantic—South America and Africa. These include osteoglossid bonytongues, cichlids, lungfishes, killifishes, and characiforms. However, the geographic and temporal distributions of the fossil record of these fishes often yield conflicting results (Lundberg, 1993; Friedman *et al*., 2013), and suggest that a continental vicariance scenario might not explain the biogeographic history of these groups. Alternative scenarios involve more recent dispersals, either *via* land (through transient land bridges or after continental collision) or sea, or a complex combination of vicariance, dispersal and regional extinction. A major issue in evaluating these different hypotheses is that the fossil record of most freshwater fishes that today inhabit tropical environments is extremely spotty, consisting mostly of fragmentary remains of difficult or dubious taxonomic attribution. Bonytongues represent a remarkable exception to this general rule.

Bonytongue fishes: a relic of Gondwana?

Bonytongue fishes (Osteoglossomorpha) are one of the earliest diverging lineages of teleost fishes, with a long evolutionary history that extends to the mid-Mesozoic (Wilson & Murray, 2008). The low species diversity of modern bonytongues contrasts with their remarkable diversity of forms, ranging from the unassuming mooneyes to the gigantic arapaima to the electrical elephantfishes. Despite this diversity, all extant species are ecologically restricted to freshwater environments in mostly tropical areas (South America, Africa, India and Southeast Asia, Australia), with the exception of two species of temperate-adapted mooneyes, which inhabit North America. In contrast to most groups of tropical freshwater fishes, bonytongues are known from numerous fossil species, many of them represented by relatively well-preserved, articulated specimens yielding abundant trait data for phylogenetic inference. In fact, extinct bonytongue genera surpass extant ones in number (Wilson & Murray, 2008; Hilton & Lavoué, 2018). Moreover, the geographic distribution of bonytongue fossils is even more widespread than that of living species (Fig. 1.1). Perhaps the most surprising feature of the paleontological record of bonytongues is the presence of several fossils (including well-preserved, articulated skeletons) in marine deposits worldwide.

The old age, widespread distribution in freshwater systems worldwide, and 'ancient' external appearance of bonytongues made them a prominent textbook example of 'Gondwanan' group that originated when all southern landmasses were joined in the supercontinent Gondwana and were carried by drifting continents to their present-day geographic distribution (Bond, 1996; Paxton & Eschmeyer, 1998; Moyle & Cech, 2000). However, the complex geographic patterns displayed by fossil bonytongues, their invasion of marine environments at a certain point of their evolutionary history, and inferred divergence times too young to be fully compatible with Gondwanan vicariance have led paleontologists and biogeographers to challenge the 'Gondwanan narrative' and consider more complex scenarios (Patterson, 1975; Lundberg, 1993; Bonde, 2008; Wilson & Murray, 2008; Lavoué, 2016; Hilton & Lavoué, 2018). Nonetheless, the biogeographic history of bonytongues remains an open question, mostly because of the uncertain phylogenetic relationships of fossil forms in respect to extant species.

Thus, for the aforementioned reasons, bonytongue fishes represent an excellent study system for the investigation of biogeographic patterns through a paleontological lens, and for an assessment

of the impact that fossil data have on the reconstruction of the biogeographic history of extant organisms.

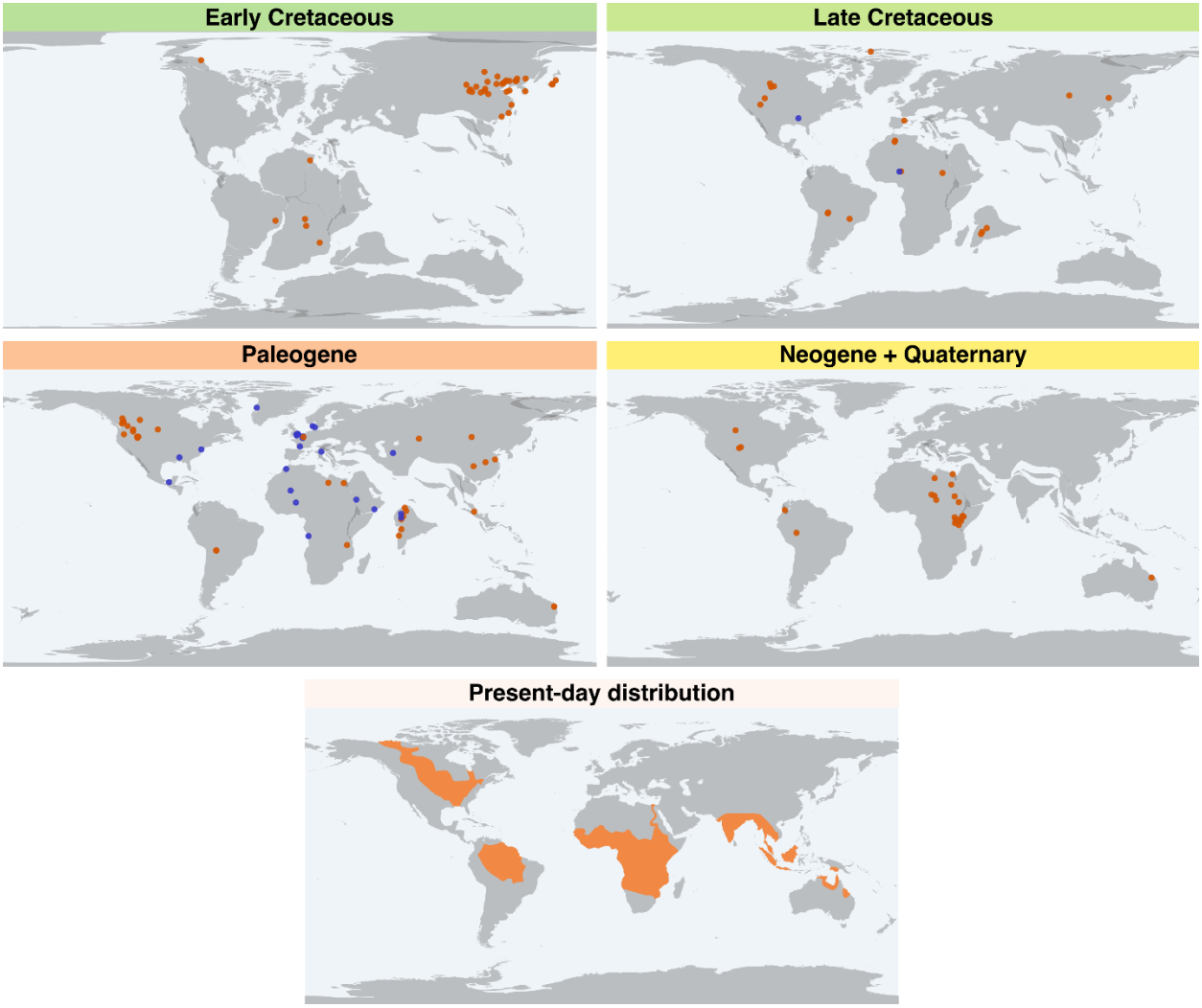

Fig. 1.1 Geographic distribution of fossil and living bonytongues (Osteoglossomorpha). Orange dots indicate freshwater fossil deposits, blue dots indicate marine fossil deposits. Notice the relative abundance and geographic spread of bonytongues in marine deposits during the Paleogene. Fossil occurrences from Chapter 2 of this Dissertation (see Table B.2). Present-day geographic range from Berra (2007). Paleogeographic maps from Müller *et al*. (2016).

Paleontological data reveal unexpected biogeographic histories of extant organisms

The overarching goal of this dissertation is to explore the paleontological record of bonytongues (Osteoglossomorpha) to derive a deep-time perspective on the biogeographic history of this ancient and iconic group of fishes. By doing so, I highlight how different approaches to fossil

data and integration with neontological data can yield unique perspectives about the evolution of extant organisms.

In Chapter 2, I review the fossil record of freshwater fishes with an intercontinental distribution in the southern hemisphere, and use its stratigraphic distribution to estimate confidence intervals on the time of origin of seven extant fish groups (including bonytongues). Origin time estimates and past and present geographic distributions provide a critical test of the vicariance scenario whereby the progressive breakup of the supercontinent Pangaea shaped the disjunct distributions that we observe today. This chapter showcases how both quantitative and qualitative evaluations of the fossil record of extant organisms improve our understanding of the biogeographic history of extant organisms.

In Chapter 3, I describe a fossil osteoglossid from early Paleocene (Danian) marine deposits of Greenland, representing the northernmost known occurrence for these animals and one of the oldest in marine environments. Moreover, I revise the fossil record of marine bonytongues, highlighting how these fishes apparently diversify and disperse globally on the wake of the Cretaceous-Paleogene mass extinction.

Chapter 4 describes a new genus and species of marine bonytongue from the early Eocene (Ypresian) of Morocco, characterized by an elongated snout that suggests higher eco-morphological diversity than previously assumed for this group of fishes. Thanks to new anatomical observations with the aid of microcomputed tomography (μCT), I evaluate the phylogenetic affinities of this new taxon and other marine bonytongues, concluding that these forms are deeply nested within extant bonytongues and likely represent close relatives of the giant arapaimas found today in South American fresh waters.

Finally, in Chapter 5, I reconstruct the biogeographic history of bonytongues and explore how the inclusion of fossil taxa in model-based biogeographic analyses impacts estimates of ancestral ranges. I discover that extant osteoglossid bonytongues likely derive from marine ancestors that dispersed circumglobally and reentered freshwater systems multiple times independently. This is the first known case in which the last common ancestor of a fish clade was marine whereas all its extant members and their closest living relatives are ecologically restricted to freshwater settings. I demonstrate how fossil data radically changes model-based biogeographic inferences, and I

show how trait-dependent dispersal models represent a promising approach to including the biology of organisms into biogeographic reconstructions.

In the concluding chapter, I summarize the main findings of this dissertation and outline future research avenues in biogeography and evolutionary biology that could open up through the integration of descriptive and analytical approaches applied to a combination of paleontological and neontological data.

REFERENCES

Berra, T. M. (2007). *Freshwater Fish Distribution*. The University of Chicago Press, Chicago.

Bond, C. E. (1996). *Biology of Fishes*, 2nd edition. Saunders College Publishing.

Bonde, N. (2008). Osteoglossomorphs of the marine Lower Eocene of Denmark–with remarks on other Eocene taxa and their importance for palaeobiogeography. *Geological Society, London, Special Publications*, 295, 253-310.

Brundin, L. Z. (1988). Phylogenetic biogeography. In: Myers, A. A. & Giller, P.S. (eds.) *Analytical biogeography*, pp. 343-369. New York: Chapman & Hall.

Buerki, S., Forest, F., Alvarez, N., Nylander, J. A., Arrigo, N., & Sanmartín, I. (2011). An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography*, 38, 531-550.

Chakrabarty, P. (2004). Cichlid biogeography: comment and review. *Fish and fisheries*, 5, 97-119.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Friedman, M., Keck, B. P., Dornburg, A., Eytan, R. I., Martin, C. H., Hulsey, C. D., Wainwright, P. C., & Near, T. J. (2013). Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20131733.

Grande, L. (1985). The use of paleontology in systematics and biogeography, and a time control refinement for historical biogeography. *Paleobiology*, 11, 234-243.

Hilton, E. J., & Lavoué, S. (2018). A review of the systematic biology of fossil and living bony-tongue fishes, Osteoglossomorpha (Actinopterygii: Teleostei). *Neotropical Ichthyology*, 16, e180031.

Humphries, C. J., & Parenti, L. R. (1986). *Cladistic Biogeography*. Oxford Monographs on Biogeography 2. Oxford: Clarendon Press.

Lavoué, S. (2016). Was Gondwanan breakup the cause of the intercontinental distribution of Osteoglossiformes? A time-calibrated phylogenetic test combining molecular, morphological, and paleontological evidence. *Molecular phylogenetics and evolution*, 99, 34-43.

Lundberg J. G. (1993). African–South American freshwater fish clades and continental drift: problems with a paradigm. In: Goldblatt, P. (ed.) *Biological Relationships Between Africa and South America*, pp. 156– 199. Yale University Press, New Haven, CT.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic biology*, 63, 951-970.

Meseguer, A. S., Lobo, J. M., Ree, R., Beerling, D. J., & Sanmartín, I. (2015). Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of Hypericum (Hypericaceae). *Systematic Biology*, 64, 215-232.

Moyle, P. B., & Cech, J. J. Jr. (2000). *Fishes. An Introduction to Ichthyology*, 4th edition. Prentice-Hall Inc., Upper Saddle River, NJ.

Müller, R. D., Seton, M., Zahirovic, S., Williams, S. E., Matthews, K. J., Wright, N. M., Shephard, G. E., Maloney, K. T., Barnett-Moore, N., Hosseinpour, M., & Bower, D. J. (2016). Ocean basin evolution and global-scale plate reorganization events since Pangea breakup. *Annual Review of Earth and Planetary Sciences*, *44*, 107-138.

Murphy, W. J., & Collier, G. E. (1997). A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism. *Molecular Biology and Evolution*, 14, 790-799.

Oliveros, C. H., Andersen, M. J., Hosner, P. A., Mauck III, W. M., Sheldon, F. H., Cracraft, J., & Moyle, R. G. (2020). Rapid Laurasian diversification of a pantropical bird family during the Oligocene–Miocene transition. *Ibis*, 162, 137-152.

Parenti, L. R. (1981). A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bulletin of the American Museum of Natural History*, 4, 341-557.

Parenti, L. R. (1991). Ocean basins and the biogeography of freshwater fishes. *Australian Systematic Botany*, 4, 137-149.

Patterson, C. (1975). The distribution of Mesozoic freshwater fishes. *Mémoires du Muséum national d'histoire naturelle Série A Zoologie*, 88, 156-174.

Patterson, C. (1981). Methods of paleobiogeography. In: Nelson G. & Rosen, D. E. (eds.) *Vicariance Biogeography: a Critique*, pp. 446-489. Columbia University Press: New York.

Paxton, J. R., & Eschmeyer, W. N., eds. (1998). *Encyclopedia of Fishes*, 2nd edition. Academic Press: San Diego

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57, 4-14.

Rosen, D. E. (1978). Vicariant patterns and historical explanation in biogeography. *Systematic zoology*, 27, 159-188.

Rosen, B. R. (1988). From fossils to earth history: applied historical biogeography. In: Myers, A. A. & Giller, P.S. (eds.) *Analytical biogeography*, pp. 437-481. New York: Chapman & Hall.

Sanmartín, I. (2012). Historical biogeography: evolution in time and space. *Evolution: Education and Outreach*, 5, 555-568.

Silvestro, D., Zizka, A., Bacon, C. D., Cascales-Minana, B., Salamin, N., & Antonelli, A. (2016). Fossil biogeography: a new model to infer dispersal, extinction and sampling from palaeontological data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150225.

Tavares, V. D. C., Warsi, O. M., Balseiro, F., Mancina, C. A., & Dávalos, L. M. (2018). Out of the Antilles: fossil phylogenies support reverse colonization of bats to South America. *Journal of Biogeography*, 45, 859-873.

Wallace, A. R. (1876). *The geographical distribution of animals; with a study of the relations of living and extinct faunas as elucidating the past changes of the Earth's surface.* Harper & Brothers, New York.

Wilson, M. V. H., & Murray, A. M. (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.

CHAPTER 2

Vicariance and Dispersal in Southern Hemisphere Freshwater Fish Clades: a Palaeontological Perspective

Note: The contents of this chapter have been published[1]. Supplementary materials for this chapter are available at:
https://onlinelibrary.wiley.com/doi/full/10.1111/brv.12473?login=true#supporting-information

ABSTRACT

Widespread fish clades that occur mainly or exclusively in fresh water represent a key target of biogeographical investigation due to limited potential for crossing marine barriers. Timescales for the origin and diversification of these groups are crucial tests of vicariant scenarios in which continental break-ups shaped modern geographic distributions. Evolutionary chronologies are commonly estimated through node-based palaeontological calibration of molecular phylogenies, but this approach ignores most of the temporal information encoded in the known fossil record of a given taxon. Here, we review the fossil record of freshwater fish clades with a distribution encompassing disjunct landmasses in the southern hemisphere. Palaeontologically derived temporal and geographic data were used to infer the plausible biogeographic processes that shaped the distribution of these clades. For seven extant clades with a relatively well-known fossil record, we used the stratigraphic distribution of their fossils to estimate confidence intervals on their times of origin. To do this, we employed a Bayesian framework that considers non-uniform preservation potential of freshwater fish fossils through time, as well as uncertainty in the absolute age of fossil horizons. We provide the following estimates for the origin times of these clades: Lepidosireniformes [125–95 million years ago (Ma)]; total-group Osteoglossomorpha (207–167 Ma); Characiformes (120–95 Ma; a younger estimate of 97–75 Ma when controversial Cenomanian fossils are excluded); Galaxiidae (235–21 Ma); Cyprinodontiformes (80–67 Ma); Channidae (79–43 Ma); Percichthyidae (127–69 Ma). These

dates are mostly congruent with published molecular timetree estimates, despite the use of semi-independent data. Our reassessment of the biogeographic history of southern hemisphere freshwater fishes shows that long-distance dispersals and regional extinctions can confound and erode pre-existing vicariance-driven patterns. It is probable that disjunct distributions in many extant groups result from complex biogeographic processes that took place during the Late Cretaceous and Cenozoic. Although long-distance dispersals shaped the distributions of several freshwater fish clades, their exact mechanisms and their impact on broader macroevolutionary and ecological dynamics are still unclear and require further investigation.

Key words: historical biogeography, vicariance, long-distance dispersal, freshwater fishes, evolutionary timescales, palaeontology, fossil record.

[1] **Capobianco, A. & Friedman, M., 2019.** Vicariance and dispersal in southern hemisphere freshwater fish clades: a palaeontological perspective. *Biological Reviews*, 94: 662–699.

I. INTRODUCTION

Freshwater fishes are a fundamental component of the biosphere, constituting more than 20% of living vertebrate species (Nelson, Grande & Wilson, 2016). Extant freshwater fish clades with intercontinental, disjunct distributions have long been model systems in historical biogeography, as seas and oceans represent a relatively strong barrier to their dispersal (Lundberg, 1993).

Continental vicariance driven by Mesozoic breakup of Pangaea is a widely cited explanation for these disjunct distributions (e.g. Novacek & Marshall, 1976; Parenti, 1981; Greenwood, 1983; Chakrabarty, 2004; Sparks & Smith, 2005; Inoue *et al.*, 2009). Alternative scenarios involve more recent long-distance dispersals, *via* land (through transient land bridges or after continental collision) or sea. Despite obvious challenges, trans-oceanic dispersal has been increasingly proposed as the probable mechanism underlying the intercontinental distributions of several terrestrial and freshwater groups (e.g. de Queiroz, 2005; Poux *et al.*, 2006; Pramuk *et al.*, 2008; Samonds *et al.*, 2012), including some freshwater fish clades (Lundberg, 1993; McDowall, 2002; Bonde, 2008; Friedman *et al.*, 2013). Time is the critical variable in testing whether distributions matching those predicted by vicariance arose by this mechanism (Upchurch & Hunn, 2002;

Donoghue & Moore, 2003). Vicariance can be ruled out if lineages with a disjunct distribution are too young to have been influenced by the corresponding geologic event (e.g. breakup between South America and Africa for a clade inhabiting both continents).

Traditionally, fossils and their stratigraphic context have been the only source of information on evolutionary timescales relevant to vicariance hypotheses. In the last 20 years, advances in molecular clock methods have revolutionized the field of evolutionary biology (Ho & Duchêne, 2014), and construction of a molecular time-calibrated tree is now the conventional approach for timing evolutionary events. However, fossils remain the principal source of time calibration for molecular trees, requiring a thorough understanding of the fossil record in order to select calibrations and appropriate parameters properly for timetree analysis (Parham *et al.*, 2012). Alternative methods for estimating the time of origin of a group rely only on palaeontological and stratigraphic data (Strauss & Sadler, 1989; Marshall, 1997; Hedman, 2010), but are used less frequently than molecular clocks.

Herein, we consider existing fossil and molecular evidence for the evolutionary timescale of freshwater fish clades with a widespread disjunct distribution that includes southern hemisphere landmasses. We use phylogenetic and palaeobiogeographic information to infer possible biogeographic patterns for these clades, and to evaluate whether vicariance associated with the Mesozoic breakup of Gondwana, dispersal, or both shaped their geographic distribution. We excluded taxa with a distribution limited to the northern hemisphere, as during the Mesozoic and Cenozoic North America and Eurasia were often connected by transient land bridges (e.g. the Beringian and Thulean land bridges; Brikiatis, 2014, 2016). Biotic exchanges between former Laurasian landmasses were relatively common in the late Mesozoic and Cenozoic and involved several freshwater fish taxa (see Cavin, 2017), including sturgeons (Choudhury & Dick, 1998), bowfins (Grande & Bemis, 1998), cypriniforms (Imoto *et al.*, 2013) and pikes (Grande, 1999).

While we cover both extant taxa with no (or limited) fossil record (Section II.1) and extinct taxa known only from the fossil record (Section II.2), particular attention is given to seven extant freshwater fish clades with more extensive fossil records: Lepidosireniformes (South American and African lungfishes), Osteoglossomorpha (bonytongues and allies), Characiformes (characins and allies), Galaxiidae, Cyprinodontiformes (killifishes), Channidae (snakeheads) and Percichthyidae (Southern temperate perches). Most of these groups (with the notable exceptions

of galaxiids and cyprinodontiforms) are usually classified as primary division freshwater fishes (Myers, 1938), an ecological term indicating low tolerance to salinity. Although widely used, Myers' (1938) classification of freshwater fishes is purely qualitative, has no bearing on ancestral environmental adaptations (i.e. whether a group of freshwater fishes derives from freshwater or marine ancestors) and does not necessary reflect the dispersal abilities of a fish clade.

For the seven focal clades listed above, we used the temporal distribution of their fossil record quantitatively to estimate their origin times, building upon the theoretical framework developed by Marshall (1997). This method utilizes an empirically derived fossil preservation potential function to assess, for a given taxon, the plausible extent of an early evolutionary history undetected by its fossil record (in other words, how much older than its oldest known fossil can a taxon plausibly be). By so doing, it accounts for non-uniform fossil preservation in time. Furthermore, we modified the method to consider uncertainty in the absolute age of fossil-bearing deposits. The origin-time estimates derived with this method were then compared with the timescale of Gondwanan fragmentation to test for vicariant scenarios, and with published molecular estimates to check for congruency or discrepancy.

II. FRESHWATER FISH CLADES WITH INTERCONTINENTAL DISTRIBUTIONS

(1) Extant taxa with disjunct distributions and no (or limited) fossil record

Biogeographic hypotheses for clades with limited palaeontological records are generally assessed through phylogenies that are time-calibrated with fossils of other groups. Many freshwater fish clades with disjunct distributions fall under this category.

(*a*) *Mordaciidae and Geotriidae*

Southern hemisphere lampreys inhabit southern South America and southern Oceania. The four species in these groups are either anadromous or secondarily restricted to freshwater (Potter *et al.*, 2015), suggesting high dispersal potential. Indeed, the monotypic *Geotria* inhabits river systems throughout southern South America, New Zealand and southern Australia, making it one

of the most widespread freshwater fishes (Berra, 2007). The unresolved phylogenetic position of *Geotria* relative to mordaciids and northern hemisphere lampreys (Potter *et al.*, 2015) and the lack of published timetrees for lampreys preclude further testing of biogeographic scenarios.

(*b*) *Atheriniformes*

Within atheriniforms (silversides), the Malagasy Bedotiidae is closely aligned to an Australasian group including Melanotaeniidae, Pseudomugilidae and Telmatherinidae. This relationship has been interpreted as evidence of Cretaceous vicariance between Indo-Madagascar and Austro-Antarctica (Sparks & Smith, 2004). However, fossil-calibrated phylogenies identify an Eocene divergence between bedotiids and Australasian taxa (Campanella *et al.*, 2015), contradicting the vicariant hypothesis. Many silverside clades show repeated freshwater invasions by marine ancestors, and the last common ancestor of bedotiids and the Australasian clades was likely marine or at least euryhaline. Marine dispersal followed by freshwater invasion better explains the biogeographic pattern seen in this group (Campanella *et al.*, 2015).

(*c*) *Synbranchidae*

Swamp eels occur in fresh and brackish waters of Central and South America, West Africa, East Asia, Indo-Malaysia and northern Oceania. Many synbranchids show broad salinity tolerance, and air breathing allows extensive survival out of water (Graham, 1997). Relationships within synbranchids are poorly known (Rosen & Greenwood, 1976). Nonetheless, a latest Cretaceous divergence of synbranchids from their closest living relatives (Near *et al.*, 2013) and the intercontinental distributions of *Monopteros* and *Ophisternon* (Rosen & Greenwood, 1976) imply multiple long-distance dispersal events.

(*d*) *Mastacembelidae*

Spiny eels inhabit Indo-Malaysia and Africa, with one species restricted to the Middle East. Phylogenetic analyses support an Indo-Malayan origin for mastacembelids, followed by dispersal to the Middle East and from there to Africa during the Miocene (Day *et al.*, 2017). This

is congruent with the African invasion of several Asian mammals starting around 18 million years ago (Ma) (Koufos, Kostopoulos & Vlachou, 2005).

(*e*) *Anabantidae*

Climbing gouramies contain the Indo-Malayan *Anabas* and three African genera. *Anabas* can tolerate long periods of air exposure, move on land, and traverse small obstacles (Davenport & Abdul Matin, 1990; Graham, 1997). A single fossil anabantid, *Eoanabas thibetana*, is known from late Oligocene deposits of central Tibet (Wu *et al.*, 2017). The basal position of *Eoanabas* and *Anabas* within anabantids, as well as their affinity to several freshwater clades endemic to Southeast Asia (Betancur-R *et al.*, 2017), implies an Indo-Malayan origin. Anabantid dispersal from Asia to Africa probably occurred during the second half of the Paleogene (Rüber, Britz & Zardoya, 2006).

(*f*) *Polycentridae*

Polycentrids include African and South American leaffishes. Collins, Britz & Rüber (2015) resolved the South American leaffishes as a clade within African leaffishes. There is no time-calibrated phylogenetic analysis targeting polycentrids, but more inclusive timetrees suggest an Eocene divergence between South American leaffishes and the African *Polycentropsis* (Near *et al.*, 2013). This would imply transoceanic dispersal from Africa to South America in the Paleogene, paralleling the well-known cases of monkeys and caviomorph rodents (Poux *et al.*, 2006).

(*g*) *Gobioidei*

Among gobies, multiple lineages with marine ancestors colonized freshwater environments. Some of these (e.g. Milyeringidae, Butidae, Eleotridae, Sicydiinae) display disjunct intercontinental distributions. The fossil record of gobies extends to the early Eocene (Bannikov & Carnevale, 2016). However, the uncertain systematic position of early fossil gobies prevents an accurate estimate of the goby evolutionary timescale based exclusively on fossils (Bannikov

& Carnevale, 2016). Molecular clock estimates indicate that crown gobies are Late Cretaceous–Paleocene in age (Alfaro *et al.*, 2018; Li *et al.*, 2018). Among goby lineages with intercontinental distribution in freshwater environments, butids and eleotrids can thrive in a wide range of salinities, with some species inhabiting coastal marine habitats (Berra, 2007). Thus, marine dispersal is a likely explanation for their widespread distribution. Sycidiines have an amphidromous life cycle. Molecular data suggest a late Miocene origin in the western Pacific Ocean and arrival in Africa and the New World through current-driven westward marine dispersal (Keith *et al.*, 2011). The most striking case is that of the blind cave gobies belonging to Milyeringidae, which includes two genera of obligate troglobic fishes: the Malagasy *Typhleotris* and Australian *Milyeringa* (Chakrabarty, Davis & Sparks, 2012). Chakrabarty *et al.* (2012) proposed a vicariant scenario with an Early Cretaceous origin of this group, but their molecular estimate for the divergence between *Typhleotris* and *Milyeringa* ranges from the Early Cretaceous to the Eocene. An Early Cretaceous origin for a goby subclade is in stark contrast not only with the known fossil record of gobies, but also with the fossil record of acanthomorphs as a whole (Patterson, 1993). More recent studies place the origin of milyeringids firmly within the Cenozoic (Li *et al.*, 2018). Although a recent milyeringid origin would imply at least one long-distance dispersal event between Madagascar and Australia, such an event seems highly unlikely for troglobites with marked physiological limitations and very restricted habitat (Chakrabarty *et al.*, 2012). The possibility of two independent invasions of the subterranean environment from extinct marine or brackish ancestors, followed by independent acquisition of characters typical to troglobic organisms (loss of functional eyes, loss of pigmentation, and so on), cannot be excluded and could explain the striking biogeographic pattern displayed by milyeringids. However, the lack of milyeringid fossils precludes further assessment of this hypothesis.

(2) Fossil taxa with disjunct distributions

Several cases of disjunct distributions in freshwater fishes are known exclusively from the fossil record. These fall into two broad categories: widespread extinct clades; or extant clades with present distribution restricted to only one landmass, but for which fossils are found on multiple continents. Most cases discussed here are associated with the opening of the South Atlantic, as

Mesozoic and early Cenozoic freshwater deposits of South America and Africa are much better sampled than those of other southern landmasses.

(*a*) *†Mawsoniidae*

Mawsoniids represent a primarily continental radiation of Mesozoic coelacanths. †*Mawsonia* and †*Axelrodichthys* have been found in South American and African deposits spanning from the Early Cretaceous to the Cenomanian (de Carvalho & Maisey, 2008). Persistence of these mawsoniid genera in South America and Africa during opening of the South Atlantic suggests vicariance. Post-Cenomanian mawsoniids are known only from Europe and Madagascar (Gottfried, Rogers & Rogers, 2004; Cavin, Valentin & Garcia, 2016), hinting at possible dispersals from Africa in the Late Cretaceous. Cretaceous mawsoniids are often found in brackish deposits and thus they could have had relatively high salinity tolerance and long-distance dispersal potential.

(*b*) *Polypteridae*

Bichirs are an exclusively freshwater clade of early diverging actinopterygians that today occurs only in Africa, where their fossil record extends back to the Cenomanian (Gayet, Meunier & Werner, 2002; Grandstaff *et al.*, 2012; Cavin *et al.*, 2015; Cavin, 2017). Fragmentary polypterid remains from the Maastrichtian and Paleocene of Bolivia reveal a more widespread distribution of this group in the past (Gayet *et al.*, 2002). Undescribed polypterid material from the Albian–Cenomanian Alcântara Formation of Brazil (Candeiro *et al.*, 2011) suggests polypterid presence in South America pre-dating South America–Africa breakup. However, the lack of a phylogenetic framework for fragmentary fossil polypterids precludes a reliable reconstruction of their biogeographic history. The recent recognition of scanilepiforms – known from Triassic freshwater deposits of North America and Eurasia – as stem polypterids (Giles *et al.*, 2017) suggests a Pangean distribution in the early Mesozoic, followed by vicariance and regional extinctions.

(*c*) *Lepisosteidae*

Gars, like the only other extant holostean lineage (the bowfin *Amia*), are now restricted to North America. Lepisosteids have a broad Late Cretaceous distribution, with North American, South American, European, Central Asian, African, Malagasy and Indian deposits yielding gar fossils of this age (Grande, 2010). The majority of the Late Cretaceous lepisosteid material is fragmentary and diagnostic only to family, so biogeographic scenarios are difficult to reconstruct. While extant gars are mainly freshwater fishes and most fossils are found in continental deposits, some living species are occasionally found in brackish and coastal marine environments (notably *Atractosteus tristoechus*, the Cuban gar; Grande, 2010). Moreover, the discovery of early lepisosteids in Late Jurassic marine deposits from Mexico (Brito, Alvarado-Ortega & Meunier, 2017) suggests that high salinity tolerance might be primitive for the group. Marine dispersal probably played a major role in the widespread distribution of lepisosteids during the Cretaceous.

(*d*) *†Obaichthyidae and other lepisosteoids*

Obaichthyids are the sister taxon to Lepisosteidae and consist of two Aptian–Cenomanian genera: †*Obaichthys* and †*Dentilepisosteus*. Like mawsoniid coelacanths, both genera are present in Brazilian and Moroccan continental and transitional deposits (Grande, 2010), suggesting vicariance during late stages of the opening of the South Atlantic. A similar pattern can be inferred for the basal lepisosteoids †*Araripelepidotes* and †*Pliodetes* from the Aptian of Brazil and Niger, respectively, which might be sister lineages (Cavin, 2010).

(*e*) *†Vidalamiinae*

Vidalamiins are a Cretaceous–early Paleogene clade of amiids closely related to the extant bowfin *Amia*. Within vidalamiins, †Calamopleurini occurs only in western Gondwana while †Vidalamiini has a broader distribution including North America, South America, Europe and the Middle East (Grande & Bemis, 1998; Brito, Yabumoto & Grande, 2008). While the geographic and temporal distribution of calamopleurine fossils is consistent with vicariance

related to the rifting of South America and Africa, the biogeographic history of Vidalamiini appears more complex and likely involved marine dispersals. Vidalamiin fossils derive from continental and coastal marine deposits, and several species were likely euryhaline (Grande & Bemis, 1998).

(*f*) *†Archaeomenidae and †Luisiellidae*

Archaeomenids and luisiellids are poorly known freshwater stem teleost groups with a southern Gondwanan distribution (Sferco, López-Arbarello & Báez, 2015; Bean, 2017). The age of these taxa (†Archaeomenidae: Early Jurassic–Early Cretaceous; †Luisiellidae: Late Jurassic–Early Cretaceous) is consistent with a continuous Jurassic range encompassing South America, Antarctica and Australia [but see Su (1994) for a putative archaeomenid from the Early Jurassic of China].

(*g*) *†Cladocyclidae*

Cladocyclids include freshwater, brackish and coastal forms belonging to the primarily marine ichthyodectiforms, a clade of predatory stem teleosts. †*Cladocyclus* and †*Chiromystus* are both known from the Early–middle Cretaceous of South America and Africa (Martill *et al.*, 2011; Cavin, Forey & Giersch, 2013), paralleling the pattern seen in mawsoniids, obaichthyids and vidalamiins. Additionally, †*Cladocyclus* is known from Albian continental deposits of Australia (Berrell *et al.*, 2014) and possibly Italy (Signore *et al.*, 2006). As cladocyclids are often found in lagoonal and coastal marine deposits, at least some species were probably euryhaline. Thus, their palaeobiogeographic distribution may have been shaped by a combination of dispersal and vicariance.

(*h*) *Chanidae*

Milkfishes, an ostariophysan clade, include the living marine *Chanos chanos* and several extinct species, with some found in continental and transitional deposits. †*Dastilbe* and †*Parachanos* are of particular interest. These are found in Aptian–Albian deposits of Brazil and Central Africa,

respectively (Fara, Gayet & Taverne, 2010), and could be sister taxa (Near, Dornburg & Friedman, 2014). †*Dastilbe batai* from the Aptian–Albian of Equatorial Guinea is poorly preserved and may belong to the genus †*Parachanos* (Dietze, 2007). Thus, the palaeobiogeographic distribution of †*Dastilbe* and †*Parachanos* at the end of the Early Cretaceous is consistent with vicariance associated with opening of the South Atlantic. Notably, †*Parachanos* is also known from Late Cretaceous deposits of Italy and Croatia (Fara *et al.*, 2010); long-distance dispersal from Africa could explain the post-Albian European distribution of this taxon, similar to mawsoniid coelacanths. Another freshwater chanid, †*Nanaichthys* from the Aptian of Brazil, reveals a possible trans-Tethyan dispersal event during the Early Cretaceous, as this genus appears to be closely related to the Berriasian–Barremian †*Rubiesichthys* and †*Gordichthys* from Spain (Amaral & Brito, 2012).

(3) Extant taxa with disjunct distributions and known fossil record

Evolutionary timescales, and associated biogeographic scenarios, for geographically widespread extant clades can be assessed by both molecular timescales and the temporal and geographic distribution of their fossils. Seven of these clades are covered in detail herein: Lepidosireniformes, Osteoglossomorpha, Characiformes, Galaxiidae, Cyprinodontiformes, Channidae and Percichthyidae (Fig. 2.1). For these taxa, we reviewed their fossil record focusing on biogeographically relevant fossils. Then, we used the stratigraphic distribution of their fossils to infer times of evolutionary origin in a Bayesian framework. Finally, biogeographic scenarios involving vicariance and dispersal were evaluated on the basis of our fossil-based estimates and published molecular timetrees.

We did not include three clades prominently featured in the historical biogeography literature: Dipnoi, Siluriformes and Cichlidae. These groups (and the reasons for exclusion from this review) will be briefly discussed here.

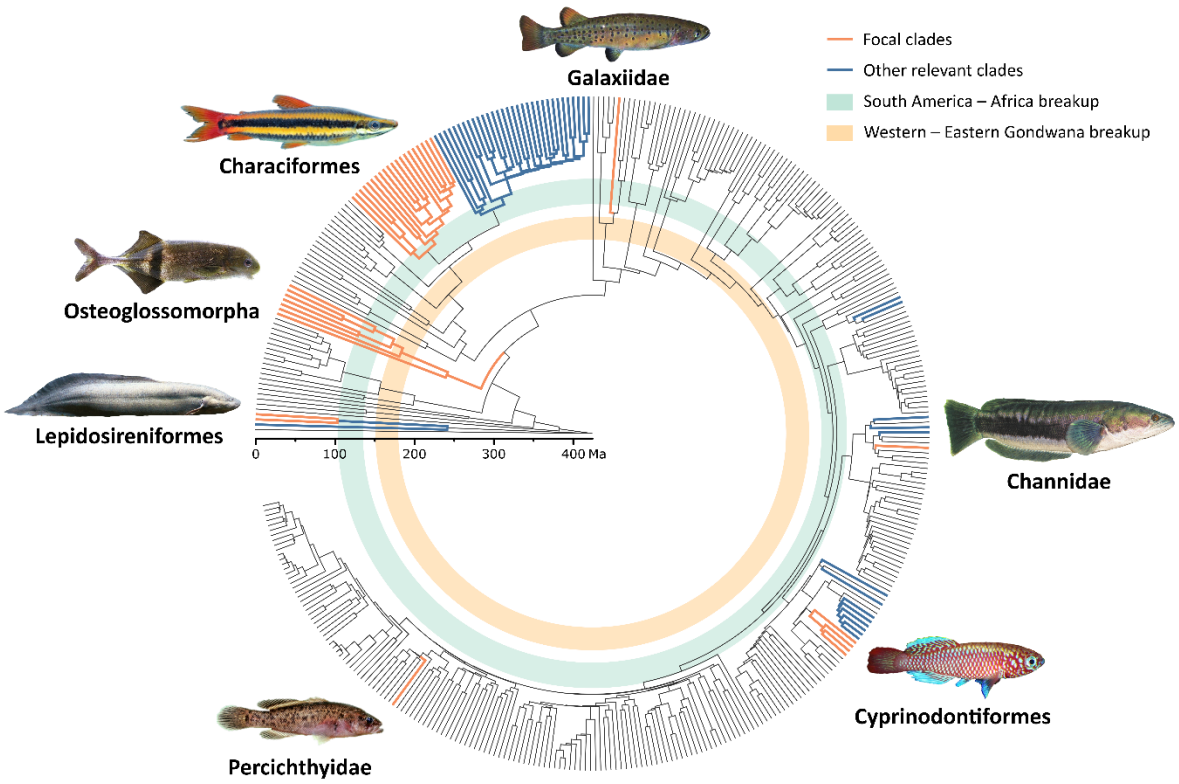

Fig. 2.1. Family-level time-calibrated molecular phylogeny of extant non-tetrapod Osteichthyes (bony fishes), modified from Betancur-R *et al.* (2015). The seven clades of widespread freshwater fishes that represent the focus of this review are highlighted in orange, while other extant clades with a disjunct distribution in the southern hemisphere that are discussed in the text are highlighted in aquamarine. Coloured bands indicate the timeframe of the Western–Eastern Gondwana break-up and the South America–Africa break-up.

Dipnoi (crown lungfishes) includes Lepidosireniformes (South American *Lepidosiren* and African *Protopterus*) and Ceratodontiformes (the Australian *Neoceratodus*). Crown lepidosireniforms are discussed below in the context of the split between South America and Africa, but the early biogeographic history of crown lungfishes has been linked to vicariance and the progressive fragmentation of Gondwana (Cavin *et al.*, 2007). The relationships of several Mesozoic lungfish genera relative to extant ones are still debated, leading to considerable uncertainty for the age of the dipnoan crown. Some phylogenetic studies recover all extinct Mesozoic genera as stem lungfishes, placing the origin of crown lungfishes in the Late Jurassic (Schultze, 2004). By contrast, other analyses find several early Mesozoic genera (e.g. †*Ceratodus*, †*Arganodus*, †*Asiatoceratodus* and †*Gosfordia*) within the lungfish crown (Cavin *et al.*, 2007; Longrich, 2017). It has even been suggested that Permian lungfishes like †*Gnathorhiza*

may be more closely related to Lepidosireniformes than to *Neoceratodus* (Kemp, Cavin & Guinot, 2017), placing the minimum age for the origin of crown lungfishes to around 300 Ma. The identification of Triassic or Permian lungfishes as stem lepidosireniforms, coupled with their widespread geographic distribution, would strongly suggest a Gondwanan (if not Pangean) distribution of early crown lungfishes, followed by a series of vicariant events and local extinctions (Cavin *et al.*, 2007). Little effort has been put into the development of a precise timescale for lungfish evolution from a molecular clock perspective, with recent estimates for crown lungfishes ranging from the Permian to the Late Jurassic (Irisarri *et al.*, 2017). Because of the uncertain affinities of early Mesozoic lungfish genera, we do not estimate the age of crown lungfishes using quantitative biostratigraphical models here. However, lepidosireniforms are considered in this framework below.

Siluriformes (catfishes) is a major clade of globally distributed otophysans that includes several thousand species. While phylogenetic analyses strongly support the South American endemics Loricarioidei and Diplomystidae as the earliest branching lineages in the siluriform tree (implying a South American origin for the group), deep-level relationships among other siluriforms – collectively grouped in Siluroidei – remain largely unknown (Betancur-R *et al.*, 2017). It is therefore not easy to identify biogeographically relevant nodes in the siluriform phylogeny (i.e. nodes corresponding to disjunct intercontinental distributions). The siluriform fossil record extends to the Late Cretaceous of South America (Gayet, 1990). However, these early fossils are fragmentary and cannot be confidently assigned to any extant lineage. Because of the uncertainties in siluriform systematics and in the affinities of the earliest siluriform fossils, we refrain from discussing the siluriform fossil record and biogeography in detail here. However, there are indications of long-distance dispersal in siluriform evolutionary history. First, several lineages of catfishes are adapted to high-salinity environments, with Ariidae and Plotosidae including mostly coastal marine species (Berra, 2007). Specifically, ariids recolonized freshwater environments after marine dispersal several times during their history, achieving a worldwide distribution in tropical fresh waters (Betancur-R, 2010). More remarkably, molecular phylogenetics resolves the recently discovered *Lacantunia enigmatica* from Mexico as deeply nested within a diverse group of African catfishes (the 'Big Africa' clade) with strong statistical support (Lundberg *et al.*, 2007). Molecular clock studies place origin of the 'Big Africa' clade during the Late Cretaceous (Lundberg *et al.*, 2007). Thus, the presence of a member of this

radiation in Mexico requires a biogeographic scenario that involves complex dispersal routes (Lundberg *et al.*, 2007). A better understanding of siluriform historical biogeography will depend on the resolution of their deep-level phylogeny and on further analysis of the early fossil record of catfishes.

Cichlidae (cichlids) is a model system for several fields in evolutionary biology, including historical biogeography. The 'Gondwanan' geographic distribution of cichlids (which includes the Neotropics, Africa, Madagascar and the Indian subcontinent) has been the focus of considerable attention among biogeographers. The topological congruence between cichlid phylogeny and Gondwanan fragmentation (with the Malagasy and Indian lineages branching first and the African clade being sister group to the South American one) has been often interpreted as evidence for vicariance (Chakrabarty, 2004; Sparks & Smith, 2005). However, this argument does not take into account the timescale of cichlid evolution, which would be necessary to test a vicariant hypothesis. Topological patterns that appear to be consistent with a vicariant scenario may arise from dispersal events, a phenomenon called pseudo-congruence (Donoghue & Moore, 2003). Most recent molecular-clock studies agree on a Late Cretaceous–Paleocene origin of crown cichlids, inconsistent with the vicariant scenario (Friedman *et al.*, 2013; Matschiner *et al.*, 2017). Matschiner (in press) reviews more than 15 years of cichlid molecular-clock studies and their implications for the group's biogeographic history. The oldest cichlid fossils are relatively recent, from middle Eocene deposits of Africa and South America (Murray, 2000*a*; Malabarba, Malabarba & López-Fernández, 2014). However, their derived anatomy suggests that a long portion of the early cichlid fossil record might be missing. Friedman *et al.* (2013) estimated the timing of cichlid origin based on the temporal distribution of their fossil record, using a comparable methodology to that applied here (see Section III). They found that, even when accounting for non-uniform fossil preservation through time, the estimated time of origin only extends to the Late Cretaceous (Campanian), around 77 Ma. While Friedman *et al.* (2013) refer to this estimate as the age for crown cichlids, it more conservatively marks divergence between South American and African cichlids, as every known cichlid fossil belongs to either Cichlinae (the Neotropical cichlid clade) or Pseudocrenilabrinae (the African cichlid clade). Nonetheless, even a Campanian age for the split between cichlines and pseudocrenilabrines would reject the hypothesis of vicariance and suggest a transatlantic dispersal event in the early history of cichlids. Because of the amount of literature discussing vicariance and dispersal in cichlid

biogeography under several different approaches, we do not consider this group in more detail here.

III. MATERIALS AND METHODS

(1) Estimation of origin times of focal clades using their fossil occurrences

We derived fossil-based estimates of the dates of origin for the seven fish groups mentioned above (Lepidosireniformes, Osteoglossomorpha, Characiformes, Galaxiidae, Cyprinodontiformes, Channidae and Percichthyidae) and for some of their sub-clades. Our method is based on the number and distribution through time of known stratigraphic horizons that yielded fossils belonging the group of interest. This approach builds upon the theoretical framework developed by Marshall (1997), which accounts for non-uniform fossil preservation and recovery through time by using an empirically derived function of recovery potential. We combined this framework with the Bayesian probability estimate for the extension of observed stratigraphic ranges developed by Strauss & Sadler (1989) to calculate 95% credibility intervals (CIs) for the origin times of focal clades.

(*a*) *Bayesian probability estimate for the extension of observed stratigraphic ranges*

Strauss & Sadler (1989) were the first to propose a Bayesian estimate for stratigraphic CIs for a given focal group. They derived the posterior density function of the endpoint θ of a stratigraphic range given the data x as:

$$h(\theta|x) = \frac{(n-2)[(\theta - y)^{-n+1} - \theta^{-n+1}]}{u_n} \qquad [1]$$

where y is the age of the last observed fossil (last appearance datum), n is the number of fossil horizons and u_n is a factor calculated by the equation:

$$u_n = (z - y)^{-n+2} - (1 - y)^{-n+2} - z^{-n+2} + 1 \qquad [2]$$

with z being the age of the first observed fossil (first appearance datum). The posterior density function given above is valid for each θ larger than z and smaller than a prior upper bound; θ, y

and z are rescaled to have the prior upper bound equal to 1. This formula assumes a uniform prior distribution of the fossil horizons bounded between 0 and 1, a condition that is almost always violated by the empirical fossil record.

The Bayesian point estimator of θ [that is, the mean of Equation (1)] is given by:

$$\frac{(n-2)u_{n-1}}{(n-3)u_n} + \frac{y[(z-y)^{-n+2} - (1-y)^{-n+2}]}{u_n} \qquad [3]$$

(*b*) *Extension to non-random distributions of fossil horizons*

In order to relax the strong assumption of uniform distribution of fossil horizons, we utilized the logical framework, introduced by Marshall (1997), of a preservation and recovery potential function. Marshall (1997) extended the use of stratigraphic confidence intervals for non-random distributions of fossil horizons by reframing the problem in terms of recovery potential rather than time. Given a function representing preservation and recovery potential over time, the area under this function between the age of the first observed fossil and the age of the last observed fossil corresponds to the duration of the focal clade (in units of preservation potential). The confidence limit for the origin time of this clade is the point at which the area under the preservation potential function between the first appearance and that point is equal to the duration of the lineage in units of preservation potential multiplied by a scaling factor that reflects the number of distinct fossil occurrences and the desired level of confidence.

Friedman *et al*. (2013) applied this framework to Strauss & Sadler's (1989) Bayesian estimate to account for heterogeneity through time in the fossil record of freshwater fishes. They measured θ, y and z of Equations (1–3) in terms of summed preservation potential and not in terms of time. In order to calculate the area under the preservation potential function easily, geological time was divided into time bins, with each bin being assigned a value equal to the proportion between the number of fossil horizons that yielded fossils of the group of interest and the total number of fossil horizons. Doing this, a uniform distribution of fossil horizons is assumed only within each time bin, and not throughout the entire fossil record. Posterior distributions, Bayesian point estimates and 95% CIs were then calculated in terms of accrued preservation potential, and later converted in terms of absolute time in light of their empirical function for preservation potential.

Herein, we applied the same method employed in Friedman *et al*. (2013) with a few adjustments. We corrected the script of Friedman *et al*. (2013) by adding a term that was missing in their calculation of u_n [Equation (2)]. However, we ascertained that this had no significant effect on the results, as that term is several orders of magnitude smaller than the resulting origin time estimate. We also employed a different empirical preservation potential function, the main difference being the use of time bins of 1 million years (Myr) each rather than corresponding to chronostratigraphic epochs (see Section III.3). Finally, we considered uncertainty in the absolute age of fossil horizons.

(*c*) *Extension to uncertain absolute age of fossil horizons*

Uncertainty in the absolute age of fossil horizons was considered by generating 1000 replicates for each Bayesian time-estimate analysis. In each replicate, every horizon was assigned an age randomly drawn from a uniform distribution bounded by minimum and maximum age of the chronostratigraphic stage (or stages) corresponding to that horizon. The absolute ages for chronostratigraphic epochs and stages were taken from the ICS International Chronostratigraphic Chart (v. 2016/12). Median and two-tailed 95% confidence intervals for the Bayesian estimates on origin times (summarized by their 95% CIs) were then calculated among the replicates.

(2) Assembly of fossil occurrence data sets

Fossil occurrences for the seven focal clades were compiled through a comprehensive literature search (see online Supporting information, Tables S1–S7). Different stratigraphic formations (or localities in cases of no formalized stratigraphy) were treated as different sampling horizons. The age of each horizon (to stage level, when possible) was assigned according to the literature. Marine fossil occurrences of the focal clades were pruned from the analysis, as accounting for marine deposits throughout the fossil record could heavily bias the recovery potential function.

(3) Estimation of the empirical recovery potential function

The recovery potential function used to account for non-uniformity in fossil preservation and recovery through time was derived empirically using a list of stratigraphic horizons (formations and/or localities) with the potential to yield fossils belonging to the group of interest. For every freshwater fish clade analysed here, this criterion was satisfied by non-marine deposits that yielded fish fossils. A list of non-marine deposits that yielded fish fossils was compiled through literature search and implemented with records from the Paleobiology Database (PBDB; https://paleobiodb.org). The beginning of the Permian (around 299 Ma) was chosen as the upper limit for the age of fossil horizons: this represents the prior upper bound on the Bayesian estimates for the origin times of the focal clades. This is a conservative prior, as it does not artificially exclude vicariance scenarios; moreover, the oldest fossils belonging to the analysed clades come from the Middle Jurassic (around 167 Ma). Although some molecular clock estimates place the origin of total-group Osteoglossomorpha in the Carboniferous (e.g. Inoue *et al.*, 2009), a Carboniferous origin for any crown-teleost clade is in strong disagreement with the fossil record (Arratia, 2015; Friedman, 2015).

The list of non-marine fossil fish deposits comprised a total of 935 unique horizons, ranging from the early Permian to the Holocene (Table S8). Fossil horizons were subdivided into seven broad, continental-scale geographic areas (North and Central America; South America; Europe and Western Asia; Africa and Arabian Peninsula; Northeastern Asia; Indo-Malaya; Oceania). For each clade, only fossil horizons from relevant geographic areas (i.e. areas in which the clade is either present today or was present in the past according to the fossil record) were included (Table 2.1). The discrete recovery potential function was built by dividing geologic time into bins of 1 Myr each, with every bin being assigned a value equal to the total number of fossil horizons present in that time interval. In so doing, uniform recovery potential was assumed within each time bin.

All calculations were performed in R version 3.4.1 (R Core Team, 2017). The script is available as Appendix S1.

	North and Central America	South America	Europe and Western Asia	Africa and Arabian Peninsula	North-Eastern Asia	Indo-Malaya	Oceania
Lepidosireniformes		**X**		**X**			
Osteoglossomorpha	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Osteoglossidae	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Characiformes	**X**	**X**	**X**	**X**			
Alestidae			**X**	**X**			
Galaxiidae		**X**		**X**			**X**
Cyprinodontiformes	**X**	**X**	**X**	**X**		**X**	
Cyprinodontoidei	**X**	**X**	**X**	**X**			
Channidae			**X**	**X**	**X**	**X**	
Percichthyidae		**X**					**X**

Table 2.1. Biogeographic areas selected for each of the analysed clades to build their empirical preservation potential function. X indicates areas in which the clade is either present today or was present in the past according to the fossil record.

IV. RESULTS AND DISCUSSION

Table 2.2 summarizes the ages of origin of the freshwater fish clades considered here, as estimated from the stratigraphic distribution of fossil occurrences. Range estimates encompass uncertainty in fossil horizon age (i.e. they span from the lower confidence interval of the lower CI of the posterior distribution to the upper confidence interval of the upper CI of the posterior distribution).

Results for seven focal clades are discussed below in the context of their fossil record and geographic distribution. Comparisons with molecular timescales permit a comprehensive view of the biogeographic history for each group at a continental scale.

	Replicates lower 95%	Median point estimate	Replicates upper 95%
Lepidosireniformes	95.05	103.51	124.93
Total-group Osteoglossomorpha	167.03	182.44	206.89
Osteoglossidae (without *Chanopsis*)	72.07	82.85	112.96
Osteoglossidae (with *Chanopsis*)	103.22	123.96	154.42
Characiformes (with Cenomanian occurrences)	95.08	102.47	119.84
Characiformes (without Cenomanian occurrences)	75.07	83.40	97.30
Alestidae	53.13	60.37	72.10
Galaxiidae	21.15	97.13	235.02
Cyprinodontiformes	67.02	70.72	79.97
Cyprinodontoidei	42.02	46.27	54.77

Channidae	43.08	53.19	78.70
Percichthyidae	69.10	87.59	127.39

Table 2.2. Fossil-based estimates for the time of origin of widespread freshwater fish clades considered in this study. Columns indicate lower 95% confidence interval of the lower credibility interval (CI) of the Bayesian posterior distribution, median point estimate, and upper 95% confidence interval of the upper CI of the Bayesian posterior distribution, respectively. Values result from 1000 replicates accounting for uncertainty in absolute age of fossil horizons. All numbers are in units of million years ago (Ma).

(1) Lepidosireniformes (South American and African lungfishes)

Lepidosireniformes (*sensu* Otero, 2011) includes two living genera, the South American *Lepidosiren* (one extant species) and the African *Protopterus* (four extant species). Molecular and morphological data support monophyly of the group (e.g. Betancur-R *et al.*, 2013; Criswell, 2015). Lepidosireniform fossils comprise mainly tooth plates and jaw fragments, with some exceptions (Table S1) (Silva Santos, 1987). Crown lepidosireniforms are distinguished on the basis of tooth plate characters (Otero, 2011; Longrich, 2017). Like modern species, fossils of the group are restricted to South America and Africa (Fig. 2.2). The oldest fossils of *Lepidosiren* derive from the Late Cretaceous El Molino Formation (Maastrichtian of Bolivia; Schultze, 1991) and Vilquechico Formation (?Coniacian–Maastrichtian of Peru; Arratia & Cione, 1996). †*Protopterus nigeriensis* from the Cenomanian Wadi Milk Formation of Sudan might represent the oldest African crown lepidosireniform (Claeson *et al.*, 2014). However, Longrich (2017) did not find conclusive evidence for assigning this species or other Late Cretaceous–Eocene African fossils to *Protopterus*, and it is not clear whether they belong within the lepidosireniform crown. Leaving aside possible polyphyly of the genus (when including fossils), *Protopterus* is represented in the African record by up to eight different species (six extinct) and hundreds of specimens without specific attribution, ranging from the Late Cretaceous to the Holocene (Otero, 2011).

Extant *Protopterus* and *Lepidosiren* are strictly freshwater (Berra, 2007) and deposits yielding fossils of these genera are generally freshwater or estuarine (Cavin *et al.*, 2007). Past work cites

these environmental associations as supporting a vicariant model for the South American–African distribution of Lepidosireniformes (Lundberg, 1993; Otero, 2011). The early Late Cretaceous age of the first crown lepidosireniform fossils is consistent with vicariance. Some Mesozoic (and many Paleozoic) lungfishes outside Lepidosireniformes are known from marine deposits, leading some to hypothesize primitive marine habits and independently acquired freshwater adaptations among the living lungfish genera (Schultze, 1991). However, most (if not all) of the marine Mesozoic fossils probably represent remains of freshwater animals that have been reworked into marine deposits (Cavin *et al.*, 2007).

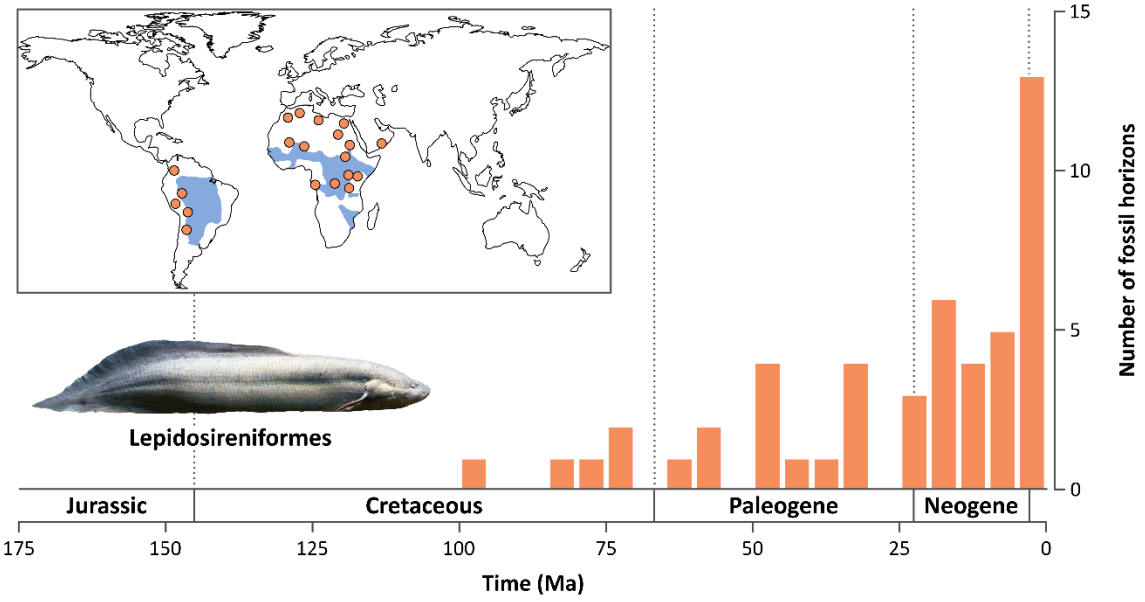

Fig. 2.2. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of crown Lepidosireniformes. Each time bin is 5 million years (Myr) in width. The inset displays the present-day geographic distribution of Lepidosireniformes (in blue), as well as the main localities in which lepidosireniform fossils have been found (orange dots). Extant geographic ranges for Figs. 2.2–2.8 were taken from Berra (2007). Photograph of West African lungfish (*Protopterus annectens*) from Wikimedia Commons.

(*a*) *Fossil-based estimate of origin times*

The origin of crown Lepidosireniformes is hereby estimated to occur between the Aptian and the Cenomanian (124.9–95.1 Ma; median point estimate: 103.5 Ma); this overlaps with fragmentation of Western Gondwana (South America + Africa; Heine, Zoethout & Müller,

2013). Our fossil-based age estimate is consistent with molecular divergence times between *Protopterus* and *Lepidosiren* (estimates centered around 112–96 Ma; Broughton *et al.*, 2013; Giles *et al.*, 2017). The limited suite of dental characters used for the systematics of extinct lepidosireniforms results in some ambiguity in the placement of some fossil remains. The possible exclusion of Late Cretaceous taxa like †*Protopterus nigeriensis* from the genus *Protopterus* (Longrich, 2017) could strongly impact the fossil-based estimate of the age of origin for the group, making it substantially younger. Nonetheless, the currently known timescale for lepidosireniform origin and evolution (based on fossil and molecular data) does not reject the vicariance hypothesis. The disjunct distribution of extant Lepidosireniformes can probably be considered as the genuine product of an ancient vicariant event.

(2) Osteoglossomorpha (bonytongues and allies)

Osteoglossomorpha is one of the earliest diverging lineages of modern teleosts (Arratia, 1999; Near *et al.*, 2012), comprising 246 living species distributed across the Americas, Africa, the Indo-Malayan region and Australia (Nelson *et al.*, 2016). The osteoglossomorph fossil record is rich (Table S2), with more than 80 extinct species, and expands the present distribution of the group to Europe and Northeastern Asia (Wilson & Murray, 2008) (Fig. 2.3).

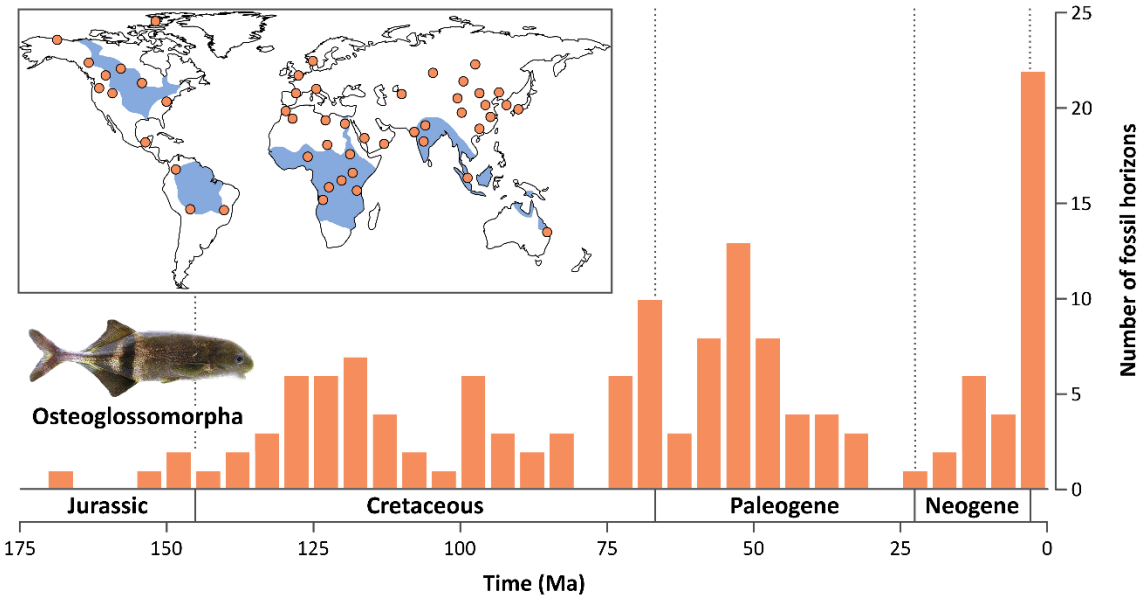

Fig. 2.3. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of total-group Osteoglossomorpha. Each time bin is 5 million years (Myr) in width. The inset displays the present-day geographic distribution of Osteoglossomorpha (in blue), as well as the main localities in which osteoglossomorph fossils have been found (orange dots). Photograph of Lokundi mormyrid (*Hippopotamyrus castor*) modified from Sullivan, Lavoué & Hopkins (2016).

Morphological and molecular data strongly support osteoglossomorph monophyly, but interpretations of intrarelationships (reviewed in Hilton, 2003) have changed considerably over time. Current classifications recognize six main lineages (Nelson *et al.*, 2016): Hiodontiformes (the sister group to all other living osteoglossomorphs; Hilton, 2003), Pantodontidae, Notopteridae, Gymnarchidae, Mormyridae and Osteoglossidae (grouped together as Osteoglossiformes).

The oldest articulated osteoglossomorph fossils belong to †*Paralycoptera* and derive from the Late Jurassic Lai Chi Chong Formation of Hong Kong and Fenshuiling Formation of Shandong, China (Tse, Pittman & Chang, 2015). The phylogenetic placement of †*Paralycoptera* is uncertain. Some analyses place it on the osteoglossomorph stem (Wilson & Murray, 2008) and others recover it as a crown osteoglossiform (Li & Wilson, 1999; Zhang, 2006). Fossil squamules from the Anoual Formation of Morocco could push back the earliest osteoglossomorph occurrence in the fossil record to the Middle Jurassic (early Bathonian; Haddoumi *et al.*, 2016). The otolith-based genus †*Archaeglossus* (Schwarzhans, 2018) from the marine Middle–Late Jurassic of England might also represent an early osteoglossomorph. The presence of early Mesozoic osteoglossomorphs in marine sediments would not be completely unexpected, as crown teleosts probably originated in marine environments (Betancur-R, Ortí & Pyron, 2015). Early Cretaceous deposits from Northeastern Asia (Russia, Mongolia, China, Korea and Japan) yield numerous early osteoglossomorphs (Wilson & Murray, 2008). Many of these fossils belong to the abundant †*Lycoptera* or closely related stem osteoglossomorphs (Li & Wilson, 1999). However, some of these Asian genera (e.g. †*Huashia*, †*Kuntulunia*, †*Xixiaichthys*) are unstable in phylogenetic analyses (Li & Wilson, 1999; Zhang, 2006; Wilson & Murray, 2008).

The oldest definitive crown osteoglossomorph is †*Yanbiania wangqinica*, a hiodontiform from the Aptian–Albian Dalazi Formation of China (Li & Wilson, 1999). Fossil hiodontiforms are also known from Late Cretaceous deposits in North America and Asia (Newbrey *et al.*, 2013;

Brinkman, Newbrey & Neuman, 2014), but extant *Hiodon* is restricted to North America. Among Osteoglossiformes, pantodontids, gymnarchids and mormyrids are endemic to Africa, and have a meagre fossil record in African Cenozoic deposits (Wilson & Murray, 2008). Notopterids show a disjunct distribution, with two African and two Indo-Malayan genera. Notopterid fossils are limited to otoliths from the latest Maastrichtian of India (Nolf, Rana & Prasad, 2008) and a few articulated specimens from the Eocene–Oligocene of Sumatra (Sangkarewang Formation; Sanders, 1934; de Smet & Barber, 2005). †*Palaeonotopterus greenwoodi* from the early Late Cretaceous (Cenomanian) Kem Kem Beds of Morocco was originally interpreted as a notopterid (Forey, 1997), but it probably represents a basal member of the clade that groups together Notopteridae, Mormyridae and Gymnarchidae (Wilson & Murray, 2008). Nonetheless, †*Palaeonotopterus* demonstrates that key divergences within crown osteoglossiforms had occurred by 100 Ma.

Extant osteoglossids comprise two sub-clades, each with an intercontinental distribution: Arapaiminae (*sensu* Forey & Hilton, 2010) inhabits South America (*Arapaima*) and Africa (*Heterotis*), while Osteoglossinae is distributed across South America (*Osteoglossum*), Southeast Asia and northern Australia (*Scleropages*). Osteoglossid fossils are known from every continent (except Antarctica) and show a higher diversity of the group in the past. †*Chanopsis lombardi* from the late Early Cretaceous (Aptian–Albian) Loia and Bokungu formations of the Democratic Republic of Congo (DRC) could represent the oldest member of Osteoglossidae (Taverne, 2016). Although †*Chanopsis* shows features characteristic of some osteoglossid sub-groups (e.g. lateral expansion of the anterior end of the frontal) it lacks definitive osteoglossid synapomorphies (Forey & Hilton, 2010) and has never been included in a formal phylogenetic analysis. Other putative early osteoglossids include †*Laeliichthys* from the Aptian of Brazil and †*Paradercetis* from the Late Cretaceous of DRC; both taxa have been assigned to Arapaiminae and feature prominently in discussions about the biogeography of the clade (Taverne, 1979; Lundberg, 1993). However, characters suggesting a relationship between *Laeliichthys* and Arapaiminae might be plesiomorphies or homoplasies (Forey & Hilton, 2010), while †*Paradercetis* is known from a poorly preserved skull roof without any clear osteoglossomorph features (Capobianco A., personal observation of MRAC RG 10.970). It is advisable to exclude these taxa from discussions about osteoglossid evolution and biogeography pending further study. †*Laeliichthys* and †*Paradercetis* aside, jaw fragments from the Maastrichtian El Molino Formation of Bolivia

could represent the oldest arapaimines (Gayet *et al.*, 2001). Osteoglossine fossils are rare, but articulated specimens of *Scleropages* from the early–middle Eocene of China (Xiawanpu and Yangxi formations; Zhang & Wilson, 2017) lie outside the current geographic range of the genus. Perhaps unexpectedly, worldwide marine deposits of Paleocene–early Eocene age yield the highest diversity of fossil osteoglossids (e.g. †*Brychaetus*, †*Furichthys*, †*Heterosteoglossum*, †*Magnigena*, †*Opsithrissops*; Bonde, 2008; Forey & Hilton, 2010). Taverne (1979) grouped some of the marine osteoglossids with the freshwater *Phareodus* in Phareodontinae. However, †*Magnigena* and †*Opsithrissops* do not seem to be closely related to †*Brychaetus* (Forey & Hilton 2010), implying multiple marine invasions. Reexamination of early Cenozoic osteoglossids (including marine forms) is necessary to untangle the complex evolutionary and biogeographic history of Osteoglossidae.

Extant osteoglossomorphs are restricted to fresh waters, with notopterids occasionally found in brackish environments (Berra, 2007). Thus, their distribution (encompassing all southern landmasses except for Antarctica) has been the subject of various biogeographic hypotheses (Nelson, 1969; Greenwood, 1973; Lundberg, 1993; Wilson & Murray, 2008). Africa has been proposed as the osteoglossomorph centre of origin (in a dispersalist scenario) due to the presence of every major extant osteoglossomorph lineage (except Hiodontidae; Darlington, 1957). However, the fossil record shows the highest diversity of Late Jurassic–Early Cretaceous osteoglossomorphs in northeastern Asia. Whether this pattern is due to an Asian origin or to geographical bias in the continental sedimentary record is not clear. Another scenario (Kumazawa & Nishida, 2000) involves a widespread Pangaean distribution during the Permian–Triassic for which there is no palaeontological evidence despite a wealth of fossil fishes of this age (Romano *et al.*, 2016). Cavin (2017) proposed a Laurasia–Gondwana vicariant event during the Jurassic corresponding to the divergence between the Laurasian Hiodontiformes and the Gondwanan Osteoglossiformes. The highly unstable phylogenetic position of several basal osteoglossomorphs (and possibly osteoglossiforms) from the Cretaceous and early Paleogene of North America and Asia (Murray *et al.*, 2018) makes this hypothesis difficult to evaluate at present.

The cosmopolitan distribution (encompassing North and South America, Africa, Europe, continental Asia, Indo-Malaya and Australia) of marine osteoglossomorphs and †*Phareodus*-like

freshwater osteoglossids in the early Paleogene suggests a role for long-distance marine dispersal (Bonde, 2008; Wilson & Murray, 2008). Thus, the disjunct modern distribution of Arapaiminae and Osteoglossinae could be explained by marine dispersal followed by colonization of freshwater environments.

(*a*) *Fossil-based estimate of origin times*

The fossil-based estimate for total-group Osteoglossomorpha ranges from the Late Triassic to the Middle Jurassic (Rhaetian–Bathonian: 206.9–167.0 Ma; median point estimate: 182.4 Ma), suggesting an early ghost lineage extending for up to 40 Myr. The time of origin of total-group Osteoglossomorpha is closely linked to the origin of the teleost crown, as either osteoglossomorphs or elopomorphs (or a clade including both) represent the sister group to all other living teleosts (Arratia, 2010; Dornburg *et al.*, 2014; Hughes *et al.*, 2018). The age discordance between the oldest crown-teleost fossils, found in Late Jurassic deposits (except for some very fragmentary Middle Jurassic remains; Haddoumi *et al.*, 2016), and molecular clock estimates, which range from the Late Carboniferous to the Late Triassic (Near *et al.*, 2012; Broughton *et al.*, 2013; Dornburg *et al.*, 2014), has been called the 'teleost gap' (Near *et al.*, 2012). It represents one of the most striking differences between fossil and molecular timescales that still remains partially unexplained. Incompleteness of the fossil record and failure to recognize early crown-teleost fossils are not sufficient explanations for this phenomenon (Sallan, 2014), and specific choices of calibration points for molecular phylogenies play some role (Friedman, 2015; Giles *et al.*, 2017). The fossil-based estimate derived here for total-group Osteoglossomorpha partially bridges that gap, possibly extending the origin of this group as far back as the latest Triassic. Still, a significant difference of at least 15–40 Myr remains, suggesting the need for a revision of molecular clock studies focused on broad-scale teleost relationships.

While total-group Osteoglossomorpha is old enough to have been affected by the breakup of Gondwana (and even Pangea), the abundance of basal osteoglossomorphs in areas not occupied by living lineages (northeastern Asia) or with low present-day diversity (North America) suggests a complex history where dispersal and/or local extinction might have played a fundamental role. Moreover, at least three subclades that are deeply nested within

Osteoglossomorpha (Notopteridae, Osteoglossinae, Arapaiminae) show disjunct distributions. The sparse fossil record of notopterids indicates that the group was already present in the Indian subcontinent by the end of the Cretaceous, but it cannot be used to derive an informative estimate for its time of origin. Molecular estimates of divergence between African and Asian notopterids show considerable variation (from the Late Jurassic to the Late Cretaceous; Inoue *et al.*, 2009; Lavoué, 2016). Thus, neither an Africa–India vicariance scenario nor a sweepstakes dispersal from Africa to India across the Mozambique Channel can be confidently rejected on the basis of the present evidence.

The fossil record of Arapaiminae and Osteoglossinae gives a minimum latest Cretaceous and early Eocene age for these two clades, respectively. However, derivation of probabilistic fossil-based estimates of their origin times is complicated by inadequate understanding of the relationships of fossil osteoglosssids (Forey & Hilton, 2010). Nonetheless, it is possible to estimate an age for Osteoglossidae as a whole. The fossil-based estimate for osteoglossid origin varies greatly depending on the inclusion or exclusion of †*Chanopsis*: Early Cretaceous and even the latest Jurassic (Tithonian–Albian: 154.4–103.2 Ma; median point estimate: 124.0 Ma) with †*Chanopsis*, or most of the Late Cretaceous (Aptian–Campanian: 113.0–72.1 Ma; median point estimate: 82.8 Ma) excluding this genus. It is clear that the phylogenetic placement of †*Chanopsis* has broad implications on the reconstruction of the early evolutionary history of the group, and a phylogenetic reassessment of this taxon is badly needed. Despite the differences in the fossil-based origin times inferred here relative to the position of †*Chanopsis*, both estimates are approximately consistent with molecular dates for crown Osteoglossidae (Early Cretaceous; Broughton *et al.*, 2013). These dates are old enough to allow for a significant role of continental vicariance, particularly involving South America–Africa drift and the fragmentation of the South America–Antarctica–Australia block, in the biogeographic history of the clade. However, the complex distributional pattern of extant and fossil osteoglossids (Wilson & Murray, 2008; Lavoué, 2016) and the presence of marine forms in the fossil record strongly suggest that dispersal has been a fundamental process during osteoglossid evolution.

(3) Characiformes (characins and allies)

Characiformes is a major clade of otophysans containing more than 2000 species, making it one of the most diverse freshwater fish lineages (Nelson *et al.*, 2016). Extant characiforms are restricted to freshwater environments of Africa and South and Central America, with one species in the southwestern USA (Fig. 2.4).

Numerous morphological characters support characiform monophyly (Wiley & Johnson, 2010), including the presence of multicuspid teeth in the jaws (lost in predators like *Hepsetus* and *Salminus*; Fink & Fink, 1981). The species-poor African Citharinoidei and species-rich Neotropical and African Characoidei represent the principal characiform lineages. Surprisingly, some molecular work questions characiform monophyly (Chen, Lavoué & Mayden, 2013; Chakrabarty *et al.*, 2017), but other analyses suggest these results are spurious (Arcila *et al.*, 2017).

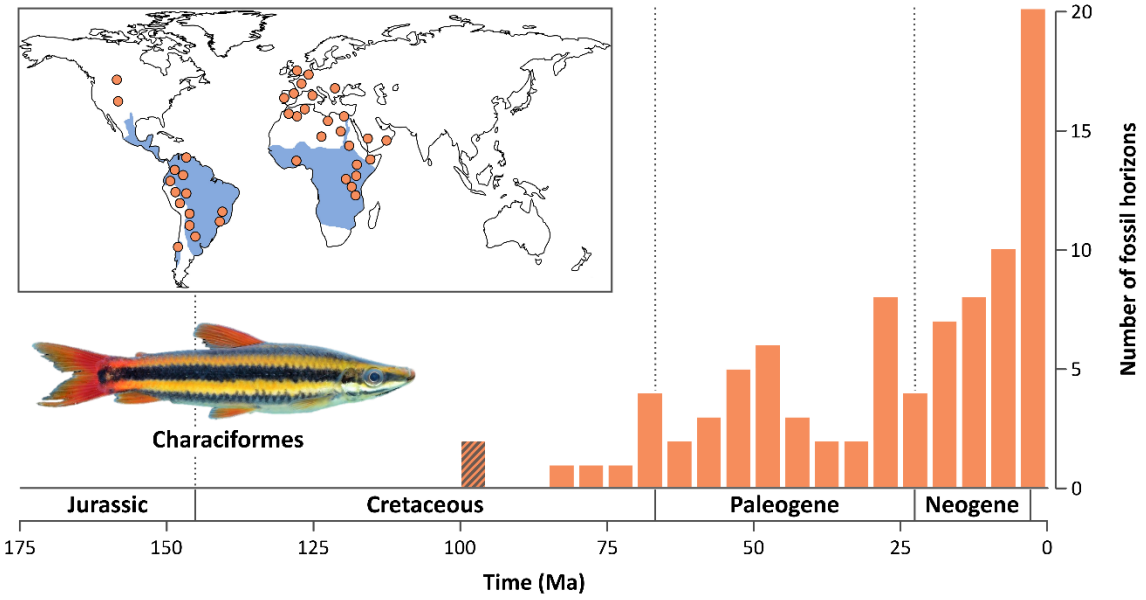

Fig. 2.4. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of Characiformes. Each time bin is 5 million years (Myr) in width. The hatched rectangle represents the doubtful occurrences of characiforms teeth in Cenomanian deposits of Africa. The inset displays the present-day geographic distribution of Characiformes (in blue), as well as the main localities in which characiform fossils have been found (orange dots). Photograph of striped headstander (*Anostomus anostomus*) by J. Armbruster from Wikimedia Commons.

Isolated teeth are the most common characiform fossils (Table S3) (Malabarba & Malabarba, 2010; Gaudant, 2014). These are sufficiently diagnostic to support a characiform attribution but often inadequate for more precise placements. The oldest putative characiform fossil teeth come from the Cenomanian of Morocco (Ifezouane Formation; Dutheil, 1999) and Sudan (Wadi Milk Formation; Werner, 1994). These occurrences would demonstrate presence of the group in Africa shortly after tectonic separation from South America. However, their attribution to characiforms has been challenged and they might instead represent ginglymodian multicuspid teeth, common in Cretaceous continental deposits of Africa, India and China (Cavin, 2017). The African record also yields the oldest articulated characiform, †*Eocitharinus macrognathus* from the middle Eocene Mahenge Formation (Lutetian of Tanzania; Murray, 2003*b*; this is also the earliest known citharinoid). Alestidae, an African subclade of Characoidei, has a relatively abundant fossil record that spans the Cenozoic. Teeth of *Hydrocynus* appear in late Paleocene–early Eocene deposits of Algeria (Hammouda *et al.*, 2016). Possible alestid fossils from the Oligocene Baid Formation of Saudi Arabia (Micklich & Roscher, 1990) and Eocene and middle Miocene deposits of southwestern Europe (Gaudant, 2014) indicate a broader distribution of this clade in the past. Fragmentary material from the Maastrichtian Maevarano Formation of Madagascar has been tentatively referred to Characiformes (Ostrowski, 2012), but requires further study.

The Maastrichtian El Molino Formation of Bolivia is the oldest horizon yielding characiform fossils in South America, which today is home to the greatest diversity of characiforms (Gayet, 1991). Various tooth morphologies are present in these latest Cretaceous Bolivian deposits, indicating that the diversification of modern lineages (characids, serrasalmids and possibly acestrorhynchids) was underway by the end of the Late Cretaceous (Gayet *et al.*, 2001, 2003). Complete fossils of South American characiforms (including bryconids, curimatids, triportheids and several characid lineages) are known from the Eocene–Oligocene Entre-Corregos Formation and the Oligocene Tremembé Formation of southeastern Brazil (Malabarba, 1998; Weiss, Malabarba & Malabarba, 2014).

The recent discovery of putative characiform dentaries and vertebral centra from late Campanian North American deposits (Dinosaur Park and Kaiparowits formations) greatly extends the known geographic range of the group and implies an elaborate biogeographic scenario (Newbrey *et al.*,

2009; Brinkman *et al.*, 2013). Although the dentaries show an interdigitating hinge joint at the symphysis (a character thought to be unique to characiforms), these identifications should be approached with caution given the limited material. Characiforms also appear in the European fossil record by the end of the Cretaceous, with teeth found in Maastrichtian deposits in France and Romania (Grigorescu *et al.*, 1985; Otero, Valentin & Garcia, 2008). There are no extant European characiforms, but fossils are found throughout the Cenozoic (including articulated specimens; Gaudant, 1980), with the youngest examples from the latest Miocene (Gaudant, 2014).

Extant characiforms are strictly freshwater (with isolated brackish records; Lundberg, 1993), and all known fossils come from freshwater or at most brackish deposits. However, marine Early and Late Cretaceous fossils from Europe and South America (†*Salminops*, †*Sorbinicharax* and †*Santanichthys*) have been aligned with characiforms in the past (Gayet, 1985; Taverne, 2003; Filleul & Maisey, 2004), leading to hypotheses of a marine origin for Characiformes and of better dispersal abilities in early characiforms than might be predicted from modern forms (Calcagnotto, Schaefer & DeSalle, 2005; Otero *et al.*, 2008). Restudy of †*Salminops* and †*Sorbinicharax* failed to find evidence that these genera are even otophysans (Mayrinck, Brito & Otero, 2015; Mayrinck *et al.*, 2017). †*Santanichthys* is better interpreted as a basal member of Otophysi or Ostariophysi rather than a stem characiform (Malabarba & Malabarba, 2010). Thus, a marine origin of Characiformes is not supported by palaeontological and phylogenetic data.

Two factors further complicate attempts to reconstruct characiform biogeographic history. First, extant African characiforms belong to three distinct clades (Citharinoidei, Alestidae and the monotypic Hepsetidae). Second, characiform fossils are found in areas outside their present distribution (Fig. 2.4). Several non-mutually exclusive hypotheses have been formulated to explain the presence of three different characiform lineages in Africa: a single vicariant event between Africa and South America when characiforms were already diversified, followed by extinction of several African lineages to account for the rarity of sister pairs between extant American and African clades; multiple vicariant events associated with the diachronous split between South America and Africa; and trans-oceanic dispersal events from South America to Africa, usually associated with the questionable hypothesis of a marine ecology in early characiforms (Lundberg, 1993; Malabarba & Malabarba, 2010). Evaluating these proposals

without a well-supported phylogenetic framework for Characiformes is prohibitive; in fact, apart from the basal split between Citharinoidei and Characoidei, there is no agreement across different analyses about the relationships among major characiform lineages (see Dahdul, 2010). Arcila *et al.* (2017) recently recovered a single African characoid clade, with a strongly supported sister-group relationship between Hepsetidae and Alestidae. Given the low support for most other basal nodes within Characoidei, an alternative hypothesis with a diverse South American characoid clade nested within an African radiation cannot be excluded *a priori*. Under this scenario, only one event (either a pre-drift dispersal, or a post-drift oceanic dispersal, or a vicariant event) would be necessary to explain the current distribution of characiforms. Characiform fossils found in Europe and North America are difficult to interpret in a biogeographic framework, as their phylogenetic affinities are unclear. It has been proposed that European characiforms, which are mainly found in Maastrichtian, early Eocene, Oligocene and middle Miocene deposits, are the result of multiple waves of immigration, presumably from Africa, instead of a single colonization of the continent (Gaudant, 2014). The North American Campanian fossils, if confirmed as characiforms, hint at possible dispersals from South America or Europe (there is evidence for both routes from early Campanian terrestrial vertebrates; Newbrey *et al.*, 2009; Cavin, 2017). The widespread distribution of characiforms in the latest Cretaceous may suggest multiple long-distance dispersal events during the biogeographic history of the clade.

(*a*) *Fossil-based estimate of origin times*

The fossil-based divergence time estimate for characiforms depends heavily on the inclusion or exclusion of the Cenomanian fossil teeth from northern Africa. When including these putative characiform occurrences, our estimate is consistent with a vicariant scenario involving the South America–Africa split, as the origin of the clade is estimated as Albian–Cenomanian (119.8–95.1 Ma; median point estimate: 102.5 Ma). This is generally congruent with molecular clock estimates for the age of crown Characiformes (mostly ranging from 120 to 80 Ma; Near *et al*., 2012; Betancur-R *et al*., 2015). Without Cenomanian occurrences, our estimate shifts forwards in time by around 20 Myr to the Late Cretaceous (Cenomanian–Campanian: 97.3–75.1 Ma; median point estimate: 83.4 Ma), rejecting the vicariant scenario. Thus, a careful taxonomic

reassessment of the Cenomanian multicuspid teeth from the Ifezouane and Wadi Milk formations could substantially impact the reconstruction of characiform biogeographic history. Particular caution should be applied when interpreting these results for two main reasons besides uncertainty on Cenomanian occurrences: the phylogenetic position of most early characiform fossils is unknown, so placement in the crown rather than on the stem is not assured; and the divergence between Citharinoidei and Characoidei may not correspond to a South America–Africa split, if South American characoids are nested within an African radiation. In this last case, the divergence between South American and African characiforms would have occurred later than the citharinoid–characoid split. Considering these two factors, together with the inclusion of Cenomanian fossils, our older estimate is more likely to be a conservative test of the vicariant scenario (i.e. it is likely to be an overestimate of true divergence time rather than underestimate). If we exclude the doubtful Cenomanian fossils, some of the oldest known characiforms – from Maastrichtian and Paleocene deposits – are unambiguous members of modern lineages that are deeply nested within characiform phylogeny (Gayet *et al.*, 2001, 2003). Hence, our younger estimate is more likely to be an underestimate of the true age of characiform origin. The apparent absence in the fossil record of early crown characiforms and the sudden appearance of several derived lineages in the Maastrichtian–Paleocene could be the result of different phenomena, which are not mutually exclusive: an early evolutionary history characterized by low diversification rates, followed by rapid diversification from the Maastrichtian onwards; the lack of appropriate depositional settings in the fossil record to recover Late Cretaceous characiforms; or a high degree of endemism before a rapid geographic expansion at the end of the Late Cretaceous (less likely under a vicariant scenario).

The fossil-based estimate for the origin of the African Alestidae could at most extend to the latest Cretaceous (Maastrichtian–Ypresian: 72.1–53.1 Ma; median point estimate: 60.4 Ma), significantly postdating the separation of South America and Africa. A stable phylogenetic placement of alestids (and of the other African characoid taxon, Hepsetidae) is needed before interpreting this result in light of a biogeographic scenario. Nonetheless, the timescale of alestid evolution is consistent with the emergence of modern characiform lineages during the Maastrichtian–Paleocene. The fossil record of characiforms in Europe hints at multiple dispersals of alestids from Africa during the Cenozoic, a pattern found in other non-marine vertebrates (Koufos *et al.*, 2005; Tabuce & Marivaux, 2005).

(4) Galaxiidae (galaxiids)

Galaxiidae includes more than 50 species of freshwater and diadromous fishes inhabiting temperate regions of the southern hemisphere (southern South America, South Africa, Australia, New Zealand and New Caledonia) (Fig. 2.5). Both morphological and molecular phylogenies strongly support galaxiid monophyly (McDowall & Burridge, 2011; Burridge *et al.*, 2012).

The galaxiid fossil record is restricted to Miocene lacustrine deposits of New Zealand (Table S4) (McDowall & Pole, 1997; Lee, McDowall & Lindqvist, 2007). The earliest examples belong to †*Galaxias effusus* from the early Aquitanian Foulden Hills Diatomite (Lee *et al.*, 2007). The Maastrichtian †*Stompooria rogersmithi* from freshwater deposits of South Africa was originally described as a galaxiid (Anderson, 1998). Although these specimens are articulated, subsequent study indicates they are too poorly preserved to permit precise taxonomic identification (Wilson & Williams, 2010). Significantly, †*Stompooria* differs from living galaxiids in several features, including the presence of scales (McDowall & Burridge, 2011).

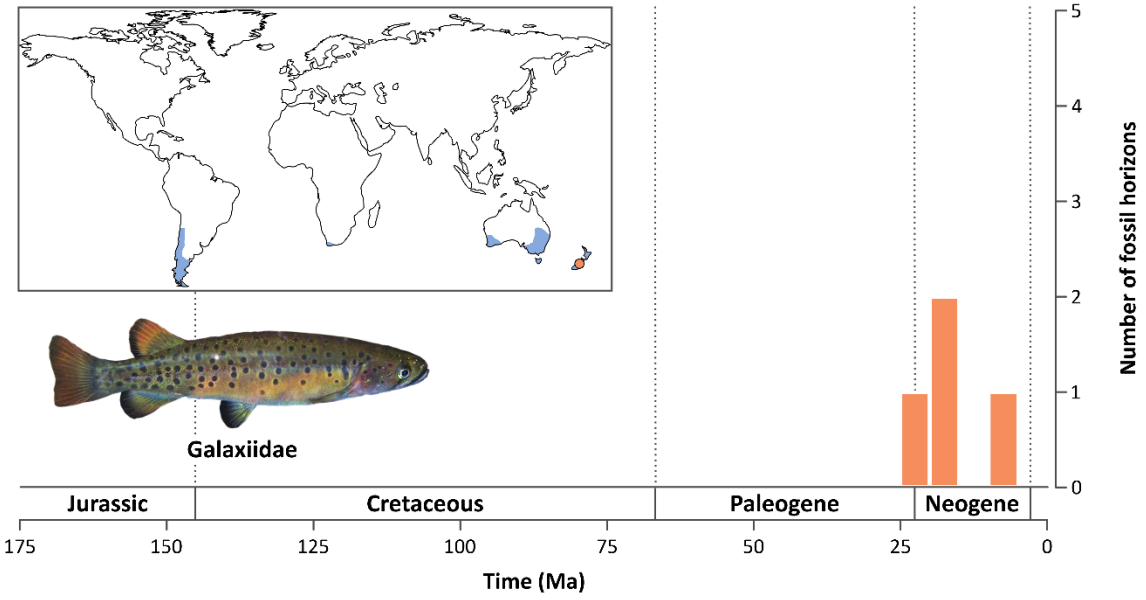

Fig. 2.5. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of Galaxiidae. Each time bin is 5 million years (Myr) in width. The inset displays the present-day geographic distribution of Galaxiidae (in blue), as well as the main localities in which galaxiid fossils have been

found (orange dots). Photograph of spotted galaxias (*Galaxias truttaceus*) by N. Litjens from Wikimedia Commons.

Because of their peculiar distribution and the complex life cycle of some species, galaxiids have been at the centre of a long-standing debate concerning the relative contributions of vicariance and sweepstakes dispersal [see McDowall (2010) for a review]. While most galaxiids are exclusively freshwater, at least 11 species are diadromous (i.e. they migrate between fresh waters and sea during their life cycle; McDowall, 2007). Some diadromous species show broad distributions (e.g. *Galaxias maculatus* occurs in Australia, New Zealand, Chatham Islands, southern South America and Falkland Islands; McDowall, 1972), implying that open seaways are not a barrier to their dispersal. Diadromy has been lost many times during galaxiid evolution, indicated by phylogenetic studies and by the existence of landlocked populations of otherwise diadromous species (Allibone & Wallis, 1993; Waters & Wallis, 2001). Time-calibrated total evidence analyses imply a complex scenario of vicariant events associated with the early divergences followed by multiple marine dispersals since the Oligocene (Burridge *et al.*, 2012). Moreover, ancestral life-history reconstructions show that diadromy cannot be rejected as the ancestral state for most of the nodes corresponding to disjunct geographic distributions (Burridge *et al.*, 2012).

(*a*) *Fossil-based estimate of origin times*

The fossil-based estimate for the origin time of galaxiids is extremely broad and spans the whole Mesozoic and most of the Cenozoic (235.0 – 21.2 Ma; median point estimate: 97.1 Ma), failing to give insight into their biogeographic history. This is a consequence of the very low number of distinct stratigraphic horizons in which galaxiid fossils have been found (only four when excluding †*Stompooria*). Published timetrees place the origin of crown Galaxiidae in the Late Cretaceous–early Paleogene, with a very long stem lineage extending to the Early Cretaceous (Burridge *et al.*, 2012; Betancur-R *et al.*, 2017).

Fossil *Galaxias* from the early Miocene of Otago show that galaxiids were present there shortly after the Oligocene 'drowning' event that almost completely submerged New Zealand [Cooper & Cooper, 1995; Landis *et al.*, 2008; see Sharma & Wheeler (2013) for a critique of this scenario].

This is consistent with the total-evidence analysis of Burridge *et al.* (2012), which indicates that the earliest New Zealand galaxiid clades diverged from their sister groups around the Oligocene–Miocene boundary. Thus, the presence of several lineages of galaxiids in New Zealand is better explained through multiple long-distance dispersal events.

(5) Cyprinodontiformes (killifishes and allies)

Cyprinodontiformes comprises more than 1200 species occurring in the Americas, the Mediterranean region, Africa and Southeast Asia and living predominantly in freshwater and brackish environments. Cyprinodontiform monophyly – and its division into two subclades with approximately equal modern diversity: Aplocheiloidei and Cyprinodontoidei – is strongly supported by morphological and molecular studies (Parenti, 1981; Setiamarga *et al.*, 2008). However, phylogenetic relationships among major killifish lineages (especially within Cyprinodontoidei) differ wildly across studies, with recent molecular phylogenies challenging the monophyly of long-standing taxa like Cyprinodontidae and Poeciliidae (Pohl *et al.*, 2015).

European and North American cyprinodontoids dominate the cyprinodontiform fossil record (Table S5). Very few fossil occurrences are known from Africa and South America, and none from Madagascar, India and Southeast Asia (Fig. 2.6). The oldest fossils referred to Cyprinodontiformes come from the Maastrichtian El Molino Formation of Bolivia (Gayet, 1991). These articulated, poorly preserved specimens do not exhibit typical cyprinodontiform synapomorphies of the caudal skeleton (Arratia & Cione, 1996). The El Molino fossils could represent a very basal lineage of killifishes or small-bodied freshwater fishes unrelated to killifishes. Undescribed material from the middle Eocene Lumbrera Formation of Argentina was listed as an indeterminate poeciliid by Arratia & Cione (1996). The earliest definitive cyprinodontiform fossils come from early Oligocene (Rupelian) deposits of Europe (Spain, France, Switzerland and Germany) and are represented by articulated specimens (Gaudant, 1982; Frey, Maxwell & Sánchez-Villagra, 2016). Numerous killifish species were present in Europe by the end of the Oligocene, probably representing every major living lineage of Old World cyprinodontoids (*Aphanius*-like cyprinodontids, valenciids and procatopodine poeciliids; Costa, 2012). The European genera *Aphanius* and *Valencia* have fossil records that extend to the early and middle Miocene, respectively (Reichenbacher & Kowalke, 2009; Gaudant *et al.*, 2015).

Killifishes also appear in the Oligocene of North America (Coatzingo Formation of Mexico; Guzmán, 2015), and the genus *Fundulus* is first found in early Miocene (Burdigalian) deposits of Nevada (Lugaski, 1977). Other extant killifish genera (*Cyprinodon* and several goodeids) have been found in Pliocene and Pleistocene deposits of the southern USA and Mexico (Smith, 1981; Miller & Smith, 1986). Only one fossil aplocheiloid species has ever been formally described (†*Kenyaichthys kipkechi* from the late Miocene Lukeino Formation of Kenya; Altner & Reichenbacher, 2015). Several fossil aplocheiloid specimens are also known from the Oligocene Daban Formation of Somalia (Van Couvering, 1982), but remain undescribed. These two cases represent the only examples of fossil killifishes in Sub-Saharan Africa.

Several killifishes live in brackish environments, and some fundulids and cyprinodontids inhabit coastal marine settings (Berra, 2007). Nonetheless, the widespread distribution of cyprinodontiforms has been interpreted as a 'reduced Pangaean' distribution by Parenti (1981, p. 534), who argued that the origin of Cyprinodontiformes should extend to the Late Triassic. Similarly, the origins of both cyprinodontids and aplocheiloid killifishes have been hypothesized to have occurred in the Late Jurassic–Early Cretaceous based on modern geographic distributions (Parker & Kornfield, 1995; Murphy & Collier, 1997). Others emphasized the high salinity tolerance shown by several cyprinodontiforms in arguing for marine dispersal, with a South American origin and successive dispersals to Africa during the Late Cretaceous to early Paleogene (Lundberg, 1993; Briggs, 2003).

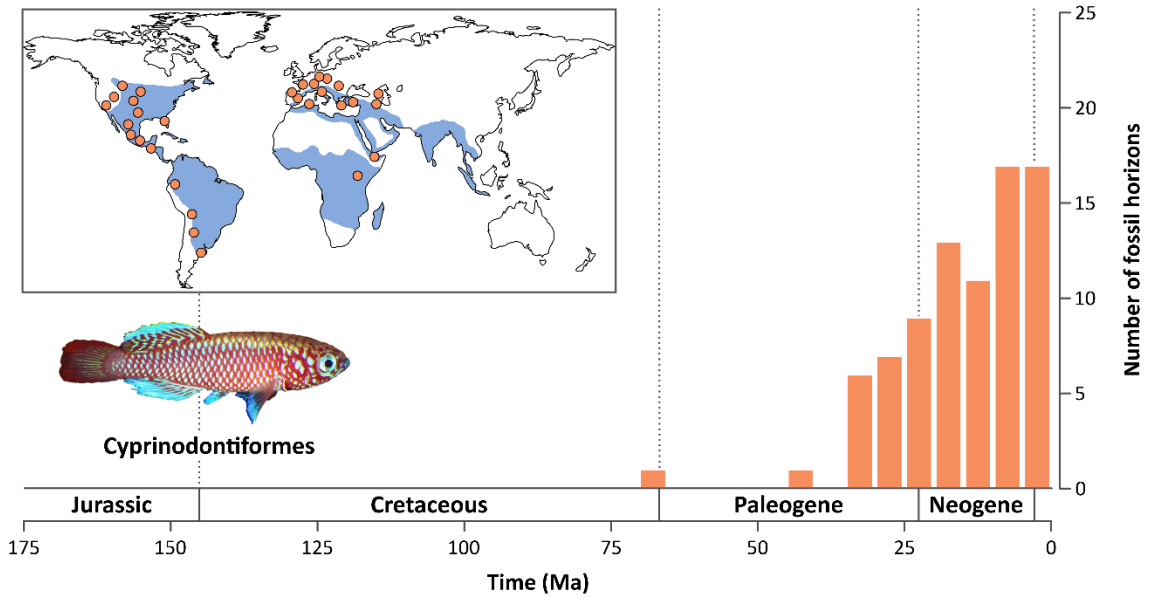

Fig. 2.6. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of Cyprinodontiformes. Each time bin is 5 million years (Myr) in width. The inset displays the present-day geographic distribution of Cyprinodontiformes (in blue), as well as the main localities in which cyprinodontiform fossils have been found (orange dots). Photograph of an African killifish (*Nothobranchius kilomberoensis*) from Wikimedia Commons.

Traditional classifications place Old World cyprinodontoids in three unrelated lineages, but molecular phylogenies resolve them as a clade nested within an American radiation (Pohl *et al.*, 2015). This topology implies only one event (either vicariance or long-distance dispersal) to explain the presence of cyprinodontoids on both sides of the Atlantic.

Recent phylogenies of Aplocheiloidei indicate that African and Indo-Malayan cyprinodontoids are sister lineages (Furness *et al.*, 2015; Pohl *et al.*, 2015), contradicting a hypothesized South American and African clade (Murphy & Collier 1997). The branching order of major clades within Aplocheiloidei is incongruent with the sequence of Gondwanan breakup, suggesting that a purely vicariant scenario is overly simplistic. Unfortunately, the scant aplocheiloid fossil record provides few temporal and biogeographic constraints.

(*a*) *Fossil-based estimate of origin times*

Cyprinodontiformes (with the inclusion of the El Molino fossils) is estimated to originate during the Late Cretaceous (Campanian–Maastrichtian: 80.0–67.0 Ma; median point estimate: 70.7 Ma), whereas its major sub-clade Cyprinodontoidei probably appeared during the early-middle Eocene (Ypresian–Lutetian: 54.8–42.0 Ma; median point estimate: 46.3 Ma).

The fossil-based time estimate for Cyprinodontiformes rejects the vicariant hypothesis for this group, as South America, Africa and the Indo-Malagasy block were already separated from each other by seaways during the Campanian–Maastrichtian (Ali & Aitchison, 2008; Granot & Dyment, 2015). This timescale agrees with recent molecular studies that put the origin of killifishes in the Late Cretaceous (Near *et al.*, 2013; Matschiner *et al.*, 2017). However, this result should be treated with caution for two reasons. First, the fossil-based estimate is strongly reliant on the Maastrichtian El Molino Formation material, whose cyprinodontiform affinity is dubious at best; the next oldest occurrence is around 20 Myr younger than the El Molino fossils. Additionally, the taxonomic distribution of fossil cyprinodontiforms among the two main sub-clades – Cyprinodontoidei and Aplocheiloidei – is extremely uneven, so that the two aplocheiloid occurrences in the Oligocene–Miocene do not contribute to the time estimate derived here. Thus, a time estimate focused only on the cyprinodontoid fossil record may be more reliable than a cyprinodontiform estimate.

The estimated age for Cyprinodontoidei strongly rejects the vicariant hypothesis by placing cyprinodontoid origin in the early–middle Eocene. This is congruent with some molecular estimates (Near *et al.*, 2013; Betancur-R *et al.*, 2017), but significantly younger than others (Matschiner *et al.*, 2017). In any case, a latest Cretaceous–early Paleogene origin for this transatlantic clade strongly suggests a key role of long-distance dispersal in its biogeographic history.

A higher probability of long-distance dispersal events in killifishes compared to other freshwater fishes should be expected on the basis of remarkable physiological, behavioural and life-history traits, including not only high salinity tolerance, but also a facultative amphibious lifestyle, desiccation-resistant eggs and developmental diapause, that are present in at least some members of this group (Turko & Wright, 2015; Furness, 2016). In this regard, killifishes could represent a valuable biogeographic model system to study the timing and directionality of rare biotic exchanges among geographically separated landmasses during the last 80 Myr.

(6) Channidae (snakeheads)

The freshwater, predatory Channidae includes two extant genera: *Parachanna* (Western and Central Africa) and *Channa* (Indo-Malayan region and East Asia) (Fig. 2.7). Together with anabantoids (gouramies and allies), snakeheads are labyrinth fishes (Anabantiformes = Anabantoidei + Channoidei; Wiley & Johnson, 2010). This group is characterized by the presence of the suprabranchial organ, an accessory air-breathing apparatus (Wiley & Johnson, 2010). Channid monophyly is supported by numerous morphological synapomorphies (Wiley & Johnson, 2010; Murray, 2012) and molecular phylogenetic analyses (e.g. Li, Musikasinthorn & Kumazawa, 2006).

The earliest snakehead fossils come from middle Eocene (Lutetian) deposits of Indo-Pakistan and consist mainly of cranial material (Table S6) (Khare, 1976; Murray & Thewissen, 2008). The channid affinity of these middle Eocene fossils is clear, but their exact relationships to modern lineages is unclear. Fragmentary fossils of *Parachanna* appear in late Eocene (early–middle Priabonian) formations of Egypt and Libya (Murray *et al.*, 2010*a*; Otero *et al.*, 2015). More complete cranial remains and isolated vertebrae are known from the latest Eocene–earliest Oligocene Jebel Qatrani Formation in the Fayum Depression (Murray, 2012). Fossil snakeheads are also found in early–middle Miocene deposits of Europe and Central Asia, areas with no extant channids (e.g. Gaudant & Reichenbacher, 1998; Kordikova, Heizmann & Pronin, 2003). Better-preserved specimens are needed to determine whether European fossils belong to *Parachanna* or *Channa* (Gaudant, 2015). The range expansion of *Channa* into East Asia appears to have happened relatively recently, as the oldest snakehead remains in this region come from early Pleistocene deposits of China (Liu & Su, 1962).

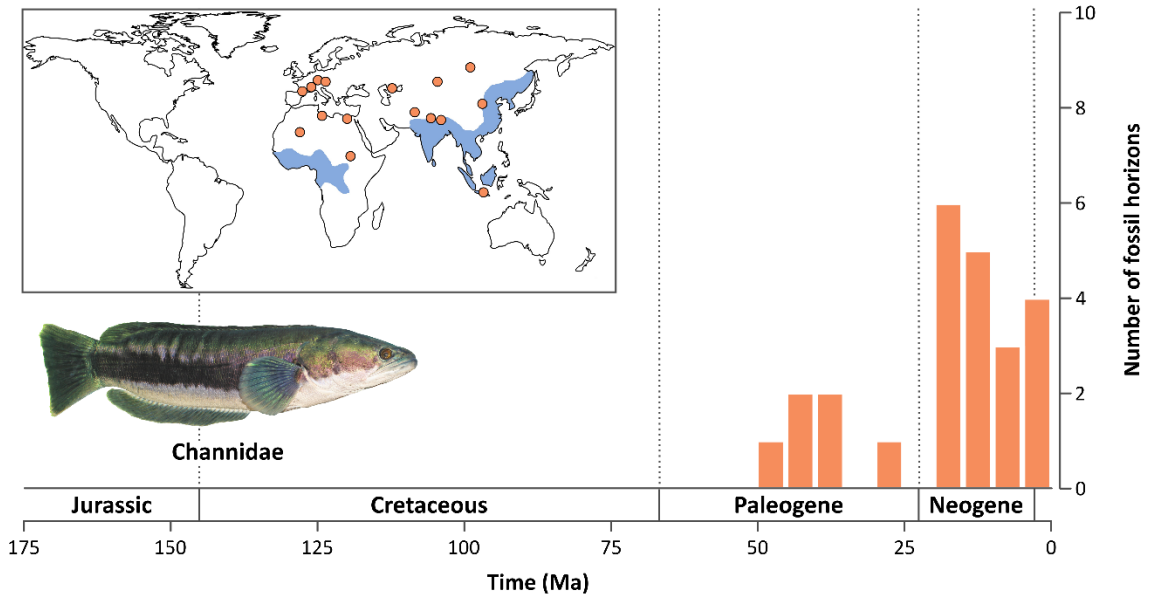

Fig. 2.7. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of Channidae. Each time bin is 5 million years (Myr) in width. The inset displays the present-day geographic distribution of Channidae (in blue), as well as the main localities in which channid fossils have been found (orange dots). Photograph of giant snakehead (*Channa micropeltes*) from Wikimedia Commons.

Snakeheads are currently restricted to freshwater environments, although at least one species (*Channa punctata*) has moderate salinity tolerance and can thrive in brackish waters (Dubey *et al.*, 2016). Fossil snakeheads are usually found in freshwater deposits, although some of the earliest representatives of the group come from estuarine/transitional deposits (Subathu and Birket Qarun formations of India and Egypt, respectively; Khare, 1976; Murray *et al.*, 2010*a*). Channids are facultatively amphibious, can survive outside of water for days in a humid environment and are capable of short bursts of overland movement (Chew *et al.*, 2003). Thus, channids probably have good dispersal potential over the mainland, but they are limited by other environmental factors including water salinity and atmospheric humidity. It has been hypothesized that the geographic distribution of channids has been strongly controlled by climatic variables (precipitation and temperature), and that their presence in Europe and Central Asia during the early–middle Miocene and recent invasion of East Asia reflect broad-scale changes in Eurasian atmospheric circulation patterns (Böhme, 2004).

Two biogeographic scenarios have been proposed for channids. The first involves an origin in the Indo-Malayan region, followed by dispersal to Africa (Briggs, 1995). Although a late

Miocene–early Pliocene age has been previously hypothesized for this dispersal event (Böhme, 2004), *Parachanna* fossils in late Eocene–early Oligocene deposits of northern Africa set a minimum age of around 40 Ma (Murray, 2012). The second scenario postulates a vicariant event between the Indo-Malagasy block and the rest of Gondwana during the Late Jurassic–Early Cretaceous (Li *et al.*, 2006).

Regardless of scenario, the fossil record of channids implies dispersal to Europe by 20 Ma. Gaudant (2015) proposed Africa as the source of immigration on the basis of palaeobiogeographic affinities between Europe and Africa during the early–middle Miocene. Specifically, European fossil channids have been found in association with specimens of alestid characiforms, a group now restricted to Africa. However, a phylogenetic appraisal of the European channids is needed to distinguish between African and Asian origins.

(*a*) *Fossil-based estimate of origin times*

The fossil-based estimate for the origin of Channidae ranges from the Late Cretaceous to the Eocene (Campanian–Lutetian: 78.7–43.1 Ma; median point estimate: 53.2 Ma), long after the separation of the Indian subcontinent from continental Africa. Thus, it rejects the hypothesis of Early Cretaceous vicariance associated with the *Parachanna–Channa* divergence. Instead, this date is consistent with the hypothesis of origin in the Indian (or Indo-Malagasy) subcontinent, followed by dispersal into Africa before the late Eocene. Although the exact timing of initial collision between India and continental Asia is still debated (ranging between 50 and 35 Ma; Ali & Aitchison, 2008; Najman *et al.*, 2010), the fossil record of terrestrial mammals shows a strong signal of biotic exchange between Southeast Asia and Africa in the middle Eocene (Tabuce & Mariveaux, 2005). It is possible that channid dispersal to Africa was coeval with this mammalian exchange.

Because of ambiguities concerning Eocene fossils from Indo-Pakistan, it is unclear whether our estimate pertains to the channid crown or total group. We therefore compare our results to molecular estimates for both clades. Only studies that used mitochondrial data and/or calibrations based on vicariance hypotheses found origin times significantly older than the fossil-based estimate (Li *et al.*, 2006; Wang & Yang, 2011). Other studies provide relatively broad

estimates that overlap with the fossil-based one and are consistent with a dispersal-to-Africa scenario (e.g. Adamson, Hurwood & Mather, 2010; Matschiner *et al.*, 2017). Surprisingly, none of these molecular timetrees has sufficient scope to estimate the origin time of channids accurately, as they are either focused on channids with sparse outgroup sampling, or they encompass the whole teleost tree and include only few channid species. A time-calibrated phylogeny focused on Anabantaria (the clade comprising synbranchiforms and anabantiforms; Betancur-R *et al.*, 2017) would be needed to assess the timescale of anabantiform – and channid – origin and diversification properly. Because most anabantarian lineages are endemic to the Indo-Malayan region, it is possible that this clade originated in the isolated Indian subcontinent during the Late Cretaceous. An anabantarian timetree would be necessary to test this hypothesis.

(7) Percichthyidae (South American and Australian temperate perches)

Percichthyidae includes more than 20 species of perch-like freshwater fishes, distributed across Australia and southern South America (Fig. 2.8). Molecular phylogenies show that Percichthyidae *sensu* Johnson (1984) is polyphyletic, with the catadromous *Percalates* distantly related to other percichthyids (e.g. Near *et al.*, 2013; Lavoué *et al.*, 2014). Thus, we use the term Percichthyidae to contain members of the group as historically construed minus *Percalates* (i.e. *sensu* Betancur-R *et al.*, 2017). *Percalates* and percichthyids share several morphological features, to the point that *Percalates* has been synonymized to the percichthyid genus *Macquaria* in the past (MacDonald, 1978); consequently, the fossil record of percichthyids is difficult to evaluate. New morphological studies are needed to identify percichthyid synapomorphies permitting correct taxonomic identification of perch-like fossil fishes found in freshwater sediments of southern continents. In fact, various fossil specimens reported in the literature as percichthyids have been referred to the non-percichthyid *Percalates* (Hills, 1934).

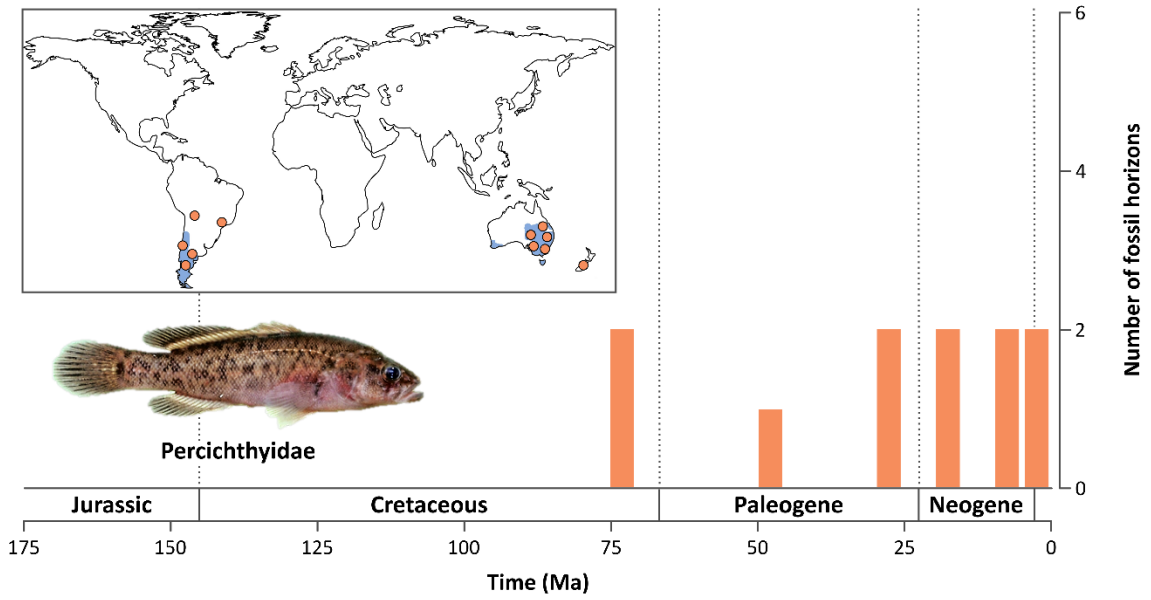

Fig. 2.8. Histogram showing the temporal distribution of distinct stratigraphic horizons bearing fossils of Percichthyidae. Each time bin is 5 million years (Myr) in width. The inset displays the present-day geographic distribution of Percichthyidae (in blue), as well as the main localities in which percichthyid fossils have been found (orange dots). Photograph of nightfish (*Bostockia porosa*) by the Australian Museum from Wikimedia Commons.

The Maastrichtian El Molino Formation of Bolivia yields the oldest putative percichthyid fossils (Table S7), including the articulated anterior half of a skeleton referred to the genus *Percichthys* (Gayet & Meunier, 1998). Other articulated percichthyid specimens have been found in deposits from the early–middle Eocene of Argentina and the Oligocene of Brazil, and in the early Miocene Río Pedregoso Formation of Chile (originally interpreted as late Paleocene in age; Arratia, 1982; Pedroza *et al.*, 2017). These fossils show a broader distribution of percichthyids in South America, where they are today restricted to the southernmost tip of the continent. Percichthyid fossils are also found in Australia, with the oldest examples being at least early-middle Miocene in age (Hills, 1946; Turner, 1982). Two scales from the early Miocene Bannockburn Formation of New Zealand show some similarities with those of percichthyids (McDowall & Lee, 2005). Although the material is too scant for precise taxonomic identification, none of the extant freshwater fishes of New Zealand shows a comparable scale morphology, suggesting the existence of an extinct lineage of perch-like fishes in New Zealand.

Berra (2007) assigned Percichthyidae to Myers' 'peripheral division' of freshwater fishes. However, this classification stemmed from the inclusion of the catadromous *Percalates* in the group. Excluding *Percalates* from Percichthyidae, extant percichthyids occur almost exclusively in freshwater environments (with a few species rarely recorded in estuaries; Arratia, 1982). Additionally, percichthyid fossils are only found in freshwater deposits. Chen *et al.* (2014) recovered an antitropical clade of temperate freshwater fishes, named Percichthyoidea, uniting the North American centrarchids and elassomatids, the East Asian sinipercids, and percichthyids. They proposed a freshwater origin for percichthyoids and a complex biogeographic history to account for its distribution. However, other studies place marine taxa (like *Enoplosus*) as deep branches within this broader clade (Near *et al.*, 2013; Betancur-R *et al.*, 2017), hinting at a marine origin followed by freshwater invasions: one in the northern hemisphere and another in the southern hemisphere, leading to percichthyids.

(*a*) *Fossil-based estimate of origin times*

Because of the relatively poor percichthyid fossil record, our fossil-based time estimate for percichthyid origin spans most of the Cretaceous, from the Barremian to the Maastrichtian (127.4–69.1 Ma; median point estimate: 87.6 Ma). Strikingly, it is significantly older than molecular clock estimates, which indicate a Paleocene–Oligocene origin for crown Percichthyidae (Near *et al.*, 2013; Chen *et al.*, 2014; the oldest known percichthyid fossils pre-date the upper bound of this range). Moreover, the South American clade including the genera *Percichthys* and *Percilia* appears to be nested within the Australian radiation (Lavoué *et al.*, 2014). This is in contrast with the early appearance of South American percichthyids, including extinct species attributed to *Percichthys*. Two hypotheses can be proposed to explain this discrepancy. First, published molecular-clock analyses underestimate the divergence times of the main lineages within Centrarchiformes (like Percichthyidae), due to the inadequate fossil calibrations. Second, the early South American fossil percichthyids may not be percichthyids at all, but rather more closely related to *Percalates* or to another lineage of perch-like fishes. Detailed anatomical studies of percichthyids and their relatives are needed to identify diagnostic characters for determining the relationship of these fossils.

Although circum-Antarctic deep water circulation was established only around 31 Ma (Lawver & Gahagan, 2003), geophysical and palaeopalynological evidence suggest that the seaway between east Antarctica and Australia formed by the beginning of the Paleocene (Woodburne & Case, 1996; Bowman *et al.*, 2012). Thus, the Maastrichtian age of the Bolivian percichthyid fossils would suggest that early percichthyids would have been able to disperse overland between South America and Australia *via* Antarctica. It is possible that the *Percichthys* + *Percilia* clade diverged from other percichthyids because of a vicariant event caused by submersion of the South Tasman Rise and the separation of Australia from Antarctica during the Paleocene.

V. HISTORICAL BIOGEOGRAPHY OF WIDESPREAD FRESHWATER FISH CLADES

(1) Biogeographic patterns and the origin of modern geographic distributions

General patterns concerning the biogeographic history of widespread freshwater fishes can be gathered from the individual study cases presented here. Continental vicariance cannot be rejected for some of these clades: lepidosireniforms, osteoglossomorphs, characiforms and percichthyids (Fig. 2.9). However, osteoglossomorphs and characiforms are probably characterized by a complex biogeographic history that involved several long-distance dispersals as well as continental vicariance and that has been partially concealed by regional extinctions. In fact, the fossil record of these two groups greatly expands their present geographic distribution, highlighting the importance of palaeontological data in reconstructing the biogeographic history of extant organisms. While the fossil record of galaxiids does not capture their early evolutionary history, molecular clock studies suggest a similar pattern of early vicariance followed by long-distance dispersals, although on a more recent timescale. Among all the extant clades examined here, crown lepidosireniforms are probably the only group whose continental geographic distribution has been driven purely by a strict vicariant event: separation of South American and African landmasses. By contrast, cyprinodontiforms and channids are likely much younger than any major continental breakup that might have affected their geographic distribution. Thus, their intercontinental distribution is probably the result of multiple dispersal events, either overland (channids) or transoceanic (cyprinodontiforms).

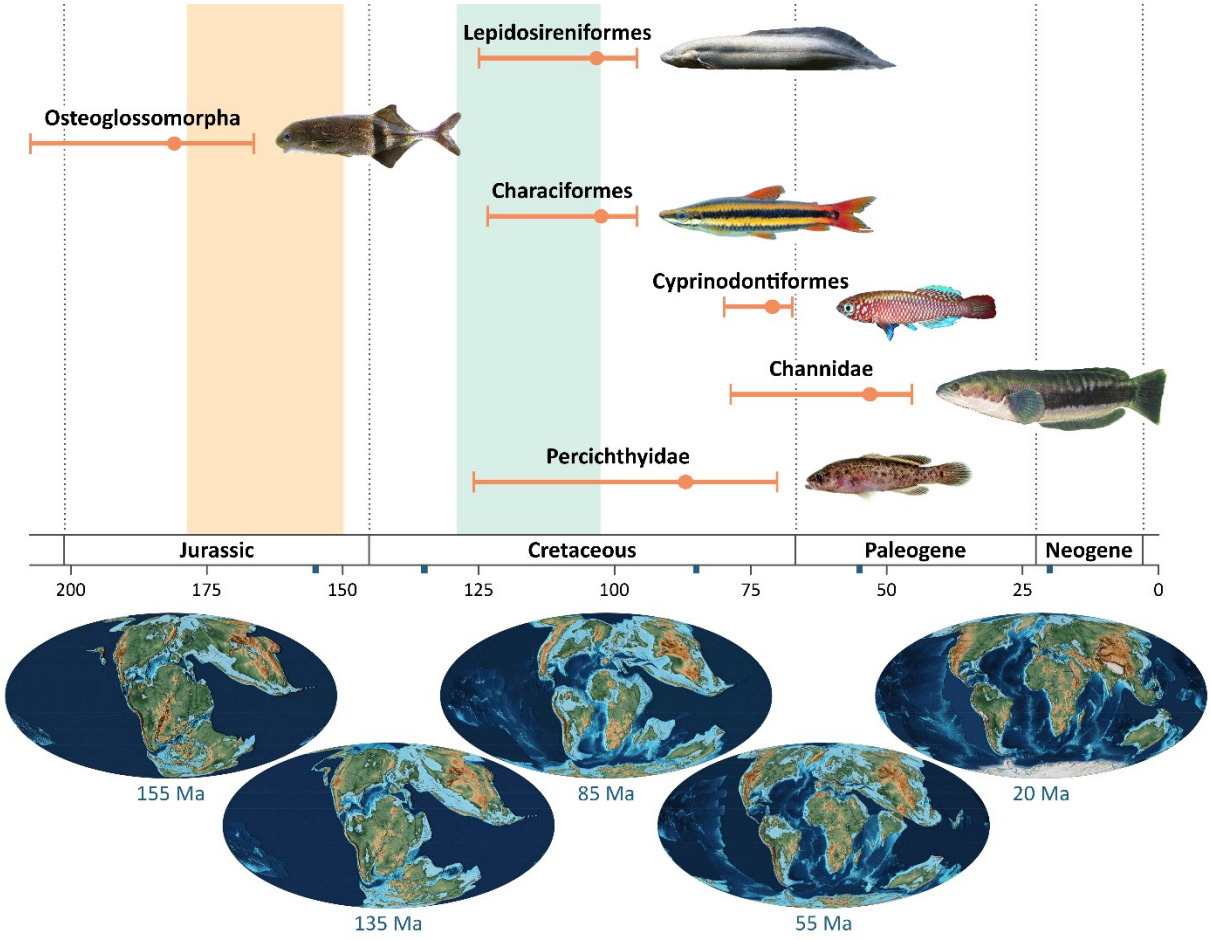

Fig. 2.9. Fossil-derived timescale for the origin of the focal clades considered in this review. Galaxiidae is not included because its estimate was not informative (see Section IV.4). The timescale for characiform origin shown here and in Fig. 2.10 is the older estimate from this study (i.e. including Cenomanian occurrences; see Section IV.3). The dot indicates the median point estimate, while the bar encompasses the range of estimates when accounting for both non-uniform distribution of the fossil record and uncertainty in the age of fossil horizons. As in Fig. 2.1, coloured bands indicate the timeframe of the Western–Eastern Gondwana breakup (in light ocre) and the South America–Africa breakup (in light green). The horizontal axis represents time, with scale provided in million years ago (Ma). Paleogeographic maps are taken from Scotese (2014). Blue boxes refer to the age of the palaeogeographic reconstructions relative to the timescale.

There is no doubt that the progressive breakup of Gondwana had a massive impact on the geographic distribution of terrestrial and freshwater organisms living at the time of these geologic events. However, it seems that, at least for freshwater fishes, the pre-existing background of vicariance-driven distributions has been progressively eroded through time by

extinctions and intercontinental dispersals. In fact, while the separation of South America and Africa corresponds to several vicariant events that can be inferred from the Aptian–Cenomanian fossil record of these continents (involving mawsoniids, lepisosteoids, amiids, cladocyclids and chanids; see Section II.2), lungfishes are the only freshwater fishes inhabiting both continents today for which the same process can be confidently identified as the primary cause of their present disjunct distribution. Together, the evidence presented here strongly suggests that rare intercontinental dispersals can have a significant effect on biogeographic patterns across continents. The relevance of long-distance dispersals in freshwater fish biogeography highlighted here parallels a growing literature supporting a prominent role of these events in the biogeographic history of a wide variety of terrestrial and freshwater organisms (de Queiroz, 2005; Gamble *et al.*, 2011; Pyron, 2014; Rota, Peña & Miller, 2016; Scheben *et al.*, 2016).

(2) Oceanic dispersal in freshwater fishes

While in some cases marine intercontinental dispersal of freshwater organisms could be explained by marine ancestry (e.g. osteoglossids), there is no evidence for past adaptations to open marine environments in several freshwater clades for which an oceanic dispersal event likely happened (e.g. cichlids, killifishes, synbranchids). The exact mechanisms by which transoceanic dispersal of freshwater fishes could happen are difficult to evaluate because this kind of dispersal is rare and relatively improbable (although it becomes almost inevitable over geological timescales). Proposed mechanisms (not mutually exclusive) include formation of giant freshwater plumes following catastrophic events like typhoons or tropical river floods; rafting of large chunks of soil and vegetation [see Houle (1998) for dispersal of terrestrial vertebrates, but these 'floating islands' might include puddles of fresh water as well]; 'stepping-stone' dispersal across island arches (Gilpin, 1980; however, this mechanism may be unfeasable for freshwater organisms); or bird-mediated zoochory of fish eggs (Hirsch *et al.*, in press). Strikingly, most freshwater fish taxa for which transoceanic dispersal has been inferred possess peculiar physiological or behavioural adaptations (e.g. high salinity tolerance, drought-resistant eggs, air-breathing and amphibious lifestyle) that might have increased their chance of surviving such an improbable journey. A similar pattern is also seen in terrestrial vertebrates for which sweepstakes dispersal has been inferred. For example, small body size, arboreal habits and

heterothermy are common features of mammals that survived transoceanic journeys (Kappeler, 2000; Nowack & Dausmann, 2015), while drought- and salinity-resistant eggs and adhesive fingers are probably some of the adaptations that allowed geckos to disperse multiple times across oceans and to colonize oceanic islands (Gamble *et al.*, 2011). In this sense, while long-distance dispersals have a stochastic nature, we would expect a strong phylogenetic component for these events, which should be clustered within clades possessing those traits mentioned above. Among freshwater fishes examined here, the only exception to this general pattern seems to be represented by the poorly studied polycentrid leaffishes (see Section II.1), thus encouraging further investigation of this clade's natural history.

(3) Congruence and discrepancy between the fossil record and molecular divergence-time estimates

The fossil-based age estimates inferred herein for several clades of widespread freshwater fishes are generally congruent with molecular timescales published in the last 10 years (Fig. 2.10). This is a striking result, as these two different approaches draw upon semi-independent data: although time calibration of molecular phylogenies commonly employs fossil data, these are usually limited to a very small subset of the known fossil record of a clade (Parham *et al.*, 2012). Moreover, molecular timescales of some taxa are often estimated using exclusively external fossil calibrations – that is, fossils belonging to other, closely related taxa. As a result, there is very minor overlap between the data informing our fossil-based age estimates and the data informing evolutionary timescales in molecular phylogenies. Yet, for several taxa (Lepidosireniformes, Osteoglossidae, Characiformes, Cyprinodontiformes, Cyprinodontoidei, Channidae), the fossil-based timescales inferred in this study are not significantly different from published molecular ones, providing support for the evolutionary timescales presented here.

Deviations are worth discussing, as they might highlight problematic issues in either of these approaches for estimating evolutionary timescales. The origin of Percichthyidae estimated here is significantly older than corresponding molecular estimates; this may be due to the misidentification of some articulated specimens from the Maastrichtian El Molino Formation as belonging to the genus *Percichthys* (see Section IV.7). The most striking discrepancy is represented by the age that we derived for total-group Osteoglossomorpha (latest Triassic–

Middle Jurassic), which is significantly younger than most recent molecular estimates. This relates to a broader discrepancy between the oldest crown teleost fossils (Middle–Late Jurassic) and the age of crown teleosts inferred by molecular clock studies: the so-called 'teleost gap' (Near *et al.*, 2012).

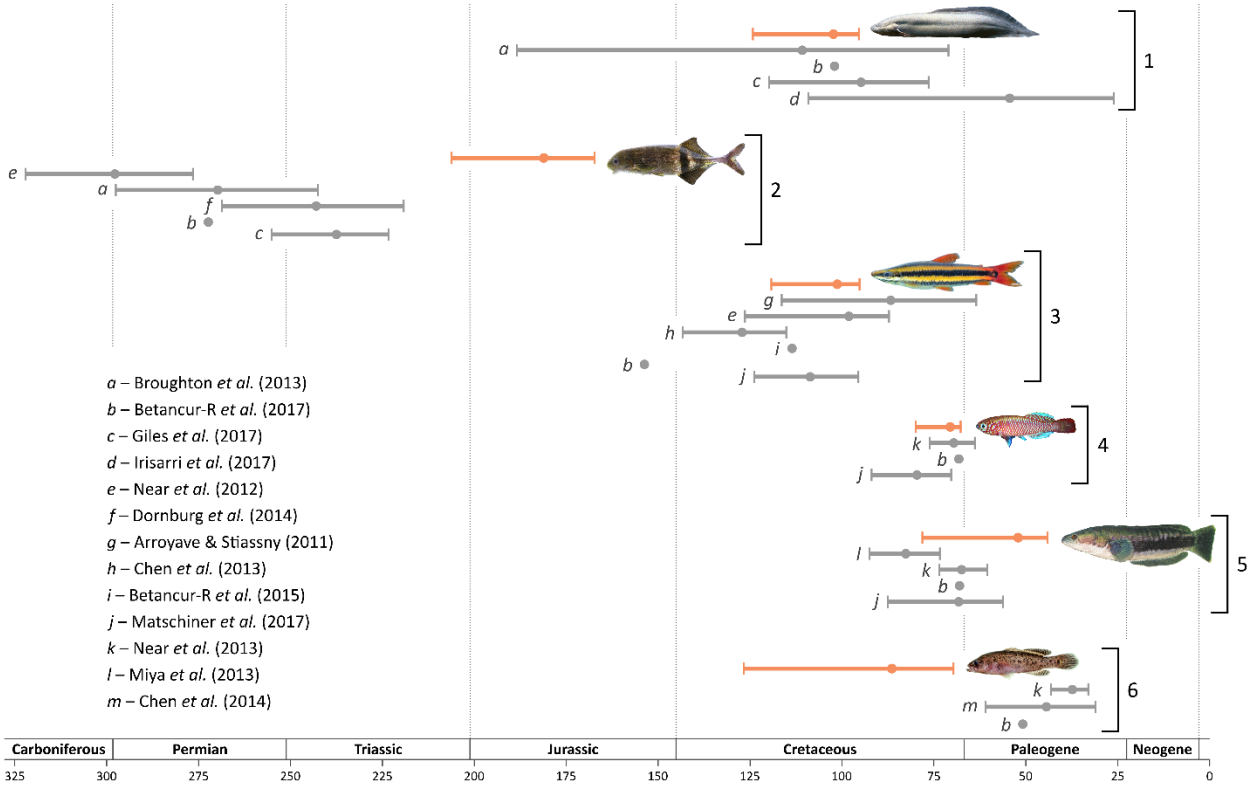

Fig. 2.10. Comparison between fossil-derived estimates (in orange) and recently published molecular estimates (in grey) for the origin times of: 1, Lepidosireniformes; 2, total-group Osteoglossomorpha; 3, Characiformes; 4, Cyprinodontiformes; 5, Channidae; 6, Percichthyidae. Molecular estimates for channids refer to stem Channidae (see Section IV.6). Galaxiidae is not included because its fossil-based estimate was not informative (see Section IV.4). The dot indicates the point estimate, while the bar (when present) encompasses 95% confidence or credibility interval. The horizontal axis represents time, with scale provided in million years ago (Ma).

While the use of rapidly evolving molecular markers and misidentified fossil calibrations can yield unrealistically old estimates for the crown teleost radiation, correcting for these factors still results in an inferred Permo-Triassic origin of crown teleosts (Dornburg *et al.*, 2014; Giles *et al.*, 2017). The wealth of stem teleosts found in Middle Triassic–Middle Jurassic formations (Arratia, 2015; López-Arbarello & Sferco, 2018) suggests that it should not be impossible (at least theoretically) to find crown teleost fossils in deposits of that age. Incompleteness of the fossil

record can only partially account for this gap. According to our fossil-based estimates, even when accounting for non-uniform fossil preservation potential through time it would be very unlikely to find any stem osteoglossomorph fossils older than 207 Ma. However, it should be noted that this estimate is based on the temporal distribution of non-marine deposits, which is likely not appropriate when trying to derive age estimates for the earliest divergences in the teleost tree, as the early evolutionary history of teleosts probably occurred in marine environments (Betancur-R *et al.*, 2015; Guinot & Cavin, 2018). In summary, the gap between the earliest molecular divergence estimates within crown teleosts and the oldest crown teleost fossils can be only partially explained by an incomplete fossil record or by failure to recognize crown teleosts among known Triassic fossils. It is possible that high heterogeneity in the rates of molecular evolution at the base of the teleost radiation or biased effective calibration prior densities are responsible for pushing molecular estimates towards older dates, but more studies about the impact of prior specification on the molecular timescale of early teleost evolution are needed to test these hypotheses.

(4) Limitations of the stratigraphic approach to infer origin times and test biogeographic hypotheses

The stratigraphic approach utilized here presents several limitations. Firstly, at least 15–20 distinct fossil horizons are needed in order to obtain an informative range of age estimates, meaning a range that is precise enough not to encompass several geologic periods and to provide some insight on evolutionary timescales. Several clades have a very limited fossil record and are often concentrated in a few distinct fossil horizons, as in the cases of galaxiids and percichthyids. Additionally, many of the estimates derived here rely heavily on the correct taxonomic identification of the oldest known representatives of a clade. This can be particularly problematic when the oldest putative fossils of a clade are very fragmentary (e.g. teeth, scales or isolated otoliths, as in lepidosireniforms, osteoglossomorphs and characiforms) or when, even with better preserved articulated fossils, their phylogenetic affinities are dubious (e.g. El Molino cyprinodontiforms and percichthyids, or of the putative osteoglossid *Chanopsis*). While a possible solution to the former could be to restrict the analysis to articulated fossils only, with the preservation potential function based upon fossil horizons that can yield articulated specimens

(Friedman *et al.*, 2013), this approach ignores considerable information coming from microfossil assemblages and, more importantly, drastically lowers the number of distinct stratigraphic horizons from which the focal clade is known.

Another possible issue stems from the phylogenetic interpretation of the results of this type of analysis – in other words, the phylogenetic node to which an age estimate pertains. We believe that it is more appropriate to refer this estimate to the least inclusive clade containing all the fossils considered in the analysis (see Section II.3 for an example involving cichlids).

While we used the estimated origin time of widespread freshwater fish clades as a test of a simple vicariant scenario for each of these clades, it is clear that our approach is very limited in scope and can only test whether the evolutionary timescale of the group of interest is compatible with the timescale of relevant continental breakups. Model-based biogeographic analyses that include fossil taxa in a phylogenetic framework, allow for heterogeneity in dispersal rates through time, and constrain vicariant events to the known timescales of underlying geologic events are needed to reconstruct the biogeographic history of these clades in more detail. While significant progress has been made towards the development of complex biogeographic models [Ronquist & Sanmartín, 2011; Matzke, 2014; but see Ree & Sanmartín, 2018 for a critique of the Dispersal-Extinction-Cladogenesis + Jump dispersal (DEC+J) model], two major challenges remain: the inclusion of fossil taxa in a 'total-evidence' phylogeny (Ronquist, Lartillot & Phillips, 2016), which requires the collection of morphological data for both extant and extinct taxa – a complex and time-consuming task that requires high levels of taxon-specific expertise; and the lack of models accounting for taphonomic biases and the incomplete nature of the fossil record in phylogeny-based biogeographic reconstruction software. It is worth noting that fragmentary fossil specimens that can be assigned to broad clades but are not sufficiently diagnostic to permit finer taxonomic resolution can often provide invaluable geographic and temporal information. These specimens have very few informative morphological characters, so they will likely be ignored in any phylogeny-based biogeographic reconstruction [although see Silvestro *et al.* (2016) for a way to estimate biogeographic parameters using fossil data without phylogenies, and Cau (2017) for an approach towards specimen-level phylogenetics in palaeontology). Consequently, even a qualitative assessment of the geographic and temporal

distribution of fossils belonging to a certain clade – including fragmentary specimens – has the potential to greatly improve our understanding of its biogeographic history.

(5) Future directions

Stressing the importance of the fossil record in biogeographic reconstruction, we hope that further attention will be directed towards ways of integrating fossil data into analytically explicit biogeographic reconstructions. Ultimately, a better understanding of the early biogeographic history of freshwater fishes will come from detailed morphological studies able to solve the systematics of some key fossil taxa. For example, the Maastrichtian El Molino Formation in Bolivia records the first occurrence of several freshwater fish lineages that still occur in South American freshwater environments (Gayet, 1991; Gayet *et al.*, 2001), and thus represents one of the oldest fossil fish assemblages with a modern taxonomic composition in southern landmasses. Moreover, it is one of the very few freshwater fish communities known from around the Cretaceous–Paleogene boundary in the southern hemisphere. Yet, despite the biogeographic and palaeoenvironmental importance of these fossils, the systematic position of the El Molino fishes (including those for which articulated specimens are known) is still highly uncertain.

Time-calibrated phylogenetic trees based mainly (if not exclusively) on molecular data will remain, for the foreseeable future, the primary way to derive evolutionary timescales for a group of organisms and thus test alternative biogeographic hypotheses. Accuracy and precision of molecular timescales strongly depend on the choices made for time calibration (Duchêne, Lanfear & Ho, 2014). The fossil-based estimates derived here for the origin of widespread freshwater fish taxa could be used in future studies as calibration priors for the relevant nodes, with the advantage that soft maximum bounds were objectively inferred from the temporal distribution of the fossil record and not arbitrarily decided (as often happens in node calibrations; Bromham *et al*., in press). The use of analytically derived calibration distributions removes a layer of subjectivity in the process of molecular dating and can potentially yield timescales that better reflect what we know from the palaeontological record [see also Hedman (2010) and Matschiner *et al.* (2017) for different approaches to deriving fossil-based origin time distributions].

Comparing separate molecular evolutionary timescales across freshwater fish taxa can be problematic because available analyses are usually focused on specific clades. These commonly differ in the methods employed, in the kind of data analysed and in prior assumptions – which, in the case of Bayesian dating, include priors on distribution of node times, branch-rates and calibration distributions, among others. Thus, it might be expected that different studies do not show comparable timescales, making the task of building a comprehensive timescale of biogeographic evolution in freshwater fishes particularly challenging. While substantial progress has been made towards the reconstruction of a fish timetree encompassing every major fish lineage (Near *et al.*, 2012; Betancur-R *et al.*, 2017), these studies are not targeted towards the reconstruction of intercontinental biogeographic patterns and so they lack several key taxa and internal nodes. A possible solution could be to perform a 'fish-wide' time-calibrated phylogenetic analysis that specifically targets every biogeographically relevant freshwater taxon, in order to derive a unified timescale of continental-scale biogeographic events across freshwater fishes.

Finally, among freshwater fishes, descendants of past long-distance dispersals play a fundamental role in freshwater communities and can be subject to spectacular radiations, as in the cases of cichlids in the Neotropics, galaxiids in New Zealand and killifishes in Africa. Several recent studies suggest that ecological opportunity through invasion of new adaptive zones – including colonization of new geographic areas – can influence diversification patterns (e.g. Burbrink & Pyron, 2010; Burress & Tan, 2017). However, the impact of long-distance dispersal events on macroevolutionary dynamics – including diversification rates and modes – and continental-scale biotic assemblages is still largely unexplored.

VI. CONCLUSIONS

(1) Vicariance and dispersal both played crucial roles in structuring the distribution of modern freshwater fishes. However, even when clades are old enough to have experienced continental vicariance, the pre-existing vicariance-driven distribution is often confounded and eroded through time by successive dispersals and regional extinctions during the Late Cretaceous and Cenozoic. The only known examples of present-day disjunct intercontinental distributions

consistent with pure vicariance are South American and African lungfishes (Lepidosireniformes) and, possibly, Southern temperate perches (Percichthyidae).

(2) The evidence presented here shows that oceanic long-distance dispersal likely happened in several freshwater fish taxa. This complements recent studies stressing the importance of long-distance dispersal in terrestrial lineages. However, the means by which oceanic dispersal by freshwater fishes is achieved, and the impact of these rare events on macroevolutionary dynamics are still relatively unknown and could represent important future areas of investigation in biogeographic research.

(3) Fossils can provide invaluable temporal, geographic and environmental information that can be used to reconstruct the biogeographic history of a clade. Specifically, fossil data can expand the present geographic distribution of a clade and reveal past dispersal or vicariant events that have been obscured by regional extinction. Moreover, fossils can show that extinct members of a clade had environmental tolerances differing from modern species. For example, while all living osteoglossomorphs are restricted to freshwater habitats, several fossil osteoglossomorphs were found in marine deposits of Paleocene–early Eocene age all over the world, suggesting a substantial role of marine dispersal in the past (if not present) geographic distribution of the group.

(4) Methods to infer origin times using the temporal distribution of the known fossil record of a clade complement time-calibrated molecular phylogenies as means to establish evolutionary timescales. Fossil-based estimates can be compared with molecular estimates and, when conflicts between the two arise, can point out problematic issues in either evaluation of the fossil record or the methods used to infer molecular timetrees. Fossil-based age ranges can be also used to calibrate relevant nodes on molecular phylogenies, avoiding the necessity to specify user-defined, subjective calibration parameters.

VII. ACKNOWLEDGEMENTS

The core of this review started as a class project for the University of Michigan “Data Analysis and Model Estimation” course by Eric Hetland, whom we wish to thank. We thank the contributors to the Paleobiology Database, without which this work would not have been

possible. We also thank F. Mees (Royal Museum of Central Africa, Tervuren) for allowing us to examine the holotype of *Paradercetis kipalaensis*, C. Abraczinskas (University of Michigan Museum of Paleontology) for improving the figures and Lionel Cavin and an anonymous reviewer for valuable comments that improved the manuscript. This work was supported by funding from by the Department of Earth and Environmental Sciences of the University of Michigan (Scott Turner Student Research Grant Award 2017, to A.C.) and by the Society of Systematic Biologists (2017 SSB Graduate Student Research Award, to A.C.).

VIII. REFERENCES

References marked with asterisk have been cited only within supporting information.

*Adaci, M., Tabuce, R., Mebrouk, F., Bensalah, M., Fabre, P. H., Hautier, L., Jaeger, J. J., Lazzari, V., Mahboubi, M. H., Marivaux, L. & Otero, O. (2007). Nouveaux sites à vertébrés paléogènes dans la région des Gour Lazib (Sahara Nord–occidental, Algérie). *Comptes Rendus Palevol* **6**, 535–544.

Adamson, E. A., Hurwood, D. A. & Mather, P. B. (2010). A reappraisal of the evolution of Asian snakehead fishes (Pisces, Channidae) using molecular data from multiple genes and fossil calibration. *Molecular Phylogenetics and Evolution* **56**, 707–717.

*Agnolin, F. L., Bogan, S., Tomassini, R. L. & Manera, T. (2014). Nuevo Percichthyidae (Teleostei, Percoidei) del Plioceno temprano de la provincia de Buenos Aires (Argentina) y sus implicancias biogeográficas. *Revista del Museo Argentino de Ciencias Naturales* **16**, 19–31.

*Aguilera, O. & Lundberg, J. (2010). Venezuelan Caribbean and Orinocoan Neogene fish. In: Sanchez–Villagra, M., Aguilera, O. & Carlini, F. (eds.) *Urumaco and Venezuelan Paleontology*, 129–152. Indiana University Press, Bloomington.

*Akhmetiev, M. A., Lopatin, A. V., Sytchevskaya, E. K. & Popov, S. V. (2005). Biogeography of the northern Peri-Tethys from the late Eocene to the early Miocene: Part 4. Late Oligocene–Early Miocene: Terrestrial Biogeography, Conclusions. *Palaeontological Journal* **39**, *Supplement 1*, S1–S54.

Alfaro, M.E., Faircloth, B.C., Harrington, R.C., Sorenson, L., Friedman, M., Thacker, C.E., Oliveros, C.H., Černý, D. & Near, T.J. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nature Ecology & Evolution* **2**, 688–696.

Ali, J. R. & Aitchison, J. C. (2008). Gondwana to Asia: plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews* **88**, 145–166.

Allibone, R. M. & Wallis, G. P. (1993). Genetic variation and diadromy in some native New Zealand galaxiids (Teleostei: Galaxiidae). *Biological Journal of the Linnean Society* **50**, 19–33.

Altner, M. & Reichenbacher, B. (2015). †Kenyaichthyidae fam. nov. and †*Kenyaichthys* gen. nov.–First Record of a Fossil Aplocheiloid Killifish (Teleostei, Cyprinodontiformes). *PLoS One* **10**, e0123056.

*Alvarado-Ortega, J., Cuevas-García, M., del Pilar Melgarejo-Damián, M., Cantalice, K. M., Alaniz-Galvan, A., Solano-Templos, G. & Than-Marchese, B. A. (2015). Paleocene fishes from Palenque, Chiapas, southeastern Mexico. *Palaeontologia Electronica* **18**, 1–22.

*Alvarez, J. & Aguilar, F. (1957). Contribución al estudio de la suspensión gonopódica del género *Poeciliopsis* con una descripción de una nueva especie fósil procedente de El Salvador, Centro América. *Revista de la Sociedad Méxicana de Historia Natural* **18**, 153–171.

*Alvarez, J. & Arriola-Longoria, J. (1972). Primer goodeido fosil procedente del Plioceno jalisciense (Pisces, Teleostomi). *Boletin de la Sociedad de Ciencias Naturales de Jalisco*, *6*, 6–15.

*Alvarez, J. & Moncayo, M.E. (1976). Contribución a la Paleoictiología de la Cuenca de México. *Anales del Instituto Nacional de Antropología e Historia* **7**, 191–242.

Amaral, C. R. & Brito, P. M. (2012). A new Chanidae (Ostariophysii: Gonorynchiformes) from the Cretaceous of Brazil with affinities to Laurasian gonorynchiforms from Spain. *PLoS One* **7**, e37247.

Anderson, M. E. (1998). A late Cretaceous (Maastrichtian) galaxiid fish from South Africa. *Special Publication*, *J.L.B. Smith Institute of Ichthyology* **60**, 1–8.

*Antoine, P. O., Abello, M. A., Adnet, S., Sierra, A. J. A., Baby, P., Billet, G., Boivin, M., Calderon, Y., Candela, A., Chabain, J. & Corfu, F. (2016). A 60–million–year Cenozoic history of western Amazonian ecosystems in Contamana, eastern Peru. *Gondwana Research* **31**, 30–59.

*Antunes, M. T., Balbino, A. & Gaudant, J. (1995). Découverte du plus récent poisson Characiforme européen dans le Miocène terminal du Portugal. *Communications de l'Institut géologique des Mines* **81**, 79–84.

*Arambourg, C. (1952). Les vertébrés fossiles des gisements de phosphates (Maroc-Algérie-Tunisie). *Notes et Mémoires du Service géologique du Maroc* **92**, 1–372.

Arcila, D., Ortí, G., Vari, R., Armbruster, J. W., Stiassny, M. L., Ko, K. D., Sabaj, M. H., Lundberg, J., Revell, L. J. & Betancur-R, R. (2017). Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology & Evolution* **1**, 0020.

*Argyriou, T. (2014). Description, paleoenvironmental and paleobiogeographical implications of Miocene fish faunas from Jabal Zaltan and Sahabi (Libya). MSc thesis, University of Alberta, 310 pp.

*Argyriou, T., Cook, T. D., Muftah, A. M., Pavlakis, P., Boaz, N. T. & Murray, A. M. (2015). A fish assemblage from an early Miocene horizon from Jabal Zaltan, Libya. *Journal of African Earth Sciences* **102**, 86–101.

Arratia, G. (1982). A review of freshwater percoids from South America. *Abhandlungen der Senckenbergischen Naturforschenden Gesellschaft* **540**, 1–52.

Arratia, G. (1999). The monophyly of Teleostei and stem-group teleosts. In: Arratia, G. & Schultze, H.-P. (eds.), *Mesozoic Fishes 2 – Systematics and Fossil Record*, 265–334. Verlag Dr. F. Pfeil, München.

Arratia, G. (2010). Critical analysis of the impact of fossils on teleostean phylogenies, especially that of basal teleosts. In: Elliott, D. K., Maisey, J. G., Yu, X. & Miao, D. (eds.), *Morphology, Phylogeny and Paleobiogeography of Fossil Fishes*, 247–274. Verlag Dr. F. Pfeil, München.

Arratia, G. (2015). Complexities of early Teleostei and the evolution of particular morphological structures through time. *Copeia* **103**, 999–1025.

Arratia, G. & Cione, A. (1996). The record of fossil fishes of southern South America. *Münchner Geowissenschaftliche Abhandlungen (A)* **30**, 9–72.

Arroyave, J. & Stiassny, M. L. (2011). Phylogenetic relationships and the temporal context for the diversification of African characins of the family Alestidae (Ostariophysi: Characiformes): Evidence from DNA sequence data. *Molecular Phylogenetics and Evolution* **60**, 385–397.

Bannikov, A. F. & Carnevale, G. (2016). *Carlomonnius quasigobius* gen. et sp. nov.: the first gobioid fish from the Eocene of Monte Bolca, Italy. *Bulletin of Geosciences* **91**, 13–22.

*Bannikov, A. F. & Parin, N. N. (1997). The list of marine fishes from Cenozoic (Upper Paleocene-Middle Miocene) localities in southern European Russia and adjacent countries. *Journal of Ichthyology* **37**, 133–146.

Bean, L. B. (2017). Reappraisal of Mesozoic fishes and associated invertebrates and flora from Talbragar and Koonwarra, eastern Australia. *Proceedings of the Royal Society of Victoria* **129**, 7–20.

*Bennett, D. K. (1979). Three late Cenozoic fish faunas from Nebraska. *Transactions of the Kansas Academy of Science* **82**, 146–177.

Berra, T. M. (2007). *Freshwater Fish Distribution*. The University of Chicago Press, Chicago.

Berrell, R. W., Alvarado-Ortega, J., Yabumoto, Y. & Salisbury, S. W. (2014). The first record of the ichthyodectiform fish *Cladocyclus* from eastern Gondwana: a new species from the Lower Cretaceous of Queensland, Australia. *Acta Palaeontologica Polonica* **59**, 903–920.

*Bertini, R. J., Marshall, L. G., Gayet, M. & Brito, P. (1993). Vertebrate faunas from the Adamantina and Marília formations (Upper Bauru Group, Late Cretaceous, Brazil) in their stratigraphic and paleobiogeographic context. *Neues Jahrbuch für Geologie und Paläontologie, Abhandlungen* **188**, 71–101.

Betancur-R, R. (2010). Molecular phylogenetics supports multiple evolutionary transitions from marine to freshwater habitats in ariid catfishes. *Molecular Phylogenetics and Evolution* **55**, 249–258.

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., Lopez, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Cureton Ii, J. C., Zhang, F., Buser, T., Campbell, M. A., Ballesteros, J. A., Roa-Varon, A. *et al.* (2013). The tree of life and a new classification of bony fishes. *PLoS Currents: Tree of Life*. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, R., Ortí, G. & Pyron, R. A. (2015). Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. *Ecology Letters* **18**, 441–450.

Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G. & Orti, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology* **17**, 162.

*Blanco, A., Szabó, M., Blanco–Lapaz, À. & Marmi, J. (2017). Late Cretaceous (Maastrichtian) Chondrichthyes and Osteichthyes from northeastern Iberia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **465**, 278–294.

*Boeseman, M. (1949). On pleistocene remains of *Ophicephalus* from Java, in the "Collection Dubois". *Zoologische Mededenlingen* **30**, 83–94.

*Bogachev, V. V. (1936). New data on the Transcaucasian Miocene. *Proceedings of Azerbaijan Scientific− Research Oil Institute* **31**, 1–34.

*Bogan, S., de los Reyes, M. L. & Cenizo, M. M. (2009). Primer registro del género *Jenynsia* Günther, 1866 (Teleostei: Cyprinodontiformes) en el Pleistoceno Medio tardío de la provincia de Buenos Aires (Argentina). *Papéis Avulsos de Zoologia (São Paulo)* **49**, 81–86.

*Bogan, S., Sidlauskas, B., Vari, R. P. & Agnolin, F. (2012). *Arrhinolemur scalabrinii* Ameghino, 1898, of the late Miocene: a taxonomic journey from the Mammalia to the Anostomidae (Ostariophysi: Characiformes). *Neotropical Ichthyology* **10**, 555–560.

Böhme, M. (2004). Migration history of air-breathing fishes reveals Neogene atmospheric circulation patterns. *Geology* **32**, 393–396.

*Böhme, M. (2010). Ectothermic vertebrates (Actinopterygii, Allocaudata, Urodela, Anura, Crocodylia, Squamata) from the Miocene of Sandelzhausen (Germany, Bavaria) and their implications for environment reconstruction and palaeoclimate. *Paläontologische Zeitschrift* **84**, 3–41.

Bonde, N. (2008). Osteoglossomorphs of the marine Lower Eocene of Denmark–with remarks on other Eocene taxa and their importance for palaeobiogeography. *Geological Society*, *London*, *Special Publications* **295**, 253–310.

Bowman, V. C., Francis, J. E., Riding, J. B., Hunter, S. J. & Haywood, A. M. (2012). A latest Cretaceous to earliest Paleogene dinoflagellate cyst zonation from Antarctica, and implications for phytoprovincialism in the high southern latitudes. *Review of Palaeobotany and Palynology* **171**, 40–56.

Briggs, J. C. (1995). *Global Biogeography*. Elsevier, Amsterdam.

Briggs, J. C. (2003). Fishes and birds: Gondwana life rafts reconsidered. *Systematic Biology* **52**, 548–553.

Brikiatis, L. (2014). The De Geer, Thulean and Beringia routes: key concepts for understanding early Cenozoic biogeography. *Journal of Biogeography* **41**, 1036–1054.

Brikiatis, L. (2016). Late Mesozoic North Atlantic land bridges. *Earth-Science Reviews* **159**, 47–57.

*Brinkman, D. B. & Neuman, A. G. (2002). Teleost centra from uppermost Judith River Group (Dinosaur Park Formation, Campanian) of Alberta, Canada. *Journal of Paleontology* **76**, 138–155.

Brinkman, D. B., Newbrey, M. G. & Neuman, A. G. (2014). Diversity and paleoecology of actinopterygian fish from vertebrate microfossil localities of the Maastrichtian Hell Creek Formation of Montana. *Geological Society of America Special Papers* **503**, 247–270.

Brinkman, D. B., Newbrey, M. G., Neuman, A. G. & Eaton, J. G. (2013). Freshwater osteichthyes from the Cenomanian to late Campanian of Grand Staircase-Escalante National Monument, Utah. In: Titus, A. L. & Loewen, M. A. (eds.) *At the Top of the Grand Staircase: The Late Cretaceous of Southern Utah*, 195–236. Indiana University Press, Bloomington.

*Bristow, C. R. (1973). *Guide to the geology of the Cuenca Basin, southern Ecuador*. Ecuadorian Geological and Geophysical Society.

Brito, P.M., Alvarado-Ortega, J. & Meunier, F.J. (2017). Earliest known lepisosteoid extends the range of anatomically modern gars to the Late Jurassic. *Scientific Reports* **7**, 17830.

*Brito, P. M. & Richter, M. (2016). The contribution of Sir Arthur Smith Woodward to the palaeoichthyology of Brazil–Smith Woodward's types from Brazil. *Geological Society, London, Special Publications* **430**, 201–217.

Brito, P. M., Yabumoto, Y. & Grande, L. (2008). New amiid fish (Halecomorphi: Amiiformes) from the Lower Cretaceous Crato formation, Araripe Basin, northeast Brazil. *Journal of Vertebrate Paleontology* **28**, 1007–1014.

Bromham, L., Duchêne, S., Hua, X., Ritchie, A. M., Duchêne, D. A. & Ho, S. Y. (in press). Bayesian molecular dating: opening up the black box. *Biological Reviews*.

*Brooks, A. S., Helgren, D. M., Cramer, J. S. & Franklin, A. (1995). Dating and context of three Middle Stone Age sites with bone points in the Upper Semliki Valley, Zaire. *Science* **268**, 548–553.

Broughton, R. E., Betancur-R, R., Li, C., Arratia, G. & Ortí, G. (2013). Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLOS Currents: Tree of Life*. doi: 10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e.

*Brzobohatý, R. (1969). Die Fischfauna des südmährischen Untermiozäns. *Folia Facultatis Scientiarum Naturalium Universitatis Purkynianae Brunensis* **10**, 1–49.

*Brzobohatý, R. (1992). Otolithen aus dem Obermiozän, Pontien, des Wiener Beckens (Götzendorf und Stixneusiedl, NÖ). *Annalen des Naturhistorischen Museums in Wien. Serie A für Mineralogie und Petrographie, Geologie und Paläontologie, Anthropologie und Prähistorie* **94**, 1–6.

*Brzobohatý, R. & Stancu, J. (1974). Die Fischfauna des Sarmatien s. str. In: Papp, A., Marinescu, F. & Senes, J. (eds.) *Chronostratigraphie und Neostratotypen, Miozän der Zentralen Paratethys. Sarmatien.–4 (M5)*, 492–515. Slowak. Akad. Wiss., Bratislava.

*Buffetaut, E., Bussert, R. & Brinkmann, W. (1990). A new nonmarine vertebrate fauna in the Upper Cretaceous of northern Sudan. *Berliner Geowissenschaftlische Abhandlungen* **120**, 183–202.

Burbrink, F. T. & Pyron, R. A. (2010). How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in New World ratsnakes (tribe Lampropeltini)? *Evolution* **64**, 934–943.

Burress, E. D. & Tan, M. (2017). Ecological opportunity alters the timing and shape of adaptive radiation. *Evolution* **71**, 2650–2660.

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

Calcagnotto, D., Schaefer, S. A. & DeSalle, R. (2005). Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. *Molecular Phylogenetics and Evolution* **36**, 135–153.

Campanella, D., Hughes, L. C., Unmack, P. J., Bloom, D. D., Piller, K. R. & Ortí, G. (2015). Multi-locus fossil-calibrated phylogeny of Atheriniformes (Teleostei, Ovalentaria). *Molecular Phylogenetics and Evolution* **86**, 8–23.

Candeiro, C. R. A., Fanti, F., Therrien, F. & Lamanna, M. C. (2011). Continental fossil vertebrates from the mid-Cretaceous (Albian–Cenomanian) Alcântara Formation, Brazil, and their relationship with contemporaneous faunas from North Africa. *Journal of African Earth Sciences* **60**, 79–92.

*Cao, Z., Xing, L. & Yu, Q. (1985). The magnetostratigraphic age and boundaries of the Yushe Formation. *Bulletin of the Institute of Geomechanics, Chinese Academy of Geological Science* **6**, 144–154.

*Cappetta, H. (1972). Les poissons crétacés et tertiaires du bassin des Iullemmeden (République du Niger). *Palaeovertebrata* **5**, 179–251.

*Cappetta, H., Jaeger, J. J., Sabatier, M., Sige, B., Sudre, J. & Vianey–Liaud, M. (1978). Decouverte dans le Paleocene du Maroc des plus anciens mammiferes eutheriens d'Afrique. *Geobios* **11**, 257–263.

*Carnevale, G., Landini, W. & Sarti, G. (2006). Mare versus Lago-mare: marine fishes and the Mediterranean environment at the end of the Messinian Salinity Crisis. *Journal of the Geological Society* **163**, 75–80.

*Case, G. R. (1994). Fossil fish remains from the late Paleocene Tuscahoma and early Eocene Bashi formations of Meridian, Lauderdale County, Mississippi. Part II. Teleosteans. *Palaeontographica Abteilung A* **230**, 139–153.

*Casier, E. (1966). *Faune icthyologique du London Clay: Atlas* (Vol. 2). British Museum (Natural History), London.

Cau, A. (2017). Specimen-level phylogenetics in paleontology using the Fossilized Birth-Death model with sampled ancestors. *PeerJ* **5**, e3055.

*Cavender, T. M. (1968). Freshwater fish remains from the Clarno Formation, Ochoco Mountains of north–central Oregon. *Ore Bin* **30**, 125–141.

*Cavender, T. M. (1986). Review of the fossil history of North American freshwater fishes. In: Hocutt, C. H. & Wiley, E. O. (eds.) *The zoogeography of North American freshwater fishes*, 699–724. John Wiley & Sons, New York.Cavin, L. (2010). Diversity of Mesozoic semionotiform fishes and the origin of gars (Lepisosteidae). *Naturwissenschaften* **97**, 1035–1040.

Cavin, L. (2017). *Freshwater Fishes: 250 Million Years of Evolutionary History*. ISTE Press-Elsevier, London and Oxford.

*Cavin, L., Boudad, L., Duffaud, S., Kabiri, L., Le Lœuff, J., Rouget, I. & Tong, H. (2001). L'évolution paléoenvironnementale des faunes de poissons du Crétacé supérieur du bassin du Tafilalt et des régions avoisinantes (Sud–Est du Maroc): implications paléobiogéographiques. *Comptes Rendus de l'Académie des Sciences. Série 2* **333**, 677–683.

Cavin, L., Boudad, L., Tong, H., Läng, E., Tabouelle, J. & Vullo, R. (2015). Taxonomic composition and trophic structure of the continental bony fish assemblage from the early Late Cretaceous of southeastern Morocco. *PLoS One* **10**, e0125786.

*Cavin, L. & Forey, P. L. (2001). Osteology and systematic affinities of *Palaeonotopterus greenwoodi* Forey 1997 (Teleostei: Osteoglossomorpha). *Zoological Journal of the Linnean Society* **133**, 25–52.

Cavin, L., Forey, P. L. & Giersch, S. (2013). Osteology of *Eubiodectes libanicus* (Pictet & Humbert, 1866) and some other ichthyodectiformes (Teleostei): phylogenetic implications. *Journal of Systematic Palaeontology*, *11*(2), 115–177.

Cavin, L., Suteethorn, V., Buffetaut, E. & Tong, H. (2007). A new Thai Mesozoic lungfish (Sarcopterygii, Dipnoi) with an insight into post-Palaeozoic dipnoan evolution. *Zoological Journal of the Linnean Society* **149**, 141–177.

Cavin, L., Valentin, X. & Garcia, G. (2016). A new mawsoniid coelacanth (Actinistia) from the Upper Cretaceous of Southern France. *Cretaceous Research* **62**, 65–73.

Chakrabarty, P. (2004). Cichlid biogeography: comment and review. *Fish and Fisheries* **5**, 97–119.

Chakrabarty, P., Davis, M. P. & Sparks, J. S. (2012). The first record of a trans-oceanic sister-group relationship between obligate vertebrate troglobites. *PLoS One* **7**, e44083.

Chakrabarty, P., Faircloth, B. C., Alda, F., Ludt, W. B., Mcmahan, C. D., Near, T. J., Dornburg, A., Albert, J. S., Arroyave, J., Stiassny, M. L. J., Sorenson, L. & Alfaro, M. J. (2017). Phylogenomic systematics of ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of Characiformes. *Systematic Biology* **66**, 881–895.

*Chang, A., Arratia, G. & Alfaro, G. (1978). *Percichthys lonquimayiensis* n. sp. from the upper Paleocene of Chile (Pisces, Perciformes, Serranidae). *Journal of Paleontology* **52**, 727–736.

*Chang, M. M. & Chou, C. C. (1976). Discovery of *Plesiolycoptera* in Songhuajiang–Liaoning Basin and origin of Osteoglossomorpha. *Vertebrata PalAsiatica* **14**, 146–153.

*Chang, M. M. & Chou, C. C. (1977). On Late Mesozoic fossil fishes from Zhejiang Province, China. *Memoirs of Institute of Vertebrate Paleontology and Paleoanthropology, Academia Sinica* **12**, 1–59.

*Chang, M. M. & Miao, D. (2004). An overview of Mesozoic fishes in Asia. In: Arratia, G. & Tintori, A. (eds.) *Mesozoic Fishes 3 – Systematics, Paleoenvironments and Biodiversity*, 535–563. Verlag Dr. Friedrich Pfeil, München.

*Chen, P. J., Li, J., Matsukawa, M., Zhang, H., Wang, Q. & Lockley, M. G. (2006). Geological ages of dinosaur–track–bearing formations in China. *Cretaceous Research* **27**, 22–32.

Chen, W. J., Lavoué, S., Beheregaray, L. B. & Mayden, R. L. (2014). Historical biogeography of a new antitropical clade of temperate freshwater fishes. *Journal of Biogeography* **41**, 1806–1818.

Chen, W. J., Lavoué, S. & Mayden, R. L. (2013). Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). *Evolution* **67**, 2218–2239.

Chew, S. F., Wong, M. Y., Tam, W. L. & Ip, Y. K. (2003). The snakehead *Channa asiatica* accumulates alanine during aerial exposure, but is incapable of sustaining locomotory activities on land through partial amino acid catabolism. *Journal of Experimental Biology* **206**, 693–704.

Choudhury, A. & Dick, T. A. (1998). The historical biogeography of sturgeons (Osteichthyes: Acipenseridae): a synthesis of phylogenetics, palaeontology and palaeogeography. *Journal of Biogeography* **25**, 623–640.

*Churcher, C. S. & De Iuliis, G. (2001). A new species of *Protopterus* and a revision of *Ceratodus humei* (Dipnoi: Ceratodontiformes) from the Late Cretaceous Mut Formation of eastern Dakhleh Oasis, Western Desert of Egypt. *Palaeontology* **44**, 305–323.

*Cione, A. L. & Azpelicueta, M. M. (2013). The first fossil species of *Salminus*, a conspicuous South American freshwater predatory fish (Teleostei, Characiformes), found in the Miocene of Argentina. *Journal of Vertebrate Paleontology* **33**, 1051–1060.

*Cione, A. L., Azpelicueta, M. M., Bond, M., Carlini, A. A., Casciotta, J. R., Cozzuol, M. A., de la Fuente, M., Gasparini, Z., Goin, F. J., Noriega, J. & Scillato-Yané, G. J. (2000). Miocene vertebrates from Entre Ríos province, eastern Argentina. *El Neógeno de Argentina* **14**, 191–237.

*Cione, A. L. & Báez, A. M. (2007). Peces continentales y anfibios cenozoicos de Argentina: los últimos cincuenta años. *Publicación Electrónica de la Asociación Paleontológica Argentina* **11**, 195–220.

*Cione, A. L. & Casciotta, J. R. (1995). Freshwater teleostean fishes from the Miocene of the Quebrada de la Yesera, Salta, Northwestern Argentina. *Neues Jahrbuch fur Geologie und Palaontologie–Abhandlungen* **196**, 377–394.

*Cione, A. L., Dahdul, W. M., Lundberg, J. G. & Machado-Allison, A. (2009). *Megapiranha paranensis*, a new genus and species of Serrasalmidae (Characiformes, Teleostei) from the upper Miocene of Argentina. *Journal of Vertebrate Paleontology* **29**, 350–358.

*Cione, A. L. & Torno, A. E. (1988). Assignment of the bony fish "*Propygidium primaevus*" (a supposed siluriform from the Tertiary of Patagonia) to the Order Perciformes. *Journal of Paleontology* **62**, 656–657.

*Cione, A. L., Vergani, G., Starck, D. & Herbst, R. (1995). Los peces del Mioceno de la quebrada de La Yesera, provincia de Salta, Argentina. Su valor como indicadores ambientales y su antiguedad. *Ameghiniana* **32**, 129–140.

Claeson, K. M., Sallam, H. M., O'Connor, P. M. & Sertich, J. J. (2014). A revision of the Upper Cretaceous lepidosirenid lungfishes from the Quseir Formation, Western Desert, central Egypt. *Journal of Vertebrate Paleontology* **34**, 760–766.

*Cockerell, T. D. A. (1922). Some fossil fish scales from Peru. *Proceedings of the U.S. National Museum* **59**, 19–20.

*Codrea, V., Barbu, O. & Bedelean, H. (2007). Middle Miocene diatomite–bearing formations from western Romania. *Bullettin of the Geological Society of Greece* **40**, 21–30.

Collins, R. A., Britz, R. & Rüber, L. (2015). Phylogenetic systematics of leaffishes (Teleostei: Polycentridae, Nandidae). *Journal of Zoological Systematics and Evolutionary Research* **53**, 259–272.

Cooper, A. & Cooper, R. A. (1995). The Oligocene bottleneck and New Zealand biota: genetic record of a past environmental crisis. *Proceedings of the Royal Society of London B: Biological Sciences* **261**, 293–302.

*Coryndon, S. C. (1966). Preliminary report on some fossils from the Chiwondo Beds of the Karonga District, Malawi. *American Anthropologist* **68**, 59–66.

*Costa, W. J. E. M. (2011). Redescription and phylogenetic position of the fossil killifish *Carrionellus diumortuus* White from the Lower Miocene of Ecuador (Teleostei: Cyprinodontiformes). *Cybium* **35**, 181–187.

Costa, W. J. E. M. (2012). Oligocene killifishes (Teleostei: Cyprinodontiformes) from southern France: relationships, taxonomic position, and evidence of internal fertilization. *Vertebrate Zoology* **62**, 371–386.

Criswell, K. E. (2015). The comparative osteology and phylogenetic relationships of African and South American lungfishes (Sarcopterygii: Dipnoi). *Zoological Journal of the Linnean Society* **174**, 801–858.

*Dahdul, W. M. (2004). Fossil serrasalmine fishes (Teleostei: Characiformes) from the Lower Miocene of north–western Venezuela. In: Sánchez-Villagra, M. R. & Clack, J. A. (eds.) *Fossils of the Miocene Castillo Formation, Venezuela: Contributions on Neotropical Palaeontology*, Special Papers in Palaeontology **71**, 23–28. Palaeontological Association.

Dahdul, W. M. (2010). Review of the phylogenetic relationships and fossil record of Characiformes. In: Grande, T., Poyato-Ariza, F. J. & Diogo, R. (eds.) *Gonorynchiformes and Ostariophysan Relationships: A Comprehensive Review*, 441–464. Science Publishers, Enfield.

*Danilchenko, P. G. (1968). Fishes of the upper Paleocene of Turkmenia. In: Obruchev, D. V. (ed.) *Essays on the Phylogeny and Systematics of Fossil Fishes and Agnathans*, 113–156. Nauka Press, Moscow.

Darlington, P. J. (1957). *Zoogeography: the Geographical Distribution of Animals*. John Wiley, New York.

*Dartevelle, E. & Casier, E. (1949). Les Poissons fossils du Bas-Congo et des regions voisines (deuxieme partie). *Annales du Musée du Congo Belge A, Série 3* **2**, 201–256.

*Dartevelle, E. & Casier, E. (1959). Les Poissons fossils du Bas–Congo et des regions voisines (treuxieme partie). *Annales du Musée du Congo Belge A, Série 3* **2**, 257–368.

Davenport, J. & Abdul Matin, A. K. M. (1990). Terrestrial locomotion in the climbing perch, *Anabas testudineus* (Bloch)(Anabantidea, Pisces). *Journal of Fish Biology* **37**, 175–184.

*Daxner-Höck, G., Böhme, M. & Kossler, A. (2013). New data on Miocene biostratigraphy and paleoclimatology of Olkhon Island (Lake Baikal, Siberia). In: Wang, X., Flynn, L. J. & Fortelius, M. (eds.) *Fossil Mammals of Asia: Neogene Biostratigraphy and Chronology,* 508–519. Columbia University Press, New York.

Day, J. J., Fages, A., Brown, K. J., Vreven, E. J., Stiassny, M. L., Bills, R., Friel, J. P. & Rüber, L. (2017). Multiple independent colonizations into the Congo Basin during the continental radiation of African *Mastacembelus* spiny eels. *Journal of Biogeography* **44**, 2308–2318.

*de Broin, F., Buffetaut, E., Koeniguer, J. C., Rage, J. C., Russell, D., Taquet, P., Vergnaud–Grazzini, C. & Wenz, S. (1974). La faune de vertébrés continentaux du gisement d'In Beceten (Sénonien du Niger). *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* **279**, 469–472.

de Carvalho, M. S. & Maisey, J. G. (2008). New occurrence of *Mawsonia* (Sarcopterygii: Actinistia) from the Early Cretaceous of the Sanfranciscana Basin, Minas Gerais, southeastern Brazil. *Geological Society*, *London*, *Special Publications* **295**, 109–144.

*de la Peña Zarzuelo, A. (1996). Characid teeth from the Lower Eocene of the Ager Basin (Lérida, Spain): paleobiogeographical comments. *Copeia* **1996**, 746–750.

de Queiroz, A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution* **20**, 68–73.

de Smet, M. E. M. & Barber, A. J. (2005). Tertiary stratigraphy. In: Barber, A. J., Crow, M. J, & Milsom, J. S. (eds.) *Sumatra. Geology*, *Resources and Tectonic Evolution*. *Geological Society of London Memoirs* **31**, 86–97.

*DeCelles, P. G. & Horton, B. K. (2003). Early to middle Tertiary foreland basin development and the history of Andean crustal shortening in Bolivia. *Geological Society of America Bulletin* **115**, 58–77.

Dietze, K. (2007). Redescription of *Dastilbe crandalli* (Chanidae, Euteleostei) from the Early Cretaceous Crato Formation of north-eastern Brazil. *Journal of Vertebrate Paleontology* **27**, 8–16.

*Divay, J. D. & Murray, A. M. (2013). A mid-Miocene ichthyofauna from the Wood Mountain Formation, Saskatchewan, Canada. *Journal of Vertebrate Paleontology* **33**, 1269–1291.

*Divay, J. D. & Murray, A. M. (2016). The fishes of the Farson Cutoff Fishbed, Bridger Formation (Eocene), Greater Green River Basin, Wyoming, USA. *Journal of Vertebrate Paleontology* **36**, e1212867.

Donoghue, M. J. & Moore, B. R. (2003). Toward an integrative historical biogeography. *Integrative and Comparative Biology* **43**, 261–270.

Dornburg, A., Townsend, J. P., Friedman, M. & Near, T. J. (2014). Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evolutionary Biology* **14**, 169.

Dubey, S. K., Trivedi, R. K., Chand, B. K., Mandal, B. & Rout, S. K. (2016). The effect of salinity on survival and growth of the freshwater stenohaline fish spotted snakehead *Channa punctata* (Bloch, 1793). *Zoology and Ecology* **26**, 282–291.

Duchêne, S., Lanfear, R. & Ho, S. Y. (2014). The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution* **78**, 277–289.

Dutheil, D. B. (1999). An overview of the freshwater fish fauna from the Kem Kem beds (Late Cretaceous: Cenomanian) of southeastern Morocco. In: Arratia, G. & Schultze, H.-P. (eds.), *Mesozoic Fishes 2 – Systematics and Fossil Record*, 553–563. Verlag Dr. F. Pfeil, München.

*Eastman, C. R. (1917). Fossil fishes in the collection of the United States National Museum. *Proceedings of the U.S. National Museum* **52**, 235–304.

*Estes, R. (1969). Two new late Cretaceous fishes from Montana and Wyoming. *Breviora* **335**, 1–15.

*Estes, R. (1984). Fish, amphibians and reptiles from the Etadunna Formation, Miocene of South Australia. *The Australian Zoologist* **21**, 335–343.

Fara, E., Gayet, M. & Taverne, L. (2010). The fossil record of Gonorynchiformes. In: Grande, T., Poyato-Ariza, F. J., & Diogo, R. (eds.) *Gonorynchiformes and Ostariophysan Relationships: A Comprehensive Review*, 173–226. Science Publishers, Enfield.

*Feibel, C. S. (1988). Paleoenvironments of the Koobi Fora Formation along the western Koobi Fora Ridge, East Turkana, Kenya. Unpublished doctoral thesis, University of Utah.

*Fernández, J. (1975). Hallazgo de peces pulmonados fósiles en la puna jujena. *Sociedad Cientifica Argentina, Anales Serie II* **41**, 13–18.

*Fernández, J., Pascual, R. & Bondesio, P. (1973). Restos de *Lepidosiren paradoxa* (Osteichthyes, Dipnoi) de la Formation Lumbrera (Eógeno, Eoceno?) de Jujuy. Consideraciones estratigráficas, paleoecológicas y paleogeográficas. *Ameghiniana* **10**, 152–172.

*Figueiredo, F. J. & Costa-Carvalho, B. C. M. (1999). *Plesiocurimata alvarengai* gen. et sp. nov.(Teleostei: Ostariophysi: Curimatidae) from the Tertiary of Taubaté Basin, São Paulo State, Brazil. *Anais da Academia Brasileira de Ciências* **71**, 885–893.

*Filippov, A. G., Erbaeva, M. A. & Sychevskaya, E.K. (2000). Miocene deposits in Aya cave near Baikal. *Geologiya i Geofizika* **41**, 755–764.

Filleul, A. & Maisey, J. G. (2004). Redescription of *Santanichthys diasii* (Otophysi, Characiformes) from the Albian of the Santana Formation and comments on its implications for otophysan relationships. *American Museum Novitates* **3455**, 1–21.

Fink, S. V. & Fink, W. L. (1981). Interrelationships of the ostariophysan fishes (Teleostei). *Zoological Journal of the Linnean Society* **72**, 297–353.

Forey, P.L. (1997). A Cretaceous notopterid (Pisces: Osteoglossomorpha) from Morocco. *South African Journal of Science* **93,** 564–569.

Forey, P. L. & Hilton, E. J. (2010). Two new tertiary osteoglossid fishes (Teleostei: Osteoglossomorpha) with notes on the history of the family. In: Elliott, D. K., Maisey, J. G., Yu, X. & Miao, D. (eds.), *Morphology*, *Phylogeny and Paleobiogeography of Fossil Fishes*, 215–246. Verlag Dr. F. Pfeil, München.

*Försterling, G. & Reichenbacher, B. (2002). Lithofazies und fischfaunen der Mittleren und Oberen Cerithien-Schichten (Ober-Oligozän–Unter-Miozän) im Mainzer Becken-paläoökologische und paläogeographische Implikationen. *Courier Forschungsinstitut Senckenberg* **237**, 293–314.

Frey, L., Maxwell, E. E. & Sánchez-Villagra, M. R. (2016). Intraspecific variation in fossil vertebrate populations: Fossil killifishes (Actinopterygii: Cyprinodontiformes) from the Oligocene of Central Europe. *Palaeontologia Electronica* **19**, 1–27.

Friedman, M. (2015). The early evolution of ray-finned fishes. *Palaeontology* **58**, 213–228.

Friedman, M., Keck, B. P., Dornburg, A., Eytan, R. I., Martin, C. H., Hulsey, C. D., Wainwright, P. C. & Near, T. J. (2013). Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20131733.

*Friedman, M., Tarduno, J. A. & Brinkman, D. B. (2003). Fossil fishes from the high Canadian Arctic: further palaeobiological evidence for extreme climatic warmth during the Late Cretaceous (Turonian-Coniacian). *Cretaceous Research* **24**, 615–632.

*Frost, G. A. (1924). Otoliths of fishes from the Upper Kimmeridgian of Buckinghamshire and Wiltshire. *Annals and Magazine of Natural History* **14**, 139–143.

Furness, A. I. (2016). The evolution of an annual life cycle in killifish: adaptation to ephemeral aquatic environments through embryonic diapause. *Biological Reviews* **91**, 796–812.

Furness, A. I., Reznick, D. N., Springer, M. S. & Meredith, R. W. (2015). Convergent evolution of alternative developmental trajectories associated with diapause in African and South American killifish. *Proceedings of the Royal Society of London B: Biological Sciences* **282**, 20142189.

Gamble, T., Bauer, A. M., Colli, G. R., Greenbaum, E., Jackman, T. R., Vitt, L. J. & Simons, A. M. (2011). Coming to America: multiple origins of New World geckos. *Journal of Evolutionary Biology* **24**, 231–244.

*Gaudant, J. (1965). *Lycoptera wangi* nov. sp. (Poisson téléostéen) dans le Jurassique des environs de Hêngshan (Shensi, Chine). *Compte Rendu Sommaire des Séances de la Société Géologique de France* **10**, 337–339.

*Gaudant, J. (1978*a*). L'ichthyofaune des marnes messiniennes des environs de Senigallia (Marche, Italie): Signification paleoecologique et paleogeographique. *Geobios* **11**, 913–919.

*Gaudant, J. (1978*b*). Sur une nouvelle espece de poissons téléostéens Cyprinodontiformes de l'Oligocene des environs de Manosque (Alpes de Haute-Provence). *Géologie Méditerranéenne* **5**, 281–290.

*Gaudant, J. (1979*a*). "*Pachylebias crassicaudus*" (Agassiz)(Poisson téléostéen, cyprinodontiforme), un constituant majeur de l'ichthyofaune du Messinien continental du bassin méditerranéen. *Geobios* **12**, 47–73.

*Gaudant, J. (1979*b*). Sur la présence de dents de Characidae (poissons téléostéens, Ostariophysi) dans les "Calcaires à Bythinies" et les "Sables bleutés" du Var. *Geobios* **12**, 451–457.

Gaudant, J. (1980). *Eurocharax tourainei* nov. gen., nov. sp. (poisson teleosteen, Ostariophysi): Nouveau Characidae fossile des "Calcaires a Bythinies" du Var. *Geobios* **13**, 683–703.

*Gaudant, J. (1981). Un nouveau Cyprinodontidae (poisson téléostéen) de l'Oligocène inférieur de Kleinkems (Pays de Bade, Allemagne): *Prolebias rhenanus* nov. sp. *Sciences géologiques Bulletin* **34**, 3–12.

Gaudant, J. (1982). *Prolebias catalaunicus* nov. sp.: une nouvelle espèce de Poissons cyprinodontidae de l'Oligocène de Sarreal (Province de Tarragona, Catalogne). *Estudios Geológicos* **38**, 95–102.

*Gaudant, J. (1988). Les Cyprinodontiformes (poissons téléostéens) oligocènes de Ronzon, Le Puy–en–Velay (Haute–Loire): anatomie et signification paléoécologique. *Geobios* **21**, 773–785.

*Gaudant, J. (1989). Découverte d'une nouvelle espèce de poissons cyprinodontiformes (*Prolebias delphinensis* nov. sp.) dans l'Oligocène du bassin de Montbrun-les-Bains (Drôme). *Géologie Méditerranéenne* **16**, 355–367.

*Gaudant, J. (1991). *Prolebias hungaricus* nov. sp. une nouvelle espèce de poissons Cyprinodontidae des diatomites Miocènes de Szurdokpüspöki (comté de Nógrád, Hongrie). *Magyar Állami Földtani Intézet Évi Jelentése az 1989*, 481–493.

*Gaudant, J. (1993). Un exemple de "régression évolutive" chez des poissons Cyprinodontidae du Miocène supérieur d'Espagne: *Aphanius illunensis* nov. sp. *Geobios* **26**, 449–454.

*Gaudant, J. (1996). Signification paléobiogéographique de la découverte de dents de Characiformes (Poissons téléostéens) dans le Miocène moyen de Sansan (Gers). *Comptes Rendus de l'Académie des Sciences. Série 2. Sciences de la Terre et des Planets* **322**, 799–803.

*Gaudant, J. (2000). L'ichthyofaune de Sansan: signification paléoécologique et paléobiogéographique. *Mémoires du Muséum National d'Histoire Naturelle* **183**, 155–175.

*Gaudant, J. (2002). La crise messinienne et ses effets sur l'ichthyofaune néogène de la Méditerranée: le témoignage des squelettes en connexion de poissons téléostéens. *Geodiversitas* **24**, 691–710.

*Gaudant, J. (2003*a*). *Prolebias euskadiensis* nov. sp., nouvelle espèce de poissons cyprinodontidae apodes de l'Oligo–Miocène d'Izarra (Province d'Álava, Espagne). *Revista Española de Paleontología* **18**, 171–178.

*Gaudant, J. (2003*b*). Sur quelques restes squelettiques de Channidae (poissons téléostéens) du Miocène moyen du Locle (Canton de Neuchâtel, Suisse). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* **126**, 115–119.

*Gaudant, J. (2006*a*). Occurrence of skeletal remains of channids (teleostean fishes) in the Ottnangian (Lower Miocene) of Langenau, near Ulm (Württemberg, Germany). *Stuttgarter Beiträge zur Naturkunde Serie B* **361**, 1–15.

*Gaudant, J. (2006*b*). Occurrence of the genus *Aphanius* Nardo (teleostean fishes, Cyprinodontidae) in the evaporitic Upper Badenian of Eastern Czech Republic. *Časopis Slezského zemského muzea Série A* **55**, 97–104.

*Gaudant, J. (2008). L'ichthyofaune messinienne du bassin de Sorbas (Almería, Espagne) et ses rapports avec l'environnement sédimentaire. *Revista Española de Paleontología* **23**, 211–223.

*Gaudant, J. (2009). Occurrence of the genus *Aphanius* Nardo (cyprinodontid fishes) in the Lower Miocene of the Cheb basin (Czech Republic), with additional notes on *Prolebias egeranus* Laube. *Journal of the National Museum (Prague), Natural History Series* **177**, 83–90.

*Gaudant, J. (2011). *Aphanius persicus* (Priem, 1908)(Pisces, Teleostei, Cyprinodontidae): une nouvelle combinaison pour *Brachylebias persicus* Priem, 1908, du Miocène supérieur des environs de Tabriz (Iran). *Geodiversitas* **33**, 347–356.

*Gaudant, J. (2012). Révision de *Prolebias stenoura* Sauvage, 1874 du Stampien (= Rupélien) de Limagne (centre de la France), espèce type du genre *Prolebias* (poisson téléostéen, Cyprinodontiformes). *Geodiversitas* **34**, 409–423.

*Gaudant, J. (2013). Occurrence of poeciliid fishes (Teleostei, Cyprinodontiformes) in the European Oligo–Miocene: the genus *Paralebias* nov. gen. *Neues Jahrbuch für Geologie und Paläontologie–Abhandlungen* **267**, 215–222.

Gaudant, J. (2014). Characiform fishes (Teleosts, Ostariophysi): how many waves of immigration into Europe during the Cenozoic? *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen* **273**, 319–326.

Gaudant, J. (2015). Revision of the poorly known Middle Miocene freshwater fish fauna from Beuern (Hesse, Germany). *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen* **278**, 291–302.

*Gaudant, J. (2016). *Francolebias arvernensis* n. sp., une nouvelle espèce de poissons cyprinodontiformes oligocènes de Chadrat (Saint-Saturnin, Puy-de-Dôme, France), avec une brève notice sur un Umbridae fossile du même gisement. *Geodiversitas* **38**, 435–449.

Gaudant, J., Barrón, E., Anadón, P., Reichenbacher, B. & Peñalver, E. (2015). Palaeoenvironmental analysis of the Miocene Arcas del Villar gypsum sequence (Spain), based on palynomorphs and cyprinodontiform fishes. *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen* **277**, 105–124.

Gaudant, J. & Reichenbacher, B. (1998). Première découverte d'un squelette de Channidae (Poisson téléostéen) dans le Miocène inférieur d'Illerkirchberg, près d'Ulm (Wurtemberg, Allemagne). *Paläontologische Zeitschrift* **72**, 383–388.

*Gaudant, J. & Reichenbacher, B. (2002). Anatomie et affinités des *Prolebias* aff. *weileri* Von Salis (Poissons téléostéens, Cyprinodontidae) du Miocène inférieur à moyen du Randecker Maar (Wurtemberg, Allemagne). *Stuttgarter Beiträge zur Naturkunde B* **331**, 1–11.

*Gaudant, J., Saint-Martin, J. P., Bessedik, M., Mansour, B., Moissette, P. & Rouchy, J. M. (1997). Découverte d'une frayère de poissons téléostéens dans des diatomites messiniennes du Djebel Murdjadjo (environs d'Oran, Algérie). *Journal of African Earth Sciences* **24**, 511–529.

*Gaudant, J. & Smith, R. (2008). Des dents de poissons Characiformes dans l'Eocène basal de Dormaal (niveau proche de la limite Paléocène–Éocène, Brabant flamand, Belgique). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* **78**, 269–275.

Gayet, M. (1985). Contribution à l'étude anatomique et systématique de l'ichthyofaune cénomanienne du Portugal. Troisième partie: complément à l'étude des Ostariophysaires. *Communicaçoes dos Serviços Geológicos de Portugal* **71**, 91–118.

*Gayet, M. (1987). Lower vertebrates from the early-middle Eocene Kuldana Formation of Kohat (Pakistan): Holostei and Teleostei. *Contributions from the Museum of Paleontology, University of Michigan* **27**, 151–168.

*Gayet, M. (1988). Découverte du plus ancien Channiforme (Pisces, Teleostei): *Parachannichthys ramnagarensis* n. g., n. sp., dans le Miocène moyen des Siwaliks (Ramnagar, Jammu et Cachemire, Inde): implications paléobiogéographiques. *Comptes Rendus de l'Académie des Sciences. Série 2* **307**, 1033–1036.

Gayet, M. (1990). Nouveaux Siluriformes du Maastrichtien de Tiupampa (Bolivie). *Comptes Rendus de l'Académie des Sciences. Série 2* **310**, 867–872.

Gayet, M. (1991). "Holostean" and teleostean fishes of Bolivia. *Revista Técnica de YPFB* **12**, 453–494.

Gayet, M., Jégu, M., Bocquentin, J. & Negri, F. R. (2003). New characoids from the Upper Cretaceous and Paleocene of Bolivia and the Mio-Pliocene of Brazil: phylogenetic position and paleobiogeographic implications. *Journal of Vertebrate Paleontology* **23**, 28–46.

Gayet, M., Marshall, L. G., Sempere, T., Meunier, F. J., Cappetta, H. & Rage, J. C. (2001). Middle Maastrichtian vertebrates (fishes, amphibians, dinosaurs and other reptiles, mammals) from Pajcha Pata (Bolivia). Biostratigraphic, palaeoecologic and palaeobiogeographic implications. *Palaeogeography*, *Palaeoclimatology*, *Palaeoecology* **169**, 39–68.

*Gayet, M. & Meunier, F. J. (1983). Ecailles actuelles et fossiles d'Osteoglossiformes (Pisces, Teleostei). *Comptes-Rendus des Séances de l'Académie des Sciences. Série 2* **297**, 867–870.

Gayet, M. & Meunier, F. J. (1998). Maastrichtian to early late Paleocene freshwater Osteichthyes of Bolivia: additions and comments. In: Malabarba, L. R., Reis, R. E., Vari, R. P., Lucena, Z. M. & Lucena, C. A. S. (eds.), *Phylogeny and Classification of Neotropical Fishes*, 85-110. EdiPUCRS, Porto Alegre.

Gayet, M., Meunier, F. J. & Werner, C. (2002). Diversification in Polypteriformes and special comparison with the Lepisosteiformes. *Palaeontology* **45**, 361–376.

*Getz, L. L. & Hibbard, C. W. (1965). A molluscan faunule from the Seymour Formation of Baylor and Knox counties, Texas. *Michigan Academy of Science, Arts, and Letters* **50**, 275–297.

Giles, S., Xu, G. H., Near, T. J. & Friedman, M. (2017). Early members of 'living fossil' lineage imply later origin of modern ray-finned fishes. *Nature* **549**, 265–268.

Gilpin, M. E. (1980). The role of stepping-stone islands. *Theoretical Population Biology* **17**, 247–253.

*González-Rodríguez, K. A., Espinosa-Arrubarrena, L. & González-Barba, G. (2013). An overview of the Mexican fossil fish record. In: Arratia, G., Schultze, H.-P. & Wilson, M. V. H. (eds.) *Mesozoic Fishes 5 – Global Diversity and Evolution*, 9–34. Verlag Dr. F. Pfeil, München.

Gottfried, M. D., Rogers, R. R. & Rogers, K. C. (2004). First record of Late Cretaceous coelacanths from Madagascar. In: Arratia, G., Wilson, M. V. H., & Cloutier, R. (eds.), *Recent Advances in the Origin and Early Radiation of Vertebrates*, 687–691. Verlag Dr. F. Pfeil, München.

Graham, J. B. (1997). *Air-Breathing Fishes. Evolution*, *Diversity and Adaptation*. Academic Press, San Diego.

*Grande, L. (1979). *Eohiodon falcatus*, a new species of hiodontid (Pisces) from the late Early Eocene Green River Formation of Wyoming. *Journal of Paleontology* **53**,103–111.

*Grande, L. (1986). The first articulated freshwater teleost fish from the Cretaceous of North America. *Palaeontology* **29**, 365–371.

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

Grande, L. (2010). An empirical synthetic pattern study of gars (Lepisosteiformes) and closely related species, based mostly on skeletal anatomy. The resurrection of Holostei. *Copeia* **2010**, *Supplement 2A*, 1–871.

Grande, L. & Bemis, W. E. (1998). A comprehensive phylogenetic study of amiid fishes (Amiidae) based on comparative skeletal anatomy. An empirical search for interconnected patterns of natural history. *Journal of Vertebrate Paleontology* **18**, *Supplement 1*, 1–696.

*Grande, L. & Cavender, T. M. (1991). Description and phylogenetic reassessment of the monotypic Ostariostomidae (Teleostei). *Journal of Vertebrate Paleontology* **11**, 405–416.

Grandstaff, B. S., Smith, J. B., Lamanna, M. C., Lacovara, K. J. & Abdel-Ghani, M. S. (2012). *Bawitius*, gen. nov., a giant polypterid (Osteichthyes, Actinopterygii) from the Upper Cretaceous Bahariya Formation of Egypt. *Journal of Vertebrate Paleontology* **32**, 17–26.

Granot, R. & Dyment, J. (2015). The Cretaceous opening of the South Atlantic Ocean. *Earth and Planetary Science Letters* **414**, 156–163.

*Greenwood, P. H. (1951). Fish remains from Miocene deposits of Rusinga Island and Kavirondo Province, Kenya. *Annals and Magazine of Natural History* **12**, 1192–1201.

*Greenwood, P. H. (1959). Quaternary fish-fossils. *Institut des Parcs Nationaux du Congo Belge, Exploration du Parc National Albert. Mission Journal of De Heinzelein de Beaucourt (1950)* **4**, 1–80.

*Greenwood, P. H. (1972). New fish fossils from the Pliocene of Wadi Natrun, Egypt. *Journal of Zoology* **168**, 503–519.

Greenwood, P. H. (1973). Interrelationships of osteoglossomorphs. In: Greenwood, P. H., Miles, R. S. & Patterson, C. (eds.), *Interrelationships of Fishes*, 307–332. Academic Press, London.

*Greenwood, P. H. (1976). Notes on *Sindacharax* Greenwood and Howes, 1975, a genus of fossil African characid fishes. *Revue Zoologique Africaine* **90**, 1–13.

Greenwood, P. H. (1983). The zoogeography of African freshwater fishes: bioaccountancy or biogeography. In: Sims, R. W., Price, J. H. & Whalley, P. E. S. (eds.), *Evolution*, *Time and Space: The Emergence of the Biosphere*, 179–199. Academic Press, London.

*Greenwood, P. H. & Howes, G. J. (1975). Neogene Fossil Fishes from the Lake Albert–lake Edward Rift (Zaire). *Bulletin of the British Museum of Natural History (Geology)* **26**, 72–127.

*Greenwood, P. H. & Patterson, C. (1967). A fossil osteoglossoid fish from Tanzania (E. Africa). *Journal of the Linnean Society of London, Zoology* **47**, 211–223.

Grigorescu, D., Hartenberger, J. L., Radulescu, C., Samson, P. & Sudre, J. (1985). Découverte de mammifères et dinosaures dans le Crétacé supérieur de Pui (Roumanie). *Comptes Rendus de l'Académie des Sciences. Série 2* **301**, 1365–1368.

Guinot, G. & Cavin, L. (2018). Body size evolution and habitat colonization across 100 million years (Late Jurassic–Paleocene) of the actinopterygian evolutionary history. *Fish and Fisheries* **19**, 577–597.

*Günther, A. (1876). Contributions to our knowledge of the fish-fauna of the Tertiary deposits of the highlands of Padang, Sumatra. *Geological Magazine* **3**, 433–440.

Guzmán, A. F. (2015). El registro fósil de los peces mexicanos de agua dulce. *Revista Mexicana de Biodiversidad* **86**, 661–673.

*Guzmán, A. F. & Polaco, O. J. (2009). Peces fósiles mexicanos de agua dulce. *Setenta y cinco anos de la Escuela Nacional de Ciencias Biológicas*, 316–340.

*Haaland, R. & Magid, A. A. (1995). *Aqualithic sites along the rivers Nile and Atbara, Sudan*. Alma Mater, Bergen.

Haddoumi, H., Allain, R., Meslouh, S., Metais, G., Monbaron, M., Pons, D., Rage, J.-C., Vullo, R., Zouhri, S. & Gheerbrant, E. (2016). Guelb el Ahmar (Bathonian, Anoual Syncline, eastern Morocco): First continental flora and fauna including mammals from the Middle Jurassic of Africa. *Gondwana Research* **29**, 290–319.

Hammouda, S. A., Murray, A. M., Divay, J. D., Mebrouk, F., Adaci, M. & Bensalah, M. (2016). Earliest occurrence of *Hydrocynus* (Characiformes, Alestidae) from Eocene continental deposits of Méridja Hamada, northwestern Sahara, Algeria. *Canadian Journal of Earth Sciences* **53**, 1042–1052.

*Haug, E. (1905). Paléontologie. In: Foureau, F. (ed.), *Documents scientifiques de la mission saharienne (Mission Foureau Lamy) d'Alger au Congo par le Tchad*. Paléontologie Société de Géographie, Paris.

Hedman, M. M. (2010). Constraints on clade ages from fossil outgroups. *Paleobiology* **36**, 16–31.

Heine, C., Zoethout, J. & Müller, R. D. (2013). Kinematics of the South Atlantic rift. *Solid Earth* **4**, 215–253.

*Hibbard, C. W. & Dalquest, W. W. (1966). Fossils from the Seymour Formation of Knox and Baylor Counties, Texas, and their bearing on the late Kansan climate of that region. *Contributions from the Museum of Paleontology, University of Michigan* **21**, 1–66.

*Hibbard, C. W. & Dunkle, D. H. (1942). A new species of cyprinodontid fish from the Middle Pliocene of Kansas. *Kansas Geological Survey, Bulletin* **41**, 270–276.

Hills, E. S. (1934). Tertiary fresh water fishes from southern Queensland. *Memoirs of the Queensland Museum* **10**, 163–172.

*Hills, E. S. (1943). Tertiary fresh-water fishes and crocodilian remains from Gladstone and Duaringa, Queensland. *Memoirs of the Queensland Museum* **12**, 96–100.

Hills, E. S. (1946). Fossil Murray cod (*Maccullochella macquariensis*) from diatomaceous earths in New South Wales. *Records of the Australian Museum* **21**, 380–382.

*Hilton, E. J. (2002). Osteology of the extant North American fishes of the genus *Hiodon* Lesueur, 1818 (Teleostei: Osteoglossomorpha: Hiodontiformes). *Fieldiana Zoology* **100**, 1–142.

Hilton, E. J. (2003). Comparative osteology and phylogenetic systematics of fossil and living bony-tongue fishes (Actinopterygii, Teleostei, Osteoglossomorpha). *Zoological Journal of the Linnean Society* **137**, 1–100.

*Hilton, E. J. & Grande, L. (2008). Fossil mooneyes (Teleostei: Hiodontiformes, Hiodontidae) from the Eocene of western North America, with a reassessment of their taxonomy. *Geological Society, London, Special Publications* **295**, 221–251.

Hirsch, P. E., N'Guyen, A., Muller, R., Adrian-Kalchhauser, I. & Burkhardt-Holm, P. (in press). Colonizing Islands of water on dry land—on the passive dispersal of fish eggs by birds. *Fish and Fisheries*, 1–9.

Ho, S. Y. & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology* **23**, 5947–5965.

*Hora, S. L. (1938). On some fossil fish-scales from the Inter-Trappean beds at Deothan and Kheri, Central Provinces. *Records of the Geological Survey of India* **73**, 267–294.

Houle, A. (1998). Floating islands: a mode of long-distance dispersal for small and medium-sized terrestrial vertebrates. *Diversity and Distributions* **4**, 201–216.

Hughes, L. C., Ortí, G., Huang, Y., Sun, Y., Baldwin, C. C., Thompson, A. W., Arcila, D., Betancur-R, R., Li, C., Becker, L., Bellora, N., Zhao, X., Li, X., Wang, M., Fang, C. *et al.*

(2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences*, 201719358.

Imoto, J. M., Saitoh, K., Sasaki, T., Yonezawa, T., Adachi, J., Kartavtsev, Y. P., Miya, M., Nishida, M. & Hanzawa, N. (2013). Phylogeny and biogeography of highly diverged freshwater fish species (Leuciscinae, Cyprinidae, Teleostei) inferred from mitochondrial genome analysis. *Gene* **514**, 112–124.

Inoue, J. G., Kumazawa, Y., Miya, M. & Nishida, M. (2009). The historical biogeography of the freshwater knifefishes using mitogenomic approaches: a Mesozoic origin of the Asian notopterids (Actinopterygii: Osteoglossomorpha). *Molecular Phylogenetics and Evolution* **51**, 486–499.

Irisarri, I., Baurain, D., Brinkmann, H., Delsuc, F., Sire, J.Y., Kupfer, A., Petersen, J., Jarek, M., Meyer, A., Vences, M. & Philippe, H. (2017). Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nature Ecology & Evolution* **1**, 1370–1378.

*Jakowlew, V. N. (1959). Fishes from Miocene deposits of Kirghizia. *Paleontologicheskij Zhurnal* **3**, 107–111.

*Jin, F. (1991). A new genus and species of Hiodontidae from Xintai, Shandong. *Vertebrata PalAsiatica* **29**, 46–54.

*Jin, F. (1994). A nomen novum for *Tanichthys* Jin, 1991. *Vertebrata PalAsiatica* **32**, 70.

*Jin, F., Zhang, J. & Zhou, Z. (1995). Late Mesozoic fish fauna from western Liaoning, China. *Vertebrata PalAsiatica* **33**, 169–193.

Johnson, G. D. (1984). Percoidei: development and relationships. In: Moser, H. G., Richards, W. J., Cohen, D. M., Fahay, M. P., Kendall, A. W. Jr. & Richardson, S. L. (eds.), *Ontogeny and Systematics of Fishes*, 464–498. American Society of Ichthyologists and Herpetologists.

*Jolly, A. & Bajpai, S. (1988). Fossil Osteoglossidae from the Kalakot Zone (Middle Eocene): implications for palaeoecology, palaeobiogeography and correlation. *Bulletin of the Indian Geologists' Association* **21**, 71–79.

*Jost, J., Kälin, D., Schulz-Mirbach, T. & Reichenbacher, B. (2007). Late Early Miocene lake deposits near Mauensee, central Switzerland: fish fauna (otoliths, teeth), accompanying biota and palaeoecology. *Eclogae Geologicae Helvetiae* **99**, 309–326.

Kappeler, P. M. (2000). Lemur origins: Rafting by groups of hibernators? *Folia Primatologica* **71**, 422–425.

Keith, P., Lord, C., Lorion, J., Watanabe, S., Tsukamoto, K., Couloux, A. & Dettai, A. (2011). Phylogeny and biogeography of Sicydiinae (Teleostei: Gobiidae) inferred from mitochondrial and nuclear genes. *Marine Biology* **158**, 311–326.

Kemp, A., Cavin, L. & Guinot, G. (2017). Evolutionary history of lungfishes with a new phylogeny of post-Devonian genera. *Palaeogeography*, *Palaeoclimatology*, *Palaeoecology* **471**, 209–219.

Khare, S. K. (1976). Eocene fishes and turtles from the Subathu Formation, Beragua coal mine, Jammu and Kashmir. *Journal of the Palaeontological Society of India* **18**, 36–43.

*Khosla, A. (2014). Upper Cretaceous (Maastrichtian) charophyte gyrogonites from the Lameta Formation of Jabalpur, Central India: palaeobiogeographic and palaeoecological implications. *Acta Geologica Polonica* **64**, 311–323.

Kordikova, E. G., Heizmann, E. P. & Pronin, V. G. (2003). Tertiary litho-and biostratigraphic sequence of the Ustyurt Plateau area, SW Kazakhstan, with the main focus on vertebrate faunas from the Early to Middle Miocene. *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen* **227**, 381–447.

Koufos, G. D., Kostopoulos, D. S. & Vlachou, T. D. (2005). Neogene/Quaternary mammalian migrations in eastern Mediterranean. *Belgian Journal of Zoology* **135**, 181–190.

*Kumar, K., Rana, R. S. & Paliwal, B. S. (2005). Osteoglossid and lepisosteid fish remains from the Paleocene Palana Formation, Rajasthan, India. *Palaeontology* **48**, 1187–1210.

Kumazawa, Y. & Nishida, M. (2000). Molecular phylogeny of osteoglossoids: a new model for Gondwanian origin and plate tectonic transportation of the Asian arowana. *Molecular Biology and Evolution* **17**, 1869–1878.

Landis, C. A., Campbell, H. J., Begg, J. G., Mildenhall, D. C., Paterson, A. M. & Trewick, S. A. (2008). The Waipounamu Erosion Surface: questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. *Geological Magazine* **145**, 173–197.

*Larson, D. W., Brinkman, D. B. & Bell, P. R. (2010). Faunal assemblages from the upper Horseshoe Canyon Formation, an early Maastrichtian cool–climate assemblage from Alberta, with special reference to the *Albertosaurus sarcophagus* bonebed. *Canadian Journal of Earth Sciences* **47**, 1159–1181.

*Lavocat, R. (1955). Découverte de Dipneustes du genre *Protopterus* dans le tertiaire ancien de Tamaguilelt (Soudan français). *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* **240**, 1915–1917.

Lavoué, S. (2016). Was Gondwanan breakup the cause of the intercontinental distribution of Osteoglossiformes? A time-calibrated phylogenetic test combining molecular, morphological, and paleontological evidence. *Molecular Phylogenetics and Evolution* **99**, 34–43.

Lavoué, S., Nakayama, K., Jerry, D. R., Yamanoue, Y., Yagishita, N., Suzuki, N., Nishida, M. & Miya, M. (2014). Mitogenomic phylogeny of the Percichthyidae and Centrarchiformes (Percomorphaceae): comparison with recent nuclear gene-based studies and simultaneous analysis. *Gene* **549**, 46–57.

Lawver, L. A. & Gahagan, L. M. (2003). Evolution of Cenozoic seaways in the circum-Antarctic region. *Palaeogeography*, *Palaeoclimatology*, *Palaeoecology* **198**, 11–37.

Lee, D. E., McDowall, R. M. & Lindqvist, J. K. (2007). *Galaxias* fossils from Miocene lake deposits, Otago, New Zealand: the earliest records of the Southern Hemisphere family Galaxiidae (Teleostei). *Journal of the Royal Society of New Zealand* **37**, 109–130.

*Li, G.-Q. (1987). A new genus of Hiodontidae from Luozigou Basin, east Jilin. *Vertebrata PalAsiatica* **25**, 91.

*Li, G.-Q. (1994). Systematic position of the Australian fossil osteoglossid fish †*Phareodus* (= *Phareoides*) *queenslandicus* Hills. *Memoirs of the Queensland Museum* **37**, 287–300.

*Li, G.-Q. (1996). A new species of Late Cretaceous osteoglossid (Teleostei) from the Oldman Formation of Alberta, Canada, and its phylogenetic relationships. In: Arratia, G. & Viohl, G. (eds.) *Mesozoic Fishes. Systematics and Paleoecology*, 285–298. Verlag Dr. F. Pfeil, München.

*Li, G.-Q., Grande, L. & Wilson, M. V. H. (1997). The species of †*Phareodus* (Teleostei: Osteoglossidae) from the Eocene of North America and their phylogenetic relationships. *Journal of Vertebrate Paleontology* **17**, 487–505.

*Li, G.-Q. & Wilson, M. V. H. (1994). An Eocene species of *Hiodon* from Montana, its phylogenetic relationships, and the evolution of the postcranial skeleton in the Hiodontidae (Teleostei). *Journal of Vertebrate Paleontology* **14**, 153–167.

*Li, G.-Q. & Wilson, M. V. H. (1996). The discovery of Heterotidinae (Teleostei: Osteoglossidae) from the Paleocene Paskapoo formation of Alberta, Canada. *Journal of Vertebrate Paleontology* **16**, 198–209.

Li, G.-Q. & Wilson, M. V. H. (1999). Early divergence of Hiodontiformes *sensu stricto* in East Asia and phylogeny of some Late Mesozoic teleosts from China. In: Arratia, G. and Schultze, H.-P. (eds) *Mesozoic Fishes 2 – Systematics and Fossil Record*, 369–384. Verlag Dr. F. Pfeil, München.

Li, H., He, Y., Jiang, J., Liu, Z. & Li, C. (2018). Molecular systematics and phylogenetic analysis of the Asian endemic freshwater sleepers (Gobiiformes: Odontobutidae). *Molecular Phylogenetics and Evolution* **121**, 1–11.

*Li, S., Zheng, D. R., Zhang, Q., Liao, H. Y., Wang, H., Wang, B., Wang, J., Lu, H. B., Chang, S. C. & Zhang, H. C. (2016). Discovery of the Jehol Biota from the Celaomiao region and discussion of the Lower Cretaceous of the Bayingebi Basin, northwestern China. *Palaeoworld* **25**, 76–83.

Li, X., Musikasinthorn, P. & Kumazawa, Y. (2006). Molecular phylogenetic analyses of snakeheads (Perciformes: Channidae) using mitochondrial DNA sequences. *Ichthyological Research* **53**, 148–159.

*Lindoso, R. M., Maisey, J. G. & de Souza Carvalho, I. (2016). Ichthyofauna from the Codó Formation, Lower Cretaceous (Aptian, Parnaíba Basin), Northeastern Brazil and their paleobiogeographical and paleoecological significance. *Palaeogeography, Palaeoclimatology, Palaeoecology* **447**, 53–64.

Liu, H.-T. & Su, T.-T. (1962). Pliocene fishes from Yushe basin, Shansi. *Vertebrata PalAsiatica* **6**, 1–25.

*Liu, X. T., Ma, F. Z. & Liu, Z. C. (1982). Pisces. In: Geological Bureau of Nei Mongol Autonomous Region (eds.) *The Mesozoic Stratigraphy and Paleontology of Guyang Coalbearing Basin, Nei Mongol, China*, 101–122. Geological Publishing House, Beijing.

*Liu, X. T., Ma, F. Z. & Liu, Z. C. (1985). Discovery of *Kuntulunia* from the Shanganning Basin of North China and its stratigraphic significance. *Vertebrata PalAsiatica* **23**, 255.

*Livingston, T. D. & Dattilo, B. F. (2004). Middle Miocene lacustrine strata and fossil killifish in a volcanic setting: the rocks of Pavits Spring, Nevada Test Site, NYE County, Nevada [abstract]. *Geological Society of America Abstracts with Programs* **36**, 286.

*Lockley, M. G., Li, J., Matsukawa, M. & Li, R. (2012). A new avian ichnotaxon from the Cretaceous of Nei Mongol, China. *Cretaceous Research* **34**, 84–93.

Longrich, N. R. (2017). A stem lepidosireniform lungfish (Sarcopterygia: Dipnoi) from the Upper Eocene of Libya, North Africa and implications for Cenozoic lungfish evolution. *Gondwana Research* **42**, 140–150.

López-Arbarello, A. & Sferco, E. (2018). Neopterygian phylogeny: the merger assay. *Royal Society Open Science* **5**, 172337.

Lugaski, T. (1977). *Fundulus lariversi*, a new Miocene fossil cyprinodont fish from Nevada. *Wasmann Journal of Biology* **35**, 203–211.

Lundberg, J. G. (1993). African–South American freshwater fish clades and continental drift: problems with a paradigm. In: Goldblatt, P. (ed.), *Biological Relationships between Africa and South America*, 156–199. Yale University Press, New Haven.

*Lundberg, J. G. (1997). Freshwater fishes and their paleobiotic implications. In: Kay, R. F., Madden, R. H., Cifelli, R. L. & Flynn, J. J., (eds.), *Vertebrate paleontology in the neotropics: the Miocene fauna of La Venta, Colombia*, 67–91. Smithsonian Institution Press, Washington.

*Lundberg, J. G. & Chernoff, B. (1992). A Miocene fossil of the Amazonian fish *Arapaima* (Teleostei, Arapaimidae) from the Magdalena River region of Colombia--Biogeographic and evolutionary implications. *Biotropica* **24**, 2–14.

*Lundberg, J. G., Machado-Allison, A. & Kay, R. F. (1986). Miocene characid fishes from Colombia: evolutionary stasis and extirpation. *Science* **234**, 208–210.

*Lundberg, J. G., Sabaj Pérez, M. H., Dahdul, W. M. & Aguilera, O. A. (2010). The Amazonian Neogene fish fauna. In: Hoorn, C. & Wesselingh, F.P. (eds.) *Amazonia: Landscape and Species Evolution*, 281–301. Wiley-Blackwell, Oxford.

Lundberg, J. G., Sullivan, J. P., Rodiles-Hernández, R. & Hendrickson, D. A. (2007). Discovery of African roots for the Mesoamerican Chiapas catfish, *Lacantunia enigmatica*, requires an ancient intercontinental passage. *Proceedings of the Academy of Natural Sciences of Philadelphia* **156**, 39–53.

*Ma, F. Z. (1980). A new genus of Lycopteridae from Ningxia, China. *Vertebrata PalAsiatica* **18**, 286–295.

*Ma, F. Z. (1983). Early Cretaceous primitive teleosts from the Jiaohe Basin of Jilin Province, China. *Vertebrata PalAsiatica* **21**, 17–31.

*Ma, F. Z. (1984). The study of fossil fishes from the Jiuquan Basin, Gansu. *Vertebrata PalAsiatica* **22**, 330–332.

*Ma, F. Z. (1986). On the generic status of *Lycoptera tungi*. *Vertebrata PalAsiatica* **24**, 260.

*Ma, F. Z. (1987). Review of *Lycoptera davidi*. *Vertebrata PalAsiatica* **25**, 8–19.

*Ma F. Z. (1993). *Late Mesozoic fossil fishes from the Jiuquan Basin of Gansu Province, China*, 1–118. Marine Press, Beijing.

*Ma, F. Z. & Sun, J. R. (1988). Jura-Cretaceous ichthyofaunas from Sankeyushu section of Tonghua, Jilin. *Acta Palaeontology Sinica* **27**, 694–711.

MacDonald, C. M. (1978). Morphological and biochemical systematics of australian freshwater and estuarine percichthyid fishes. *Marine and Freshwater Research* **29**, 667–698.

*Mahboubi, M., Ameur, R., Crochet, J. Y. & Jaeger, J. J. (1984). Implications paléobiogéographiques de la découverte d'une nouvelle localité éocène à vertébrés continentaux en Afrique Nord-occidentale: El Kohol (Sud-Oranais, Algérie). *Geobios* **17**, 625–629.

Malabarba, M. C. (1998). Phylogeny of fossil Characiformes and paleobiogeography of the Tremembé formation, São Paulo, Brazil. In: Malabarba, L. R., Reis, R. E., Vari, R. P., Lucena, Z. M. & Lucena, C. A. S. (eds.) *Phylogeny and Classification of Neotropical Fishes*, 69–84. EdiPUCRS, Porto Alegre.

Malabarba, M. C. & Malabarba, L. R. (2010). Biogeography of Characiformes: an evaluation of the available information of fossil and extant taxa. In: Nelson, J. S., Schultze, H.-P. & Wilson, M. V. H. (eds.) *Origin and Phylogenetic Interrelationships of Teleosts*, 317–336. Verlag Dr. F. Pfeil, München.

Malabarba, M. C., Malabarba, L. R. & López-Fernández, H. (2014). On the Eocene cichlids from the Lumbrera Formation: additions and implications for the Neotropical ichthyofauna. *Journal of Vertebrate Paleontology* **34**, 49–58.

Marshall, C. R. (1997). Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* **23**, 165–173.

Martill, D. M., Ibrahim, N., Brito, P. M., Baider, L., Zhouri, S., Loveridge, R., Naish, D. & Hing, R. (2011). A new Plattenkalk Konservat Lagerstätte in the upper Cretaceous of Gara Sbaa, south-eastern Morocco. *Cretaceous Research* **32**, 433–446.

*Martin, H. A., Worrall, L. & Chalson, J. (1987). The first occurrence of the Paleocene *Lygistepollenites balmei* Zone in the eastern highlands region, New South Wales. *Australian Journal of Earth Sciences* **34**, 359–365.

*Martin, M. (1984). Deux Lepidosirenidae (Dipnoi) crétacés du Sahara, *Protopterus humei* (Priem) et *Protopterus protopteroides* (Tabaste). *Paläontologische Zeitschrift* **58**, 265–277.

*Martin, M. (1995). Nouveaux lepidosirenides (Dipnoi) du Tertiaire africain. *Geobios* **28**, 275–280.

*Martin, M. (1997). *Protopterus nigeriensis* nov. sp., l'un des plus anciens protoptères –Dipnoi– (In Beceten, Sénonien du Niger). *Comptes Rendus de l'Académie des Sciences. Série 2* **325**, 635–638.

*Martinelli, A. & Forasiepi, A. (2004). Late Cretaceous vertebrates from Bajo de Santa Rosa (Allen Formation), Río Negro province, Argentina, with the description of a new sauropod dinosaur (Titanosauridae). *Revista del Museo Argentino de Ciencias Naturales Nueva Serie* **6**, 257-305.

Matschiner, M. (in press). Gondwanan vicariance or trans-Atlantic dispersal of cichlid fishes: A review of the molecular evidence. *Hydrobiologia*.

Matschiner, M., Musilová, Z., Barth, J. M., Starostová, Z., Salzburger, W., Steel, M. & Bouckaert, R. (2017). Bayesian phylogenetic estimation of clade ages supports trans-Atlantic dispersal of cichlid fishes. *Systematic Biology* **66**, 3–22.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* **63**, 951–970.

Mayrinck, D., Brito, P. M., Meunier, F. J., Alvarado-Ortega, J. & Otero, O. (2017). †*Sorbinicharax verraesi*: An unexpected case of a benthic fish outside Acanthomorpha in the Upper Cretaceous of the Tethyan Sea. *PLoS One* **12**, e0183879.

Mayrinck, D., Brito, P. M. & Otero, O. (2015). Anatomical review of †*Salminops ibericus*, a Teleostei *incertae sedis* from the Cenomanian of Portugal, anciently assigned to Characiformes and possibly related to crossognathiform fishes. *Cretaceous Research* **56**, 66–75.

McDowall, R. M. (1972). The species problem in freshwater fishes and the taxonomy of diadromous and lacustrine populations of *Galaxias maculatus* (Jenyns). *Journal of the Royal Society of New Zealand* **2**, 325–367.

McDowall, R. M. (2002). Accumulating evidence for a dispersal biogeography of southern cool temperate freshwater fishes. *Journal of Biogeography* **29**, 207–219.

McDowall, R. M. (2007). On amphidromy, a distinct form of diadromy in aquatic organisms. *Fish and Fisheries* **8**, 1–13.

McDowall, R. M. (2010). *New Zealand Freshwater Fishes. An Historical and Ecological Biogeography*. Springer Science & Business Media.

McDowall, R. M. & Burridge, C. P. (2011). Osteology and relationships of the southern freshwater lower euteleostean fishes. *Zoosystematics and Evolution* **87**, 7–185.

McDowall, R. M. & Lee, D. E. (2005). Probable perciform fish scales from a Miocene freshwater lake deposit, Central Otago, New Zealand. *Journal of the Royal Society of New Zealand* **35**, 339–344.

McDowall, R. M. & Pole, M. (1997). A large galaxiid fossil (Teleostei) from the Miocene of Central Otago, New Zealand. *Journal of the Royal Society of New Zealand* **27**, 193–198.

*Mead, J. G. (1975). A fossil beaked whale (Cetacea: Ziphiidae) from the Miocene of Kenya. *Journal of Paleontology* **49**, 745–751.

Micklich, N. & Roscher, B. (1990). Neue Fischfunde aus der Baid-Formation (Oligozän; Tihamat Asir, SW Saudi-Arabien). *Neues Jahrbuch für Geologie und Paläontologie* **180**, 139–175.

*Miller, R. R. (1945). Four new species of fossil cyprinodont fishes from eastern California. *Journal of the Washington Academy of Sciences* **35**, 315–321.

Miller, R. R. & Smith, M. L. (1986). Origin and geography of the fishes of central Mexico. In: Hocutt, C. H. & Wiley, E. O. (eds.) *The Zoogeography of North American Freshwater Fishes*, 487–517. John Wiley & Sons, New York.

Miya, M., Friedman, M., Satoh, T. P., Takeshima, H., Sado, T., Iwasaki, W., Yamanoue, Y., Nakatani, M., Mabuchi, K., Inoue, J. G., Poulsen, J. Y., Fukunaga, T., Sato, Y. & Nishida, M. (2013). Evolutionary origin of the Scombridae (tunas and mackerels): members of a Paleogene adaptive radiation with 14 other pelagic fish families. *PLoS One* **8**, e73535.

*Monod, T. & Gaudant, J. (1998). Un nom pour les poissons Characiformes de l'Eocène inférieur et moyen du bassin de Paris et du sud de la France: *Alestoides eocaenicus* nov. gen., nov. sp. *Cybium* **22**, 15–20.

*Monsch, K. A. (1998). Miocene fish faunas from the northwestern Amazonia basin (Colombia, Peru, Brazil) with evidence of marine incursions. *Palaeogeography, Palaeoclimatology, Palaeoecology* **143**, 31–50.

Murphy, W. J. & Collier, G. E. (1997). A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism. *Molecular Biology and Evolution* **14**, 790–799.

Murray, A. M. (2000*a*). Eocene cichlid fishes from Tanzania, East Africa. *Journal of Vertebrate Paleontology* **20**, 651–664.

*Murray, A. M. (2000*b*). The Palaeozoic, Mesozoic and Early Cenozoic fishes of Africa. *Fish and Fisheries* **1**, 111–145.

*Murray, A. M. (2003*a*). A new characiform fish (Teleostei: Ostariophysi) from the Eocene of Tanzania. *Canadian Journal of Earth Sciences* **40**, 473–481.

Murray, A. M. (2003*b*). A new Eocene citharinoid fish (Ostariophysi: Characiformes) from Tanzania. *Journal of Vertebrate Paleontology* **23**, 501–507.

*Murray, A. M. (2004). Late Eocene and early Oligocene teleost and associated ichthyofauna of the Jebel Qatrani Formation, Fayum, Egypt. *Palaeontology* **47**, 711–724.

*Murray, A. M. (2006). A new channid (Teleostei: Channiformes) from the Eocene and Oligocene of Egypt. *Journal of Paleontology* **80**, 1172–1178.

Murray, A. M. (2012). Relationships and biogeography of the fossil and living African snakehead fishes (Percomorpha, Channidae, *Parachanna*). *Journal of Vertebrate Paleontology* **32**, 820–835.

Murray, A. M., Cook, T. D., Attia, Y. S., Chatrath, P. & Simons, E. L. (2010*a*). A freshwater ichthyofauna from the late Eocene Birket Qarun Formation, Fayum, Egypt. *Journal of Vertebrate Paleontology* **30**, 665–680.

*Murray, A. M., Newbrey, M. G., Neuman, A. G. & Brinkman, D. B. (2016). New articulated osteoglossomorph from Late Cretaceous freshwater deposits (Maastrichtian, Scollard Formation) of Alberta, Canada. *Journal of Vertebrate Paleontology* **36**, e1120737.

Murray, A. M. & Thewissen, J. M. (2008). Eocene actinopterygian fishes from Pakistan, with the description of a new genus and species of channid (Channiformes). *Journal of Vertebrate Paleontology* **28**, 41–52.

*Murray, A. M. & Wilson, M. V. H. (2005). Description of a new Eocene osteoglossid fish and additional information on †*Singida jacksonoides* Greenwood and Patterson, 1967 (Osteoglossomorpha), with an assessment of their phylogenetic relationships. *Zoological Journal of the Linnean Society* **144**, 213–228.

*Murray, A. M., You, H. L. & Peng, C. (2010*b*). A new Cretaceous osteoglossomorph fish from Gansu Province, China. *Journal of Vertebrate Paleontology* **30**, 322–332.

Murray, A. M., Zelenitsky, D. K., Brinkman, D. B. & Neuman, A. G. (2018). Two new Palaeocene osteoglossomorphs from Canada, with a reassessment of the relationships of the genus †*Joffrichthys*, and analysis of diversity from articulated versus microfossil material. *Zoological Journal of the Linnean Society* **183**, 907–944.

Myers, G. S. (1938). Fresh-water fishes and West Indian zoogeography. *Annual Report of the Board of Regents of the Smithsonian Institution* **92**, 339–364.

Najman, Y., Appel, E., Boudagher-Fadel, M., Bown, P., Carter, A., Garzanti, E., Godin, L., Han, J., Liebke, U., Oliver, G., Parrish, R. & Vezzoli, G. (2010). Timing of India-Asia collision: Geological, biostratigraphic, and palaeomagnetic constraints. *Journal of Geophysical Research: Solid Earth* **115**, B12416.

Near, T. J., Dornburg, A., Eytan, R. I., Keck, B. P., Smith, W. L., Kuhn, K. L., Moore, J. A., Price, S. A., Burbrink, F. T., Friedman, M. & Wainwright, P. C. (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences* **110**, 12738–12743.

Near, T. J., Dornburg, A. & Friedman, M. (2014). Phylogenetic relationships and timing of diversification in gonorynchiform fishes inferred using nuclear gene DNA sequences (Teleostei: Ostariophysi). *Molecular Phylogenetics and Evolution* **80**, 297–307.

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

Nelson, G. J. (1969). Infraorbital bones and their bearing on the phylogeny and geography of osteoglossomorph fishes. *American Museum Novitates* **2394**, 1–37

Nelson, J. S., Grande, T. C. & Wilson, M. V. H. (2016). *Fishes of the World* (5th edit.). John Wiley & Sons, Hoboken.

*Neuman, A. G. & Brinkman, D. B. (2005). Fishes of the fluvial beds. In: Currie, P. J. & Koppelhus, E. B. (eds.) *Dinosaur Provincial Park: A Spectacular Ancient Ecosystem Revealed*, 167–185. Indiana University Press, Bloomington.

*Newbrey, M. G. & Ashworth, A. C. (2004). A fossil record of colonization and response of lacustrine fish populations to climate change. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 1807–1816.

*Newbrey, M. G. & Bozek, M. A. (2000). A new species of *Joffrichthys* (Teleostei: Osteoglossidae) from the Sentinel Butte Formation (Paleocene) of North Dakota, USA. *Journal of Vertebrate Paleontology* **20**, 12–20.

Newbrey, M. G., Brinkman, D. B., Winkler, D. A., Freedman, E. A., Neuman, A. G., Fowler, D. W. & Woodward, H. N. (2013). Teleost centrum and jaw elements from the Upper Cretaceous Nemegt Formation (Campanian-Maastrichtian) of Mongolia and a re-identification of the fish centrum found with the theropod *Raptorex kreigsteini*. In: Arratia, G., Schultze, H.-P. & Wilson, M. V. H. (eds.) *Mesozoic Fishes 5 – Global Diversity and Evolution*, 291–303. Verlag Dr. F. Pfeil, München.

Newbrey, M. G., Murray, A. M., Wilson, M. V. H., Brinkman, D. B. & Neuman, A. G. (2009). Seventy-five-million-year-old tropical tetra-like fish from Canada tracks Cretaceous global warming. *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 3829–3833.

*Newbrey, M. G., Wilson, M. V. H. & Ashworth, A. C. (2007). Centrum growth patterns provide evidence for two small taxa of Hiodontidae in the Cretaceous Dinosaur Park Formation. *Canadian Journal of Earth Sciences* **44**, 721–732.

*Nolf, D. (2013). *The diversity of fish otoliths, past and present*. Koninklijk Belgisch Instituut voor Natuurwetenschappen.

*Nolf, D. & Cappetta, H. (1976). Observations nouvelles sur les otolithes des téléostéens du calcaire Grossier (Eocène du Bassin de Paris). *Geobios* **9**, 251–277.

Nolf, D., Rana, R. S. & Prasad, G. V. (2008). Late Cretaceous (Maastrichtian) fish otoliths from the Deccan intertrappean beds, India: a revision. *Bulletin: Sciences de la Terre* **78**, 239–259.

*Nolf, D. & Stringer, G. L. (1996). Cretaceous fish otoliths–a synthesis of the North American record. In: Arratia, G. & Viohl, G. (eds.), *Mesozoic Fishes. Systematics and Paleoecology*, 433–459. Verlag Dr. F. Pfeil, München.

Novacek, M. J. & Marshall, L. G. (1976). Early biogeographic history of ostariophysan fishes. *Copeia* **1**, 1–12.

Nowack, J. & Dausmann, K. H. (2015). Can heterothermy facilitate the colonization of new habitats? *Mammal Review* **45**, 117–127.

*O'Connor, P. M., Gottfried, M. D., Stevens, N. J., Roberts, E. M., Ngasala, S., Kapilima, S. & Chami, R. (2006). A new vertebrate fauna from the Cretaceous Red Sandstone Group, Rukwa Rift Basin, southwestern Tanzania. *Journal of African Earth Sciences* **44**, 277–288.

*Ossian, C. R. (1973). Fishes of a Pleistocene lake in South Dakota. *Publications of the Museum – Michigan State University, Paleontological Series* **1**, 101–124.

Ostrowski, S. A. (2012). The teleost ichthyofauna from the Late Cretaceous of Madagascar: systematics, distributions, and implications for Gondwanan biogeography. Unpublished doctoral thesis, Michigan State University, 165 pp.

Otero, O. (2011). Current knowledge and new assumptions on the evolutionary history of the African lungfish, *Protopterus*, based on a review of its fossil record. *Fish and Fisheries* **12**, 235–255.

*Otero, O. & Gayet, M. (2001). Palaeoichthyofaunas from the Lower Oligocene and Miocene of the Arabian Plate: palaeoecological and palaeobiogeographical implications. *Palaeogeography, Palaeoclimatology, Palaeoecology* **165**, 141–169.

Otero, O., Pinton, A., Cappetta, H., Adnet, S., Valentin, X., Salem, M. & Jaeger, J. J. (2015). A fish assemblage from the Middle Eocene from Libya (Dur At-Talah) and the earliest record of modern African fish genera. *PLoS One* **10**, e0144358.

*Otero, O., Pinton, A., Mackaye, H. T., Likius, A., Vignaud, P. & Brunet, M. (2009). First description of a Pliocene ichthyofauna from Central Africa (site KL2, Kolle area, Eastern Djurab, Chad): What do we learn?. *Journal of African Earth Sciences* **54**, 62–74.

*Otero, O., Pinton, A., Mackaye, H. T., Likius, A., Vignaud, P. & Brunet, M. (2010*a*). The early/late Pliocene ichthyofauna from Koro–Toro, Eastern Djurab, Chad. *Geobios* **43**, 241–251.

*Otero, O., Pinton, A., Mackaye, H. T., Likius, A., Vignaud, P. & Brunet, M. (2010*b*). The fish assemblage associated with the Late Miocene Chadian hominid (Toros–Menalla, Western Djurab) and its palaeoenvironmental significance. *Palaeontographica Abteilung A: Paleozoology, Stratigraphy* **292**, 21–51.

Otero, O., Valentin, X. & Garcia, G. (2008). Cretaceous characiform fishes (Teleostei: Ostariophysi) from Northern Tethys: description of new material from the Maastrichtian of Provence (Southern France) and palaeobiogeographical implications. *Geological Society*, *London*, *Special Publications* **295**, 155–164.

*Pan, Y., Fürsich, F. T., Zhang, J., Wang, Y. & Zheng, X. (2015). Biostratinomic analysis of *Lycoptera* beds from the Early Cretaceous Yixian Formation, western Liaoning, China. *Palaeontology* **58**, 537–561.

Parenti, L. R. (1981). A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bulletin of the American Museum of Natural History* **168**, 341–557.

Parham, J. F., Donoghue, P. C., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patané, J. S., Smith, N. D., Tarver, J. E., van Tuinen, M., Yang, Z. *et al.* (2012). Best practices for justifying fossil calibrations. *Systematic Biology* **61**, 346–359.

Parker, A. & Kornfield, I. (1995). Molecular perspective on evolution and zoogeography of cyprinodontid killifishes (Teleostei; Atherinomorpha). *Copeia* **1995**, 8–21.

*Patterson, C. (1975). The distribution of Mesozoic freshwater fishes. *Mémoires du Muséum National d'Histoire Naturelle, Série A, Zoologie* **88**, 156–174.

Patterson, C. (1993). An overview of the early fossil record of acanthomorphs. *Bulletin of Marine Science* **52**, 29–59.

*Patterson, C. & Longbottom, A. E. (1989). An Eocene amiid fish from Mali, West Africa. *Copeia* **1989**, 827–836.

Pedroza, V., Le Roux, J. P., Gutiérrez, N. M. & Vicencio, V. E. (2017). Stratigraphy, sedimentology, and geothermal reservoir potential of the volcaniclastic Cura-Mallín succession at Lonquimay, Chile. *Journal of South American Earth Sciences* **77**, 1–20.

*Peñalver, E. & Gaudant, J. (2010). Limnic food web and salinity of the Upper Miocene Bicorb palaeolake (eastern Spain). *Palaeogeography, Palaeoclimatology, Palaeoecology* **297**, 683–696.

*Pittet, F., Cavin, L. & Poyato-Ariza, F. J. (2010). A new teleostean fish from the early Late Cretaceous (Cenomanian) of SE Morocco, with a discussion of its relationships with ostariophysans. In: Grande, T., Poyato-Ariza, F. J. & Diogo, R. (eds.) *Gonorynchiformes and Ostariophysan Relationships: A Comprehensive Review*, 339–362. Science Publishers, Enfield.

*Pledge, N. S. (1984). A new Miocene vertebrate faunal assemblage from the Lake Eyre Basin: a preliminary report. *The Australian Zoologist* **21**, 345–355.

Pohl, M., Milvertz, F. C., Meyer, A. & Vences, M. (2015). Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of *Pantanodon*. *Vertebrate Zoology* **65**, 37–44.

Potter, I. C., Gill, H. S., Renaud, C. B. & Haoucher, D. (2015). The taxonomy, phylogeny, and distribution of lampreys. In: Potter, M. F. (ed.), *Lampreys: Biology*, *Conservation and Control*, 35–73. Springer Dordrecht.

Poux, C., Chevret, P., Huchon, D., De Jong, W. W. & Douzery, E. J. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Systematic Biology* **55**, 228–244.

Pramuk, J. B., Robertson, T., Sites, J. W. & Noonan, B. P. (2008). Around the world in 10 million years: biogeography of the nearly cosmopolitan true toads (Anura: Bufonidae). *Global Ecology and Biogeography* **17**, 72–83.

*Prasad, G. V. R. (1989). Vertebrate fauna from the infra- and intertrappean beds of Andhra Pradesh: age implications. *Journal of the Geological Society of India* **34**, 161–173.

*Prendergast, M. E. & Lane, P. J. (2010). Middle Holocene fishing strategies in East Africa: zooarchaeological analysis of Pundo, a Kansyore shell midden in northern Nyanza (Kenya). *International Journal of Osteoarchaeology* **20**, 88–112.

Pyron, R. A. (2014). Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. *Systematic Biology* **63**, 779–797.

*Rana, R. S. (1988). Freshwater fish otoliths from the Deccan trap associated sedimentary (Cretaceous–Tertiary transition) beds of Rangapur, Hyderabad, District, Andhra Pradesh, India. *Geobios* **21**, 465–493.

*Rana, R. S., Kumar, K. & Singh, H. (2006). Palaeocene vertebrate fauna from the Fatehgarh Formation of Barmer District, Rajasthan, western India. In: Sinha, D. K. (ed.) *Micropalaeontology: Application in Stratigraphy and Paleoceanography,* 113–130. Narosa Publishing House, New Delhi.

Ree, R. H. & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* **45**, 741–749.

*Reichenbacher, B. (1988). Die fischfauna der Kirchberger Schichten (Unter–Miozän) an der typuslokalität Illerkirchberg bei Ulm. *Stuttgarter Beiträge zur Naturkunde* **139**, 1–53.

*Reichenbacher, B. (1989). Feinstratigraphische gliederung der Kirchberger Schichten (Unter–Miozän) an der typuslokalität Illerkirchberg bei Ulm. *Geologica Bavarica* **94**, 135–177.

*Reichenbacher, B. (1993). Mikrofaunen, paläogeographie und biostratigraphie der Miozänen Brack–und Süßwassermolasse in der westlichen Paratethys unter besonderer berücksichtigung der fisch-otolithen. *Senckenbergiana Lethaea* **73***,* 277–374.

*Reichenbacher, B. (2000). Das brackisch–lakustrine Oligozän und Unter Miozän im Mainzer Becken und Hanauer Becken: Fischfaunen, Paläoökologie, Biostratigraphie, Paläogeographie. *Courier Forschungsinstitut Senckenberg* **222**, 113–127.

*Reichenbacher, B. (2004). A partly endemic euryhaline fish fauna (otoliths, teeth) from the Early Miocene of the Aix–Basin (Provence, southern France). *Courier Forschungsinstitut Senckenberg* **246**, 113–128.

*Reichenbacher, B., Alimohammadian, H., Sabouri, J., Haghfarshi, E., Faridi, M., Abbasi, S., Matzke-Karasz, R., Fellin, M. G., Carnevale, G., Schiller, W. & Vasilyan, D. (2011). Late Miocene stratigraphy, palaeoecology and palaeogeography of the Tabriz basin (NW Iran, Eastern Paratethys). *Palaeogeography, Palaeoclimatology, Palaeoecology* **311**, 1–18.

*Reichenbacher, B., Berger, J. P. & Weidmann, M. (1996). Charophytes et otolithes de la Molasse d'eau douce inferieure oligocene de Moutier (Jura suisse). *Neues Jahrbuch fur Geologie und Palaontologie–Abhandlungen* **202**, 63–94

*Reichenbacher, B. & Gaudant, J. (2003). On *Prolebias meyeri* (Agassiz)(Teleostei, Cyprinodontiformes) from the Oligo–Miocene of the Upper Rhinegraben area, with the establishment of a new genus and a new species. *Eclogae Geologicae Helvetiae* **96**, 509–520.

Reichenbacher, B. & Kowalke, T. (2009). Neogene and present-day zoogeography of killifishes (*Aphanius* and *Aphanolebias*) in the Mediterranean and Paratethys areas. *Palaeogeography*, *Palaeoclimatology*, *Palaeoecology* **281**, 43–56.

*Reichenbacher, B. & Prieto, J. (2006). Lacustrine fish faunas (Teleostei) from the Karpatian of the northern Alpine Molasse Basin, with a description of two new species of *Prolebias* Sauvage. *Palaeontographica Abteilung A* **278**, 87–95.

*Reichenbacher, B. & Sienknecht, U. (2001). Allopatric divergence and genetic diversity of recent *Aphanius iberus* and fossil *Prolebias meyeri* (Teleostei, Cyprinodontidae) from southwest and western Europe, as indicated by otoliths. *Geobios* **34**, 69–83.

*Reichenbacher, B. & Weidmann, M. (1992). Fisch–otolithen aus der Oligo-Miozänen molasse der West-Schweiz und der Haute-Savoie (Frankreich). *Stuttgarter Beitr Naturkunde B* **184**, 1–83.

*Richter, M. (1984). Dental histology of a characoid fish from the Plio-Pleistocene of Acre, Brazil. *Zoologica Scripta* **13**, 69–79.

*Roberts, T. R. (1975*a*). Characoid fish teeth from Miocene deposits in the Cuenca Basin, Ecuador. *Journal of Zoology* **175**, 259–271.

*Roberts, T.R. (1975*b*). Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society* **57**, 249–319.

*Robertson, G. M. (1943). *Fundulus sternbergi*, a Pliocene fish from Kansas. *Journal of Paleontology* **17**, 305–306.

*Roe, L. J. (1991). Phylogenetic and ecological significance of Channidae (Osteichthyes, Teleostei) from the early Eocene Kuldana Formation of Kohat, Pakistan. *Contributions from the Museum of Paleontology, University of Michigan* **28**, 93–100.

Romano, C., Koot, M. B., Kogan, I., Brayard, A., Minikh, A. V., Brinkmann, W., Bucher, H. & Kriwet, J. (2016). Permian–Triassic Osteichthyes (bony fishes): diversity dynamics and body size evolution. *Biological Reviews* **91**, 106–147.

Ronquist, F., Lartillot, N. & Phillips, M. J. (2016). Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150136.

Ronquist, F. & Sanmartín, I. (2011). Phylogenetic methods in biogeography. *Annual Review of Ecology*, *Evolution*, *and Systematics* **42**, 441–464.

Rosen, D. E. & Greenwood, P. H. (1976). A fourth Neotropical species of synbranchid eel and the phylogeny and systematics of synbranchiform fishes. *Bulletin of the American Museum of Natural History* **157**, 1–70.

Rota, J., Peña, C. & Miller, S. E. (2016). The importance of long-distance dispersal and establishment events in small insects: historical biogeography of metalmark moths (Lepidoptera, Choreutidae). *Journal of Biogeography* **43**, 1254–1265.

*Rozefelds, A. C., Dettmann, M. E., Clifford, H. T. & Lewis, D. (2016). Macrofossil evidence of early sporophyte stages of a new genus of water fern *Tecaropteris* (Ceratopteridoideae: Pteridaceae) from the Paleogene Redbank Plains Formation, southeast Queensland, Australia. *Alcheringa: An Australasian Journal of Palaeontology* **40**, 1–11.

Rüber, L., Britz, R. & Zardoya, R. (2006). Molecular phylogenetics and evolutionary diversification of labyrinth fishes (Perciformes: Anabantoidei). *Systematic Biology* **55**, 374–397.

*Rubilar, A. (1994). Diversidad ictiológica en depósitos continentales miocenos de la Formación Cura-Mallín, Chile (37-39 S): implicancias paleogeográficas. *Andean Geology* **21**, 3–29.

*Rubilar, A. & Abad, E. (1990). *Percichthys sylviae* sp. nov. del Terciario de los Andes Sur-Centrales de Chile (Pisces, Perciformes, Percichthyidae). *Andean Geology* **17**, 197–204.

*Rückert-Ülkümen, N. (2006). Otolithen aus dem Mio–Pliozän von Yalova bei Istanbul, Türkei. *Neues Jahrbuch für Geologie Und Palaeontologie. Monatshefte* **10**, 577–594.

*Rückert-Ülkümen, N., Kaya, O. & Hottenrott, M. (1993). Neue beiträge zur Tertiär–Stratigraphie und otolithenfauna der Umgebung von Istanbul (Küçükçekmece–und

Büyükçekmece See), Türkei. *Mitteilungen der Bayerischen Staatssammlung für Paläontologie und Historische Geologie* **33**, 51–89.

*Rückert-Ülkümen, N. & Müller, E. D. (1999). Larven von *Aphanius sp*.(Teleostei, Cyprinodontidae) aus dem jungtertiären Seeton von Wemding (Nördlinger Ries). *Mitteilungen der Bayerischen Staatsammlung für Paläontologie und Historische Geologie* **39**, 51–68.

*Sach, V., Gaudant, J., Reichenbacher, B. & Böhme, M. (2003). Die Fischfaunen der Fundstellen Edelbeuren–Maurerkopf und Wannenwaldtobel 2 (Miozän, Obere Süβwassermolasse, Süddeutschland). *Stuttgarter Beiträge zur Naturkunde. Serie B, Geologie und Paläontologie* **334**, 1–25.

*Sahni, A. & Bajpai, S. (1988). Cretaceous-Tertiary boundary events: The fossil vertebrate, palaeomagnetic and radiometric evidence from peninsular India. *Journal of Geological Society of India* **32**, 382–396.

*Sahni, A. & Khare, S. K. (1977). A middle Siwalik fish fauna from Ladhyani (Haritalyangar), Himachal Pradesh. *Biological Memoires* **2**, 187–221.

*Sahni, A., Srikantia, S. V., Ganesan, T. M. & Wangdus, C. (1984). Tertiary fishes and molluscs from the Kuksho Formation of the Indus Group, near Nyoma, Ladakh. *Geological Society of India* **25**, 744–747.

*Saitô, K. (1936). Mesozoic leptolepid fishes from Jehol and Chientao, Manchuria. *Report of the Scientific Expedition, Manchoukou* **2**, 1–23.

Sallan, L. C. (2014). Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. *Biological Reviews* **89**, 950–971.

Samonds, K. E., Godfrey, L. R., Ali, J. R., Goodman, S. M., Vences, M., Sutherland, M. R., Irwin, M. T. & Krause, D. W. (2012). Spatial and temporal arrival patterns of Madagascar's vertebrate fauna explained by distance, ocean currents, and ancestor type. *Proceedings of the National Academy of Sciences* **109**, 5352–5357.

Sanders, M. (1934). Die Fossilen Fische der Alttertiären Süsswasserablagerungen aus Mittel-Sumatra. *Verhandelingen van het Geologisch-Mijnbouwkundig Genootschap voor Nederland en Koloni'n. Geologische Series* **11**, 1–144.

*Sauvage, H. E. (1869). Note sur les poissons du calcaire de Ronzon, près Le Puy-en-Velay. *Bulletin de la Société Géologique de France* **2**, 1069–1075.

*Sauvage, H. E. (1874). Notice sur les poissons tertiaires de l'Auvergne. *Bulletin de la Société d'Histoire Naturelle de Toulouse* **8**, 171–198.

*Sauvage, H. E. (1880). Notice sur les poissons tertiaires de Céreste (Basses–Alpes). *Bulletin de la Société Géologique de France* **3**, 439–451.

*Schaal, S. (1984). Oberkretazische Osteichthyes, Knochenfische, aus dem Bereich von Bahariya und Kharga, Ägypten, und ihre Aussagen zur Palökologie und Stratigraphie. *Berliner Geowissenchaftliche Abhandlungen A* **53**, 1–79.

*Schaeffer, B. (1947). An Eocene serranid from Patagonia. *American Museum Novitates* **1331**, 1–10.

*Schäfer, P., Kälin, D. & Reichenbacher, B. (2005). Beiträge zur Ostracoden–und Foraminiferen–Fauna der Unteren Süßwassermolasse in der Schweiz und in Savoyen (Frankreich). 2. La Chaux (Kanton Waadt, Schweiz). *Senckenbergiana Lethaea* **85**, 95–117.

Scheben, A., Bechteler, J., Lee, G. E., Pócs, T., Schäfer-Verwimp, A. & Heinrichs, J. (2016). Multiple transoceanic dispersals and geographical structure in the pantropical leafy liverwort *Ceratolejeunea* (Lejeuneaceae, Porellales). *Journal of Biogeography* **43**, 1739–1749.

Schultze, H.-P. (1991). Lungfish from the El Molino (Late Cretaceous) and Santa Lucia (Early Paleocene) formations in southcentral Bolivia. *Revista Técnica de YPFB* **12**, 441–448.

Schultze, H.-P. (2004). Mesozoic sarcopterygians. In: Arratia, G. & Tintori, A. (eds.) *Mesozoic Fishes 3 – Systematics*, *Paleoenvironments and Biodiversity*, 463–492. Verlag Dr. F. Pfeil, München.

*Schulz-Mirbach, T. & Reichenbacher, B. (2008). Fossil *Aphanius* (Teleostei, Cyprinodontiformes) from southwestern Anatolia (Turkey): a contribution to the evolutionary history of a hotspot of freshwater biodiversity. *Geodiversitas* **30**, 577–592.

*Schwartz, H. L. (1983). *Paleoecology of late Cenozoic fishes from the Turkana Basin, northern Kenya*. Unpublished doctoral dissertation, University of California, Santa Cruz.

Schwarzhans, W. (2018). A review of Jurassic and Early Cretaceous otoliths and the development of early morphological diversity in otoliths. *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen*, **287**, 75–121.

*Schwarzhans, W., Scofield, R. P., Tennyson, A. J., Worthy, J. P. & Worthy, T. H. (2012). Fish remains, mostly otoliths, from the non-marine early Miocene of Otago, New Zealand. *Acta Palaeontologica Polonica* **57**, 319–350.

Scotese, C. R. (2014). *PALEOMAP Atlas for ArcGIS*. PALEOMAP Project, Evanston.

*Scudder, S. J., Simons, E. H. & Morgan, G. S. (1995). Chondrichthyes and Osteichthyes from the early Pleistocene Leisey shell pit local fauna, Hillsborough country, Florida. *Bulletin of the Florida Museum of Natural History* **37**, 251–272.

Setiamarga, D. H., Miya, M., Yamanoue, Y., Mabuchi, K., Satoh, T. P., Inoue, J. G. & Nishida, M. (2008). Interrelationships of Atherinomorpha (medakas, flyingfishes, killifishes, silversides, and their relatives): the first evidence based on whole mitogenome sequences. *Molecular Phylogenetics and Evolution* **49**, 598–605.

Sferco, E., López-Arbarello, A. & Báez, A. M. (2015). Phylogenetic relationships of †*Luisiella feruglioi* (Bordas) and the recognition of a new clade of freshwater teleosts from the Jurassic of Gondwana. *BMC Evolutionary Biology* **15**, 268.

Sharma, P. P. & Wheeler, W. C. (2013). Revenant clades in historical biogeography: the geology of New Zealand predisposes endemic clades to root age shifts. *Journal of Biogeography* **40**, 1609–1618.

*Shen, M. (1989). *Eohiodon* from China and the distribution of osteoglossomorphs. *Vertebrata PalAsiatica* **27**, 237–247.

Signore, M., Pede, C., Bucci, E., & Barbera, C. (2006). First report of the genus *Cladocyclus* in the Lower Cretaceous of Pietraroja (Southern Italy). *Bollettino della Società Paleontologica Italiana* **45**, 141.

*Sigé, B. (1968). Dents de micromammiferes et fragments de coquilles d'oeufs de dinosauriens dans la faune de vertébrés du Crétacé supérieur de Laguna Umayo (Andes péruviennes). *Comptes Rendus de l'Academie des Sciences*, *Paris* **267**, 1495–1498.

*Sigé, B., Sempere, T., Butler, R. F., Marshall, L. G. & Crochet, J. Y. (2004). Age and stratigraphic reassessment of the fossil–bearing Laguna Umayo red mudstone unit, SE Peru, from regional stratigraphy, fossil record, and paleomagnetism. *Geobios* **37**, 771–794.

Silva Santos, R. (1987). *Lepidosiren megalos* n. sp. do Terciário do Estado do Acre—Brasil. *Anais da Academia Brasileira de Ciências* **59**, 375–384.

*Silva Santos, R. (1988). *Laeliichthys ancestralis*, novo gênero e espécie de Osteoglossiformes do Aptiano da Formaçao Areado, estado de Minas Gerais, Brasil. *MME–DNPM, Geologia 27, Paleontologia e estratigrafia* **2**, 161–167.

*Silva Santos, R. (1994). Ictiofáunula da Formação Codó, Cretáceo Inferior, com a descrição de um novo táxon–*Codoichthys carnavalii* (Pisces-Teleostei). *Anais da Academia Brasileira de Ciências* **66**, 131–144.

Silvestro, D., Zizka, A., Bacon, C. D., Cascales-Minana, B., Salamin, N. & Antonelli, A. (2016). Fossil biogeography: a new model to infer dispersal, extinction and sampling from palaeontological data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150225.

*Singh, R. R., Sharma, K. M. & Patnaik, R. (2014). First record of fossil otoliths from the Siwaliks of India. *Earth Science India* **7**, 49–54.

*Smith, C. L. (1962). Some Pliocene fishes from Kansas, Oklahoma, and Nebraska. *Copeia* **1962**, 505–520.

Smith, G. R. (1981). Late Cenozoic freshwater fishes of North America. *Annual Review of Ecology and Systematics* **12**, 163–193.

*Smith, G. R. & Lundberg, J. G. (1972). The Sand Draw fish fauna. In: Skinner, M. F. & Hibbard, C. W. (eds.) *Early Pleistocene Preglacial and Glacial Rocks and Faunas of North-central Nebraska*, 40–54. American Museum of Natural History, New York.

*Smith, M. L., Cavender, T. M. & Miller, R. R. (1975). Climatic and biogeographic significance of a fish fauna from the late Pliocene-early Pleistocene of the Lake Chapala basin (Jalisco, Mexico). In: Smith, G. R. & Friedland, N. E. (eds.), *Studies on Cenozoic Paleontology and Stratigraphy in Honor of Claude W. Hibbard. University of Michigan Papers in Paleontology* **12**, 29–38.

Sparks, J. S. & Smith, W. L. (2004). Phylogeny and biogeography of the Malagasy and Australasian rainbowfishes (Teleostei: Melanotaenioidei): Gondwanan vicariance and evolution in freshwater. *Molecular Phylogenetics and Evolution* **33**, 719–734.

Sparks, J. S. & Smith, W. L. (2005). Freshwater fishes, dispersal ability, and nonevidence: "Gondwana life rafts" to the rescue. *Systematic Biology* **54**, 158–165.

*Steurbaut, E. (1978). Otolithes de téléostéens de quelques formations d'âge Aquitanien du Midi de la France. *Bulletin van de Belgische Vereniging voor Geologie* **87**, 179–188.

*Steurbaut, E. (1980). Deux nouveaux gisements a otolithes de teleosteens, dont une espece nouvelle, dans l'Aquitanien continental du Midi de la France. *Geobios* **13**, 111–114.

*Stevens, W. N., Claeson, K. M. & Stevens, N. J. (2016). Alestid (Characiformes: Alestidae) fishes from the Late Oligocene Nsungwe Formation, Rukwa Rift Basin, of Tanzania. *Journal of Vertebrate Paleontology* **36**, e1180299.

*Stewart, K. M. (1989). *Fishing sites of North and East Africa in the Late Pleistocene and Holocene: Environmental change and human adaptation*. British Archaeological Reports International Series **521**, 273 pp.

*Stewart, K. M. (1990). Fossil fish from the Upper Semliki. *Virginia Museum of Natural History Memoir* **1**, 141–162.

*Stewart, K. M. (1997*a*). A new species of *Sindacharax* (Teleostei: Characidae) from Lothagam, Kenya, and some implications for the genus. *Journal of Vertebrate Paleontology* **17**, 34–38.

*Stewart, K. M. (1997*b*). Fossil fish from the Manonga Valley, Tanzania. In: Harrison, T. (ed.) *Neogene paleontology of the Manonga Valley, Tanzania: A window into the evolutionary history of East Africa*, 333–349. Plenum Press, NY.

*Stewart, K. M. (2001). The freshwater fish of Neogene Africa (Miocene–Pleistocene): systematics and biogeography. *Fish and Fisheries* **2**, 177–230.

*Stewart, K. M. (2003*a*). Fossil fish remains from Mio-Pliocene deposits at Lothagam, Kenya. In: Leakey, M. G. & Harris, J. M. (eds.) *Lothagam: the dawn of humanity in eastern Africa*, 75–111. Columbia University Press, NY.

*Stewart, K. M. (2003*b*). Fossil fish remains from the Pliocene Kanapoi site, Kenya. *Contributions in Science, Natural History Museum of Los Angeles County* **498**, 21–38.

*Stewart, K. M. (2009). Fossil fish from the Nile River and its southern basins. In: Dumont, H. J. (ed.) *The Nile: Origin, Environments, Limnology and Human Use*, 677–704. Springer Netherlands.

*Stewart, K. M. & Murray, A. M. (2008). Fish remains from the Plio-Pleistocene Shungura Formation, Omo River basin, Ethiopia. *Geobios* **41**, 283–295.

*Stewart, K. M. & Murray, A. M. (2013). Earliest fish remains from the Lake Malawi Basin, Malawi, and biogeographical implications. *Journal of Vertebrate Paleontology* **33**, 532–539.

*Stinton, F. C. (1977). Fish otoliths from the English Eocene, II. *Palaeontographical Society Monographs* **548**, 57–126.

*Stinton, F. C. & Kissling, D. (1968). Quelques otolithes de téléostéens de la Molasse oligocène de Suisse occidentale. *Comptes Rendus des Seances, Societe de Physique e d'Histoire Naturelle* **3**, 140–154.

Strauss, D. & Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology* **21**, 411–427.

*Stromer, E. V. (1910). Über das Gebiss der Lepidosirenidae und die Verbreitung tertiärer und mesozoischer Lungenfische. *Festschrift zum 60 Geburtstage der R. Hertwigs* **2**, 612–614.

*Su, D. Z. (1986). The discovery of a fossil osteoglossid fish in China. *Vertebrata PalAsiatica* **24**, 10.

*Su, D. Z. (1991). A new fossil hiodontoid fish from Fuxin Group of western Liaoning, China. *Vertebrata PalAsiatica* **29**, 38–45.

*Su, D. Z. (1992). On teleostean fossils from Nieerku Formation of Eastern Liaoning and the generic status of *Lycoptera longicephalus*. *Vertebrata PalAsiatica* **30**, 54–70.

Su, D. Z. (1994). New Early Jurassic actinopterygians from Weixin, Yunnan. *Vertebrata PalAsiatica* **32**, 151–165.

Sullivan, J. P., Lavoué, S. & Hopkins, C. D. (2016). *Cryptomyrus*: a new genus of Mormyridae (Teleostei, Osteoglossomorpha) with two new species from Gabon, West-Central Africa. *ZooKeys* **561**, 117–150.

*Sytchevskaya, E. K. (1986). Paleogene freshwater fish fauna of the USSR and Mongolia. *Proceedings of the Joint Soviet–Mongolian Paleontological Expedition* **29**, 1–157.

*Sytchevskaya, E. K. & Yakovlev, V. N. (1985). Fishes. In: Rasnitsin, A. P. (ed.), *Jurassic Continental Biocenoses of Southern Siberia and Adjacent Territories*, 132–136. Nauka, Moscow.

Tabuce, R. & Marivaux, L. (2005). Mammalian interchanges between Africa and Eurasia: an analysis of temporal constraints on plausible anthropoid dispersals during the Paleogene. *Anthropological Science* **113**, 27–32.

*Takai, F. (1944). A monograph on the lycopterid fishes from the Mesozoic of eastern Asia. *Journal of the Faculty of Science, Tokyo University, Series 2* **6**, 207–270.

*Taverne, L. (1978). Osteologie, phylogénèse et systématique des Téléostéens fossiles et actuels de super–ordre des Ostéoglossomorphes. Deuxième partie. Ostéologie des genres *Phareodus*, *Phareoides*, *Brychaetus*, *Musperia*, *Pantodon*, *Singida*, *Notopterus*, *Xenomystus* et *Papyrocranus*. *Mémoires de la Classe des Sciences, Académie Royale de Belgique, Collection in–8°, 2e série* **42**, 1–213.

Taverne, L. (1979). Osteologie, phylogénèse et systématique des Téléostéens fossiles et actuels de super-ordre des Ostéoglossomorphes. Troisième partie. Evolution des structures ostéologiques et conclusions générales relatives à la phylogénèse et à la systématique du super-ordre. Addendum. *Mémoires de la Classe des Sciences*, *Académie Royale de Belgique*, *Collection in-8°*, *2e série* **43**, 1–168.

*Taverne, L. (1998). Les ostéoglossomorphes marins de l'Éocène du Monte Bolca (Italie): *Monopteros* Volta 1796, *Thrissopterus* Heckel, 1856 et *Foreyichthys* Taverne, 1979. Considérations sur la phylogénie des téléostéens ostéoglossomorphes. *Studi e Ricerche sui Giacimenti Terziari di Bolca, Miscellanea Paleontologica* **7**, 67–158.

Taverne, L. (2003). Les poissons crétacés de Nardò. *Sorbinicharax verraesi* gen. sp. nov.(Teleostei, Ostariophysi, Otophysi, Characiformes). *Bollettino del Museo Civico di Storia Naturale di Verona* **27**, 29–45.

*Taverne, L. (2004). On a complete hyomandibular of the Cretaceous Moroccan notopterid *Palaeonotopterus greenwoodi* (Teleostei, Osteoglossomorpha). *Stuttgarter Beiträge zur Naturkunde Serie B (Geologie und Paläontologie)* **348**, 1–7.

*Taverne, L. (2009*a*). New insights on the osteology and taxonomy of the osteoglossid fishes *Phareodus*, *Brychaetus* and *Musperia* (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* **79**, 175–190.

*Taverne, L. (2009*b*). On the presence of the osteoglossid genus *Scleropages* in the Paleocene of Niger, Africa (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* **79**, 161-167.

*Taverne, L. (2009*c*). *Ridewoodichthys*, a new genus for *Brychaetus caheni* from the marine Paleocene of Cabinda (Africa): re-description and comments on its relationships within the Osteoglossidae (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* **79**, 147–153.

*Taverne, L. (2015). On the presence of a second osteoglossid fish (Teleostei, Osteoglossiformes) in the continental Lower Cretaceous of the Democratic Republic of Congo (Central Africa). *Geo-Eco-Trop* **39**, 247–254.

Taverne, L. (2016). *Chanopsis lombardi* (Teleostei, Osteoglossiformes) from the continental Lower Cretaceous of the Democratic Republic of Congo. Comments on the evolution of the caudal skeleton within osteoglossiform fishes. *Geologica Belgica* **19**, 291–301.

*Taverne, L. & Capasso, L. (2012). Osteology and relationships of *Prognathoglossum kalassyi* gen. & sp. nov.(Teleostei, Osteoglossiformes, Pantodontidae) from the marine Cenomanian (Upper Cretaceous) of En Nammoura (Lebanon). *Cybium* **36**, 563–574.

*Taverne, L., Kumar, K. & Rana, R. S. (2009). Complement to the study of the Indian Paleocene osteoglossid fish genus *Taverneichthys* (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* **79**, 155–160.

*Taverne, L. & Maisey, J. G. (1999). A notopterid skull (Teleostei, Osteoglossomorpha) from the continental early Cretaceous of southern Morocco. *American Museum Novitates* **3260**, 1–12.

*Taverne, L., Nolf, D. & Folie, A. (2007). On the presence of the osteoglossid fish genus *Scleropages* (Teleostei, Osteoglossiformes) in the continental Paleocene of Hainin (Mons Basin, Belgium). *Belgian Journal of Zoology* **137**, 89.

*Thomas, H., Roger, J., Sen, S., Bourdillon-De-Grissac, C. & Al-Sulaimani, Z. (1989). Découverte de vertébrés fossiles dans l'Oligocène inférieur du Dhofar (Sultanat d'Oman). *Geobios* **22**, 101–120.

*Toledo, C. E. V. & Bertini, R. J. (2005). Occurrences of the fossil Dipnoiformes in Brazil and its stratigraphic and chronological distributions. *Revista Brasileira de Paleontologia* **8**, 47–56.

*Trapani, J. (2008). Quaternary fossil fish from the Kibish formation, Omo Valley, Ethiopia. *Journal of Human Evolution* **55**, 521–530.

Tse, T. K., Pittman, M. & Chang, M. M. (2015). A specimen of *Paralycoptera* Chang & Chou 1977 (Teleostei: Osteoglossoidei) from Hong Kong (China) with a potential Late Jurassic age that extends the temporal and geographical range of the genus. *PeerJ* **3**, e865.

*Tulip, J. R., Taylor, G. & Truswell, E. M. (1982). Palynology of Tertiary Lake Bunyan, Cooma, New-South-Wales. *BMR Journal of Australian Geology & Geophysics* **7**, 255–268.

Turko, A. J. & Wright, P. A. (2015). Evolution, ecology and physiology of amphibious killifishes (Cyprinodontiformes). *Journal of Fish Biology* **87**, 815–835.

Turner, S. (1982). A catalogue of fossil fish in Queensland. *Memoirs of the Queensland Museum* **20**, 599–611.

*Turner, S. & Long, J. (2015). The Woodward factor: Arthur Smith Woodward's legacy to geology in Australia and Antarctica. *Geological Society, London, Special Publications* **430**, 466–481.

*Unmack, P. J. (1999). *Biogeography of Australian freshwater fishes*. MSc thesis, Arizona State University, 156 pp.

Upchurch, P. & Hunn, C. A. (2002). “Time”: the neglected dimension in cladistic biogeography? *Geobios* **35**, 277–286.

*Uyeno, T. & Miller, R. R. (1962). Relationships of *Empetrichthys erdisi*, a Pliocene cyprinodontid fish from California, with remarks on the Fundulinae and Cyprinodontinae. *Copeia* **1962**, 520–532.

*Van Couvering, J. A. H. (1977). Early records of freshwater fishes in Africa. *Copeia* **1977**, 163–166.

Van Couvering, J. A. H. (1982). Fossil cichlid fish of Africa. *Special Papers in Palaeontology* **29**. Palaeontological Association, London.

*Van Devender, T. R., Rea, A. M. & Smith, M. L. (1985). The Sangamon interglacial vertebrate fauna from Rancho la Brisca, Sonora, Mexico. *Transactions of the San Diego Society of Natural History* **21**, 23–55.

*Van Neer, W. (1986). Some notes on the fish remains from Wadi Kubbaniya (Upper Egypt, late Paleolithic). In: Brinkhuizen, D. C. & Clason, A. T. (eds.) *Fish and archaeology: Studies in osteometry, taphonomy, seasonality, and fishing methods.* British Archaeological Reports International Series **294**, 103–13.

*Van Neer, W. (1992). New late Tertiary fish fossils from the Sinda region, eastern Zaire. *African Study Monographs, Suppl. 17*, 27–47.

*Van Neer, W. (1993). Fish remains from the last interglacial at Bir Tarfawi (eastern Sahara, Egypt). In: Wendorf, F., Schild, R. & Close, A. E. (eds.), *Egypt During the Last Interglacial*, 144–155. Plenum Press, NY.

*Van Neer, W. (1994). Cenozoic fish fossils from the Albertine Rift Valley in Uganda. Geology and Palaeontology of the Albertine Rift Valley, Uganda–Zaire, II: Paleobiology. *CIFEG Occasional Publication, 1994/29*, 89–127.

*Van Neer, W. & Gayet, M. (1988). Étude des poissons en provenance des sites holocènes du bassin de Taoudenni–Araouane (Mali). *Bulletin du Muséum National d'Histoire Naturelle. Section C, Sciences de la Terre, Paléontologie, Géologie, Minéralogie* **10**, 343–383.

*Van Neer, W. & Uerpmann, H. P. (1989). Palaeoecological significance of the Holocene faunal remains of the B.O.S.–Missions. In: Kuper, R. (ed.) *Forschungen zur Umweltgeschichte der Ostsahara, Africa Praehistorica 2*, 307–341. Heinrich-Barth-Institut, Koln.

*Vasilyan, D., Reichenbacher, B. & Carnevale, G. (2009). A new fossil *Aphanius* species from the Upper Miocene of Armenia (Eastern Paratethys). *Paläontologische Zeitschrift* **83**, 511–519.

*von Salis, K. (1967). Geologische und sedimentologische Untersuchungen in Molasse und Quartär südöslich Wolhusen (Entlebuch, Kt. Luzern). *Mitteilungen der Naturforschenden Gesellschaft Luzern XXI*, 1–106.

Wang, J. & Yang, G. (2011). The complete mitogenome of the snakehead *Channa argus* (Perciformes: Channoidei): Genome characterization and phylogenetic implications. *Mitochondrial DNA* **22**, 120–129.

Waters, J. M. & Wallis, G. P. (2001). Cladogenesis and loss of the marine life-history phase in freshwater galaxiid fishes (Osmeriformes: Galaxiidae). *Evolution* **55**, 587–597.

*Weems, R. E. & Horman, S. R. (1983). Teleost fish remains (Osteoglossidae, Blochiidae, Scombridae, Triodontidae, Diodontidae) from the Lower Eocene Nanjemoy Formation of Maryland. *Proceedings of the Biological Society of Washington* **96**, 38–49.

*Weiler, W. (1963). Die Fischfauna des Tertiärs im oberrheinischen Graben, des-Mainzer Beckens, des unteren Maintals und der Wetterau,-unter besonderer Berücksichtigung des Untermiozäns. *Abhandlungen Senckenbergische naturforschende Gesellschaft* **504**, 1–75.

*Weiss, F. E., Malabarba, L. R. & Malabarba, M. C. (2012). Phylogenetic relationships of *Paleotetra*, a new characiform fish (Ostariophysi) with two new species from the Eocene-Oligocene of south-eastern Brazil. *Journal of Systematic Palaeontology* **10**, 73–86.

Weiss, F. E., Malabarba, M. C. & Malabarba, L. R. (2014). A new stem fossil characid (Teleostei: Ostariophysi) from the Eocene-Oligocene of southeastern Brazil. *Neotropical Ichthyology* **12**, 439–450.

*Werner, C. (1993). Late Cretaceous continental vertebrate faunas of Niger and northern Sudan. In: Thorweihe, U. & Schandelmeier, H. (eds.) *Geoscientific research in Northeast Africa*, 401–405. Balkema, Rotterdam.

Werner, C. (1994). Die kontinentale Wirbeltierfauna aus der unteren Oberkreide des Sudan (Wadi Milk Formation). *Berliner Geowissenschaftliche Abhandlungen E (B. Krebs-Festschrift)* **13**, 221–249.

*Werner, C. & Gayet, M. (1997). New fossil Polypteridae from the Cenomanian of Sudan. An evidence of their high diversity in the early Late Cretaceous. *Cybium* **21**, 67–81.

*West, R. M. (1984). Siwalik faunas from Nepal: Paleoecologic and paleoclimatic implications. In: Whyte, R. O. (ed.) *The evolution of the East Asian environment*, *Vol. 2*, 724–744. University of Hong Kong, Hong Kong.

*White, E. I. (1927). On a fossil Cyprinodont from Ecuador. *Journal of Natural History* **20**, 519–522.

*White, E. I. (1937). The fossil fishes of the terraces of Lake Bosumtwi, Ashanti. *Bulletin of the Gold Coast Geographical Association* **8**, 47–58.

Wiley, E. O. & Johnson, G. D. (2010). A teleost classification based on monophyletic groups. In: Nelson, J. S., Schultze, H.-P. & Wilson, M. V. H. (eds.) *Origin and Phylogenetic Interrelationships of Teleosts*, 123–182. Verlag Dr. F. Pfeil, München.

*Williams, D. L. G. (1980). Catalogue of Pleistocene vertebrate fossils and sites in South Australia. *Transactions of the Royal Society of South Australia* **104**, 101–115.

*Wilson, M. V. H. (1977). Middle Eocene freshwater fishes from British Columbia. *Life Sciences Contributions, Royal Ontario Museum* **113**, 1–61.

*Wilson, M. V. H. (1978). *Eohiodon woodruffi* n. sp.(Teleostei, Hiodontidae), from the Middle Eocene Klondike Mountain Formation near Republic, Washington. *Canadian Journal of Earth Sciences* **15**, 679–686.

Wilson, M. V. H. & Murray, A. M. (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.

Wilson, M. V. H. & Williams, R. R. G. (2010). Salmoniform fishes: key fossils, supertree, and possible morphological synapomorphies. In: Nelson, J. S., Schultze, H.-P. & Wilson, M. V. H. (eds.) *Origin and Phylogenetic Interrelationships of Teleosts*, 379–409. Verlag Dr. F. Pfeil, München.

Woodburne, M. O. & Case, J. A. (1996). Dispersal, vicariance, and the Late Cretaceous to early Tertiary land mammal biogeography from South America to Australia. *Journal of Mammalian Evolution* **3**, 121–161.

Wu, F., Miao, D., Chang, M. M., Shi, G. & Wang, N. (2017). Fossil climbing perch and associated plant megafossils indicate a warm and wet central Tibet during the late Oligocene. *Scientific Reports* **7**, 878.

*Xu, G. H. & Chang, M. (2009). Redescription of †*Paralycoptera wui* Chang. & Chou, 1977 (Teleostei: Osteoglossoidei) from the Early Cretaceous of eastern China. *Zoological Journal of the Linnean Society* **157**, 83–106.

*Yabumoto, Y. (1994). Early Cretaceous freshwater fish fauna in Kyushu, Japan. *Bulletin of the Kitakyushu Museum of Natural History* **13**, 107–254.

*Yabumoto, Y. (2008). A new Early Cretaceous osteoglossomorph fish from Japan, with comments on the origin of the Osteoglossiformes. In: Arratia, G., Schultze, H.-P., Wilson, M. V. H. (eds.) *Mesozoic Fishes 4 - Homology and Phylogeny*, 217–228. Verlag Dr. F. Pfeil, München.

*Yabumoto, Y. (2013). *Kokuraichthys tokuriki* n. gen. & sp., Early Cretaceous osteoglossomorph fish in Kyushu, Japan. *Bulletin of the Kitakyushu Museum of Natural History and Human History, Series A* **11**, 67–72.

*Yabumoto, Y. & Yang, S. Y. (2000). The first record of the Early Cretaceous freshwater fish, *Wakinoichthys aokii*, from Korea. *Bulletin of the Kitakyushu Museum of Natural History* **19**, 105–110.

*Zanata, A. M. & Vari, R. P. (2005). The family Alestidae (Ostariophysi, Characiformes): a phylogenetic analysis of a trans-Atlantic clade. *Zoological Journal of the Linnean Society* **145**, 1–144.

*Zhang, J.-Y. (1998). Morphology and phylogenetic relationships of †*Kuntulunia* (Teleostei: Osteoglossomorpha). *Journal of Vertebrate Paleontology* **18**, 280–300.

*Zhang, J.-Y. (2002). A new species of *Lycoptera* from Liaoning, China. *Vertebrata PalAsiatica* **40**, 257–266.

*Zhang, J.-Y. (2003). First *Phareodus* (Osteoglossomorpha: Osteoglossidae) from China. *Vertebrata PalAsiatica* **41**, 327–334.

*Zhang, J.-Y. (2004). New fossil osteoglossomorph from Ningxia, China. *Journal of Vertebrate Paleontology* **24**, 515–524.

Zhang, J.-Y. (2006). Phylogeny of Osteoglossomorpha. *Vertebrata PalAsiatica* **44**, 43–59.

*Zhang, J.-Y. (2010). Validity of the osteoglossomorph genus †*Asiatolepis* and a revision of †*Asiatolepis muroii* (†*Lycoptera muroii*). In: Nelson, J. S., Schultze, H.-P. & Wilson, M. V. H. (eds.) *Origin and Phylogenetic Interrelationships of Teleosts*, 239–249. Verlag Dr. F. Pfeil, München.

*Zhang, J.-Y. & Jin, F. (1999). A revision of †*Tongxinichthys* Ma 1980 (Teleostei: Osteoglossomorpha) from the Lower Cretaceous of northern China. In: Arratia, G. & Schultze, H.-P. (eds.) *Mesozoic Fishes 2 – Systematics and Fossil Record*, 385–396. Verlag Dr. F. Pfeil, München.

*Zhang, J.-Y., Jin, F. & Zhou, Z. (1994). A review of Mesozoic osteoglossomorph fish *Lycoptera longicephalus*. *Vertebrata Pal Asiatica* **32**, 41–59.

Zhang, J.-Y. & Wilson, M. V. H. (2017). First complete fossil *Scleropages* (Osteoglossomorpha). *Vertebrata PalAsiatica* **55**, 1–23.

CHAPTER 3

A Paleocene (Danian) Marine Osteoglossid (Teleostei: Osteoglossomorpha) from the Nuussuaq Basin of Greenland, with a Brief Review of Palaeogene Marine Bonytongue Fishes

Note: The contents of this chapter have been published[2].

Abstract: The early Palaeogene represents a key interval in the evolution of modern marine fish faunas. Together with the first appearances of many familiar fish lineages characteristic of contemporary marine environments, early Palaeogene marine deposits worldwide feature the occurrence of osteoglossid bonytongues. Their presence in marine rocks is surprising, as these fishes are strictly associated with freshwater environments in modern settings and other parts of the fossil record. Despite its possible relevance to faunal recovery after the K–Pg extinction, this marine osteoglossid radiation is relatively understudied. Here we describe an osteoglossid specimen from marine Danian deposits of western Greenland (Eqalulik Formation, northern Nuussuaq Peninsula). It consists of disarticulated cranial, pectoral and vertebral material belonging to a relatively large-bodied predator, similar to the widespread †*Brychaetus* but with some distinctive features. This specimen expands the geographic range of extinct osteoglossids to the Arctic and represents one of the earliest records of this group in marine deposits. We review other fossil occurrences of marine osteoglossids, highlighting temporal and biogeographic patterns that characterize their rise, diversification and sudden disappearance in the middle Eocene. It is likely that the transition from freshwater to marine environments occurred around the K–Pg boundary, possibly related to ecological replacement of predatory fish lineages that went extinct at the end of the Cretaceous. Further study of the Eqalulik Formation fauna could yield additional insight into the consequences of the end-Cretaceous extinction on marine fish evolution and on the assembly of modern marine faunas.

Key words: Osteoglossidae, Danian, Greenland, K–Pg extinction, early Palaeogene, marine fauna.

[2] **Capobianco, A., Foreman, E. & Friedman, M., 2021.** A Paleocene (Danian) marine osteoglossid (Teleostei: Osteoglossomorpha) from the Nuussuaq Basin of Greenland, with a brief review of Palaeogene marine bonytongue fishes. *Papers in Palaeontology*, 7: 625–640.

Fossil and molecular data point to substantial diversification in multiple groups of marine fishes in the wake of the Cretaceous-Palaeogene (K–Pg) mass extinction (Friedman 2010; Miya *et al*. 2013; Guinot & Cavin, 2016; Alfaro *et al*. 2018; Ribeiro *et al*. 2018; Sibert *et al*. 2018). Early Eocene (Ypresian) marine deposits demonstrate that marine fish faunas had a relatively modern phylogenetic composition by 50 Ma, with a dominance of acanthomorphs (Patterson 1993; Friedman *et al*. 2016; Friedman & Carnevale 2018). Among the new groups to appear in the early Palaeogene are multiple lineages of large, predatory acanthomorphs (and more specifically, percomorphs) that persist to the modern day: billfishes, scombroids (tunas, mackerels and relatives), carangoids and barracudas (Friedman 2009; Monsch & Bannikov 2011; Miya *et al.* 2013; Alfaro *et al*. 2018; Ribeiro *et al*. 2018). This has been interpreted through the lens of an ecological release model, with extant groups filling ecological roles vacated by the extinction of the dominant clades of large-bodied Late Cretaceous marine predators, like †ichthyodectiforms, †pachyrhizodontids, †pachycormids, and †enchodontids (Cavin & Martin 1995; Cavin 2002; Friedman 2009). The familiar modern predators that first emerge in the early Palaeogene are joined by a parallel, but short-lived, marine radiation of an unlikely group of non-acanthomorph teleosts: the osteoglossomorphs, commonly known as bonytongues. This is particularly striking because all extant osteoglossomorphs are restricted to freshwater environments, with few species occasionally found in brackish waters (Berra 2007). Similar environmental associations characterize nearly all of the osteoglossomorph fossil record, which extends to the Middle–Late Jurassic (Capobianco & Friedman 2018; Hilton & Lavoué 2018). In contrast to the strong freshwater association that characterizes roughly 130 million years of the osteoglossomorph record, the Paleocene–Eocene interval yields several species of marine osteoglossomorphs from sites across the world (Casier 1966; Danilchenko 1968; Bonde 2008; Wilson & Murray 2008; Forey & Hilton 2010; Alvarado-Ortega *et al.* 2015). Phylogenetic relationships of these marine

forms remain unresolved (Wilson & Murray 2008) and it is unclear whether they represent a polyphyletic assemblage arising from multiple freshwater-to-marine environmental transitions (as proposed by Bonde 2008) or a monophyletic radiation stemming from a single marine invasion. Nonetheless, several marine taxa can be confidently placed within the osteoglossomorph sub-clade Osteoglossidae (Forey & Hilton 2010). Extant osteoglossids are freshwater but are distributed between Africa, South America, Southeast Asia and northern Oceania. The geographically widespread occurrences of marine osteoglossids in early Palaeogene deposits hint at an intriguing scenario where marine dispersal followed by freshwater invasions played a role in shaping the modern disjunct distribution of this group (Bonde 2008; Forey & Hilton 2010; Capobianco & Friedman 2018; Hilton & Lavoué 2018), which has otherwise been interpreted in a vicariance biogeographic framework.

Here we present new material of a marine osteoglossid from early Paleocene (Danian) deposits of Greenland (Fig. 3.1). Although not diagnostic at the specific or generic level, this fossil is nevertheless significant on both geographic and stratigraphic grounds. It expands the range of marine bonytongues to high latitudes, and represents a rare example from the early Paleocene, an interval important for constraining patterns of turnover associated with the K–Pg but for which relatively little fossil fish material is known when compared to the Late Cretaceous and Eocene (Patterson 1993). We place this specimen in the broader context of the marine osteoglossomorph radiation(s) by giving an overview of known fossil occurrences. Finally, we outline outstanding questions surrounding this distinctive and unusual feature of early Palaeogene marine ichthyofaunas and compare it to other examples of short-lived lineages that are prominent after the K–Pg mass extinction and which might be interpreted as components of a recovery fauna.

MATERIAL AND METHODS

Specimens examined

Skeletonized specimens of extant taxa and fossil specimens of extinct taxa belonging to Osteoglossomorpha were examined as comparative material. Names of extinct taxa are preceded by a dagger symbol (†).

Hiodontidae. *Hiodon tergisus* UMMZ 180315.

Mormyridae. *Marcusenius macrolepidotus* UMMZ 200066; *Mormyrus lacerda* UMMZ 200084.

Osteoglossidae. *Arapaima gigas* UMMZ 177540, UMMZ 203831; †*Brychaetoides greenwoodi* MGUH 28906; †*Brychaetus muelleri* NHMUK PV 39699, NHMUK PV P641, NHMUK PV P3893, NHMUK PV P66889; cf. †*Brychaetus* sp. MGUH 28907; †*Brychaetus*? sp. NHMUK PV P73088; *Heterotis niloticus* UMMZ 213845; †*Heterosteoglossum foreyi* MGUH 28904; †*Magnigena arabica* NHMUK PV OE PAL 2007-1; Osteoglossidae indet. †NHMUK PV P66354; *Osteoglossum bicirrhosum* UF 189007, UMMZ 203832; †*Phareodus testis* NHMUK PV P61230; †*Ridewoodichthys caheni* MRAC RG 9169–70; *Scleropages formosus* UMMZ 203833, UMMZ 213853; †*Xosteoglossid rebeccae* MGUH 28905.

Osteoglossomorpha incertae sedis. †*Foreyichthys bolcensis* NHMUK PV P16821; †*Monopteros gigas* MNHN F BOL 285, MNHN F BOL 288; †*Thrissopterus catullii* IGUP 8839–8840.

In addition to the material listed here, further observations of osteoglossomorph osteology were made based on Kershaw (1970, 1976), Taverne (1977, 1978) and Hilton (2003).

Micro-computed tomography scanning

The specimen described here (NHMD 72014 A+B) and some of the comparative specimens were studied using micro-computed tomography (μCT) datasets produced using Nixon XT H 225ST industrial μCT scanners at the University of Michigan and the Natural History Museum, London. Individual scanning parameters are given below:

NHMD 72014 A+B (two halves of the specimen scanned independently). Voltage: 210 kV; current: 220 μA; filter: 2.5 mm copper; reflection target: tungsten; effective pixel size: 92 μm; scanning facility: University of Michigan CTEES. Following best practices in the accessibility of tomographic data (Davies *et al*. 2017), we have made tomograms, .mcs files, and .plys of segmented structures available on Dryad Digital Repository (Capobianco *et al*. 2019).

†*Brychaetus muelleri* NHMUK PV P641. Voltage: 190 kV; current: 305 μA; filter: 2.7 mm copper; reflection target: tungsten; effective pixel size: 62.9 μm; scanning facility: University of Michigan CTEES.

†*Brychaetus muelleri* NHMUK PV 39699. Voltage: 210 kV; current: 200 μA; filter: 1 mm tin; reflection target: tungsten; effective pixel size: 50.6 μm; scanning facility: NHM, Imaging and Analysis Centre.

Scans were acquired using Inspect-X and reconstructed using CT Pro 3D (Nikon Metrology, USA). Additionally, reconstructed tomograms for *Osteoglossum bicirrhosum* UF 189007 were downloaded from Morphosource (media M26520).

Reconstructed datasets were visualized and segmented using Mimics v. 19.0 (Materialise, Belgium). 3D models of segmented skeletal elements were exported from Mimics as surface files (.ply). Surface files of elements belonging to the two halves of NHMD 72014 A+B were reconstructed together in Blender v. 2.79 (blender.org). Two-dimensional high-resolution renderings of surface files were also acquired in Blender.

Institutional abbreviations. FUM, Fur Museum, Fur, Denmark; IGUP, Istituto Geologico dell'Università di Padova, Padova, Italy; MCSNV, Museo Civico di Storia Naturale, Verona, Italy; MGUH, Geology Museum, University of Copenhagen (stored at Fur Museum), Denmark; MNHN, Muséum National d'Histoire Naturelle, Paris, France; MRAC, Musée Royal de l'Afrique Centrale, Tervuren, Belgium; NHMD, Natural History Museum of Denmark, Copenhagen, Denmark; NHMUK, Earth Sciences collections, Natural History Museum, London, UK; UF, University of Florida, Gainsville, USA; UMMP, University of Michigan Museum of Paleontology, Ann Arbor, USA; UMMZ, University of Michigan Museum of Zoology, Ann Arbor, USA.

SYSTEMATIC PALAEONTOLOGY

TELEOSTEI Müller, 1845

OSTEOGLOSSOMORPHA Greenwood, Rosen, Weitzman & Myers, 1966

OSTEOGLOSSIFORMES Berg, 1940

OSTEOGLOSSIDAE Berg, 1940

Gen. et sp. indet.

Figures 3.2–3.5

Material. NHMD 72014 A+B, two halves of an ellipsoidal concretion with disarticulated bones, including maxilla, urohyal, scapula, several vertebral centra and fragments of lepidotrichia and radials (Fig. 3.2).

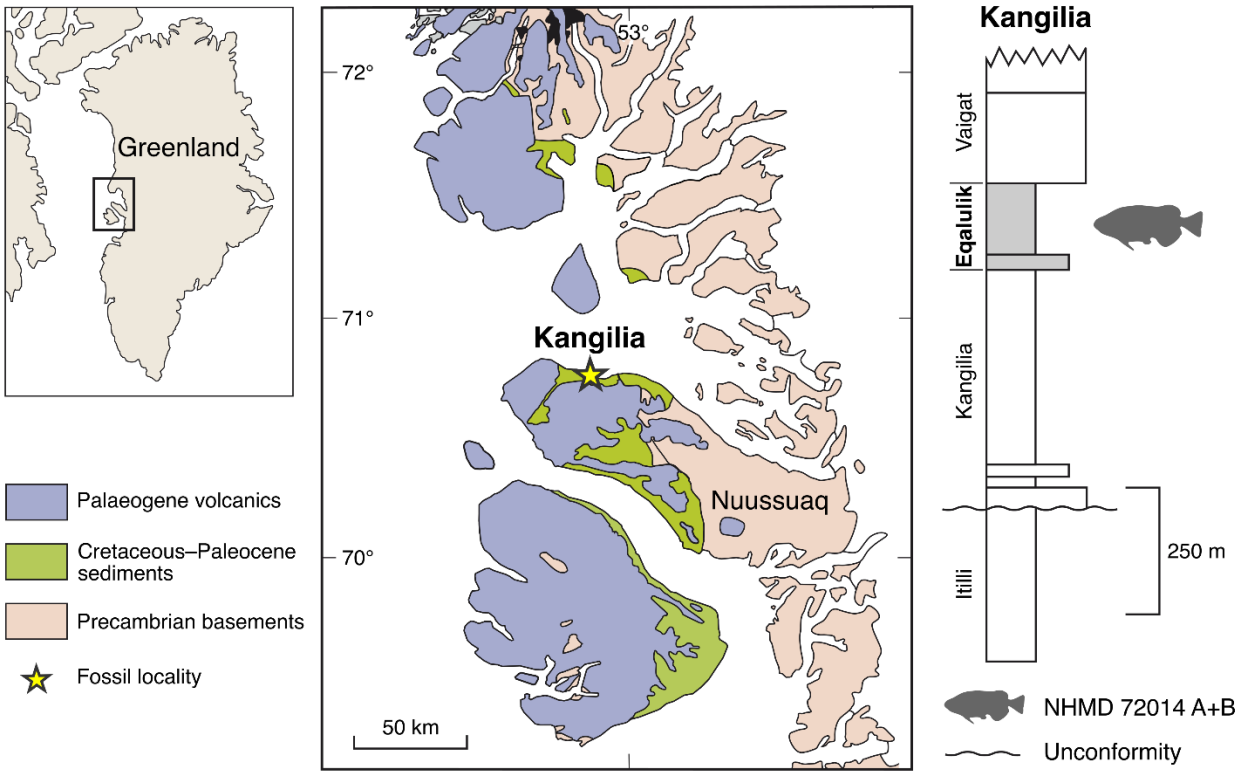

Fig. 3.1. Map of the Nuussuaq Peninsula, western Greenland, with simplified stratigraphic log of formations of the Late Cretaceous–Paleocene Nuussuaq Group outcropping in Kangilia. Modified from Dam *et al*. (2009).

Occurrence. The specimen was collected by Abraham Løvstrøm of the Geological Survey of Denmark and Greenland (GEUS) in 1957. It comes from the Kangilia Ridge in the north coast of the Nuussuaq Peninsula (western Greenland; Fig. 3.1). It was briefly mentioned – among other specimens – as indeterminate actinopterygian material from the Danian Kangilia Formation by

Bendix-Almgreen (1969) and Rosenkrantz (1970), who further refined its stratigraphic position as belonging to the '*Thyasira* Member'. The same deposits yielded the oldest known gadoid fish, still undescribed and informally known as †'*Protocodus*' (Rosen & Patterson 1969; Cohen 1984). A recent revision of the lithostratigraphy of the Nuusuuaq Basin (Dam *et al*. 2009) assigned some deposits of the Kangilia and Agatdal formations (including the '*Thyasira* Member') to the newly named Eqalulik Formation. The Eqalulik Formation was deposited in a relatively deep marine environment, with maximum water depth around 700 m (Dam *et al*. 2009). The age of the Eqalulik Formation is not well determined, partly because of its diachrony throughout the Nuussuaq Basin (Dam *et al*. 2009). While the macrofauna of the Eqalulik Formation has early Danian affinities (Rosenkrantz, 1970), palyno- and nannostratigraphy suggest a late Danian age (upper NP3 – lower NP4 nannoplankton zone, 63–62 Ma; Nøhr-Hansen & Sheldon 2000; Anthonissen & Ogg 2012).

Description. An almost complete right maxilla (97 mm in length) is preserved in the concretion (mx, Figs. 3.2–3.3). It is relatively robust and slightly dorso-ventrally bowed. It becomes more laterally compressed posteriorly, although the posterior region of the bone is not complete. Anteriorly, the maxilla extends into a short and deep antero-medial process that would have articulated with the posterior portion of the premaxilla. Posterior to this process, it presents a distinct dorsal thickening, while the medial surface of the bone bears a relatively deep elliptical pit. 20 teeth and 6 empty tooth sockets are present, arranged in a single row. Teeth increase in size from the posterior to the anterior of the jaw. The teeth are straight, antero-posteriorly compressed at the base and conical at the tip. Each tooth is hollow inside throughout its length (Fig. 3.3C–D). The tooth base appears to be sheathed in bone for a very short length (less than a third of the length of the tooth).

The urohyal (uh, Fig. 3.2) is large (around 80 mm in length) and laterally compressed with a rounded anterior head. It is much deeper posteriorly than anteriorly. Posterior to its rounded head, it bears a thickened dorsal ridge apparent in transverse cross section (Fig. 3.4A).

Part of the left pectoral girdle is preserved, including the scapula (sc, Fig. 3.2) and possible fragments of the cleithrum. The scapula is irregularly shaped with a distinct sub-circular scapular foramen (scf, Fig. 3.5A–B). Details of scapular morphology are difficult to interpret due to its

poor preservation and low contrast in tomograms. A slightly concave facet on the posterior edge of the scapula is interpreted as the articular surface for the first pectoral fin-ray (afp, Fig. 3.5A–B). Marked thickening on the meso-ventral edge is probably indicative of articulation with other radials of the pectoral fin and with the coracoid. Fragments of what are likely two lepidotrichia (fin rays) and one radial are present. The lepidotrichia (lp, Fig. 3.2) appear to be segmented and dorso-ventrally flattened. The putative radial (ra, Fig. 3.2) is sub-cylindrical in shape.

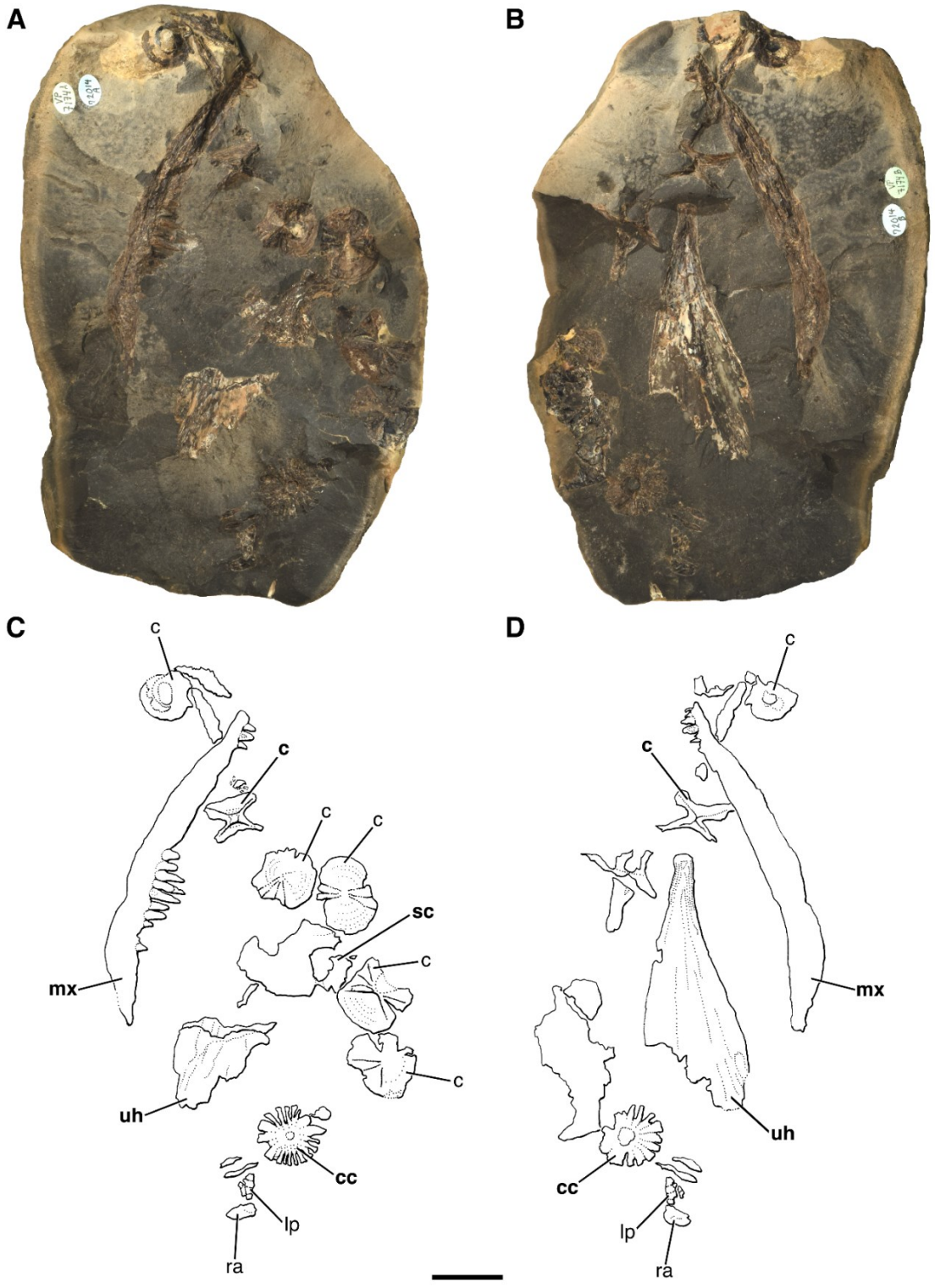

Fig. 3.2. Photographs of NHMD 72014 A+B (A–B), with interpretative line drawings (C–D). Labels in bold indicate elements rendered in Figs. 3.3–3.4. Unlabeled elements have uncertain identity. Scale bar represents 20 mm. *Abbreviations*: c, centrum; cc, caudal centrum; lp, lepidotrichia; mx, maxilla; ra, radial; sc, scapula; uh, urohyal.

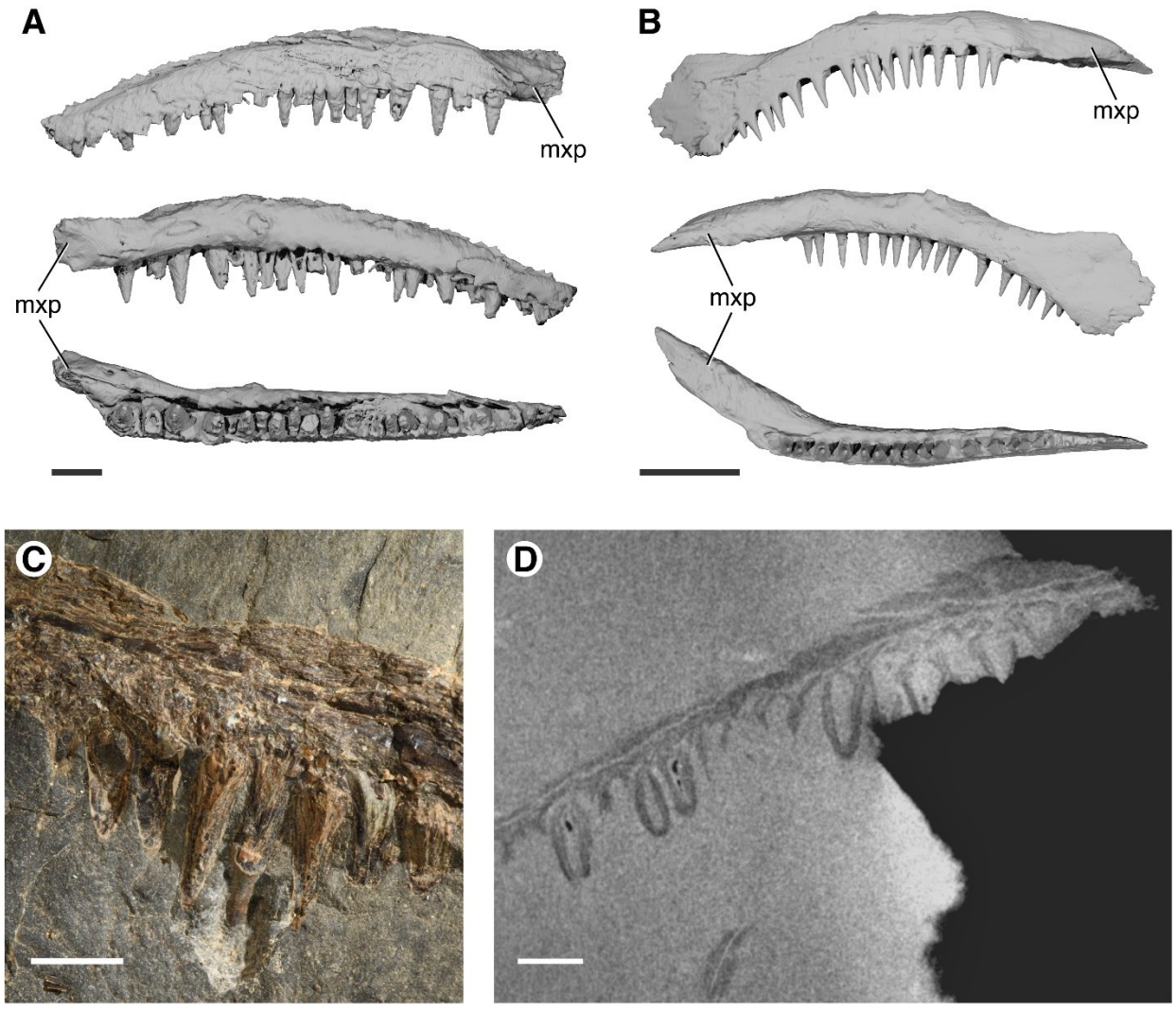

Fig. 3.3. Maxillary and dental morphology of NHMD 72014 A+B. A–B, renderings comparing the maxilla of NHMD 72014 A+B (A) with the maxilla of †*Brychaetus muelleri* NHMUK PV 39699 (B) in lateral, mesial and ventral views (from top to bottom); C, close-up photograph of maxillary teeth; D, tomogram showing maxillary teeth in sagittal section. Scale bars represent: 10 mm (A–B), 5 mm (C–D). *Abbreviations*: mxp, anterior articular process of maxilla.

Seven vertebral centra are preserved in the concretion, ranging from very fragmentary to almost complete (c and cc, Fig. 3.2). All centra are amphicoelous, sub-circular in transverse section and their width and depth is larger than their length. There is substantial morphological variability among the preserved centra. An almost complete abdominal centrum (Fig. 3.4C) displays oval, deep neural arch pits, probably separated by a shallow and narrow mid-dorsal pit. The neural arch is absent, suggesting that it was an autogenous element in life. A pair of deep ventral pits is

also present. The lateral surface of the centrum is pitted by small circular pores. No parapophysis is evident on this centrum. Due to the combination of these features, this centrum could represent one of the anteriormost centra of the vertebral column. An almost complete caudal centrum is preserved in the concretion, together with fragments of its neural and haemal arches (Fig. 3.4D). Each lateral surface of the centrum presents six or seven longitudinal sulci. A deeper ventral pit is bordered by the haemal arch.

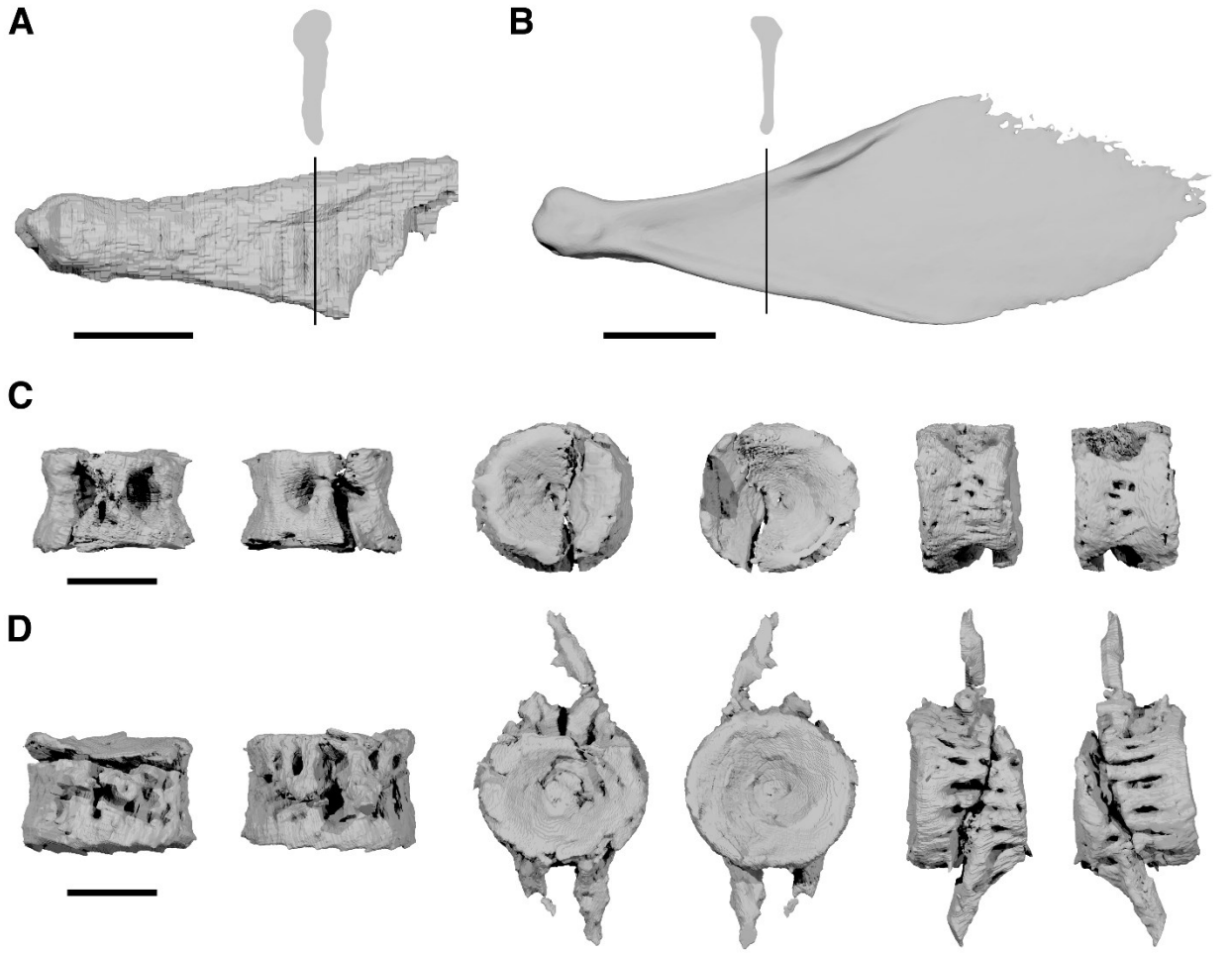

Fig. 3.4. Isolated skeletal elements of NHMD 72014 A+B. A, rendering of the anterior portion of the urohyal in left lateral view, with silhouette of its cross section. B, rendering of the urohyal of *Osteoglossum bicirrhosum* (UF 189007) in left lateral view, with silhouette of its cross section; C–D, renderings of an abdominal vertebral centrum (C) and a caudal vertebra (D) in dorsal, ventral, cranial, caudal, right lateral and left lateral views (from left to right). Scale bars represent: 10 mm (A, C–D), 2 mm (B).

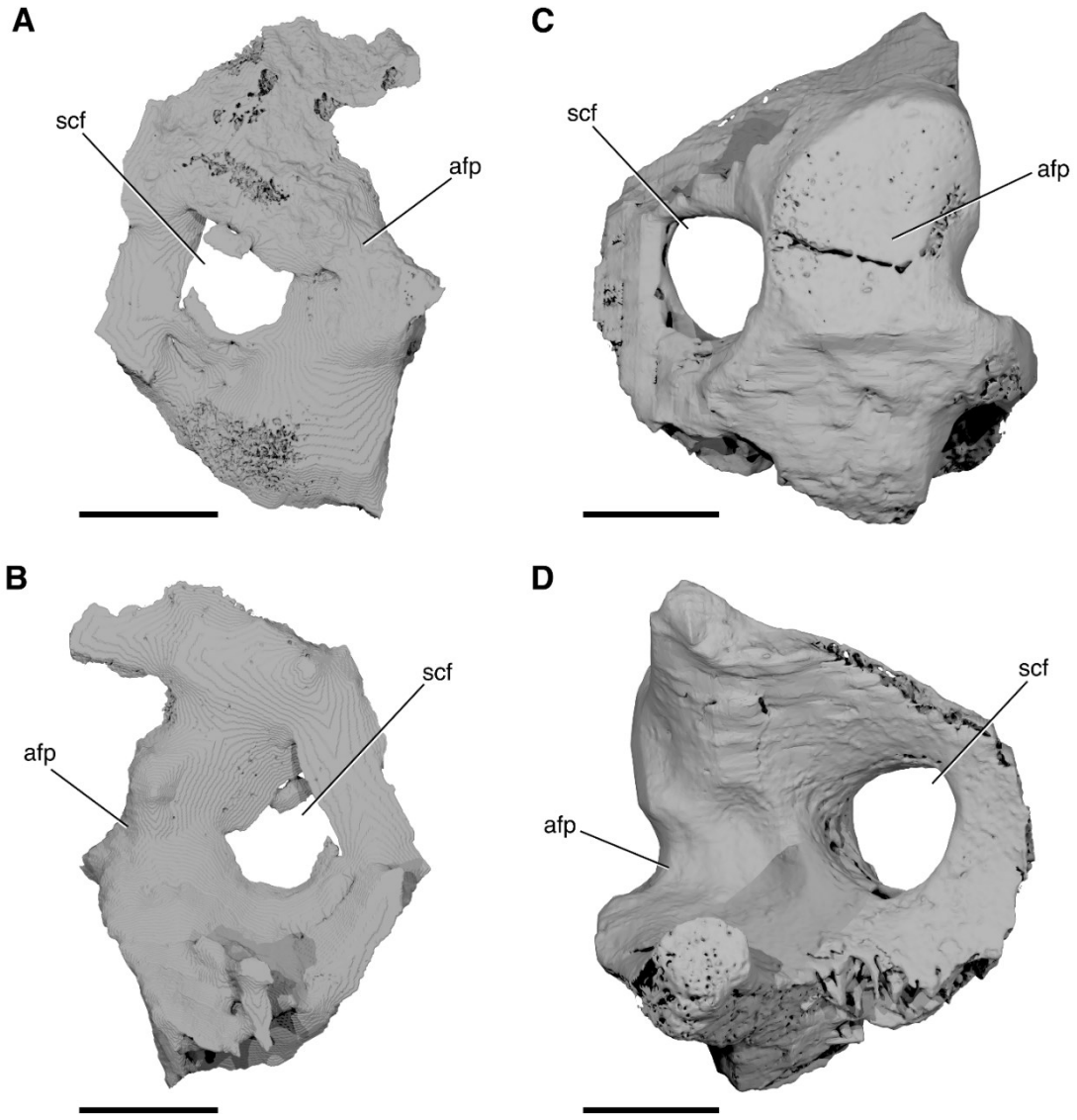

Fig. 3.5. Left scapula of NHMD 72014 A+B. A–B, renderings of the scapula in lateral (A) and mesial (B) views; C–D, renderings of the left scapula of †*Brychaetus muelleri* (paratype, NHMUK PV P641) in lateral (C) and mesial (D) views. Scale bars represent: 10 mm (A–B), 5 mm (C–D). *Abbreviations*: afp, articular facet for the first pectoral fin-ray; scf, scapular foramen.

DISCUSSION

Osteoglossid identity of the Greenland specimen and comparison with other osteoglossids

Antero-posteriorly compressed large teeth with a conical tip and a hollow cavity throughout their length are characteristic of some members of Osteoglossidae, especially of the marine fossil taxon †*Brychaetus* (Casier 1966; Forey & Hilton 2010). Comparable teeth have been found

worldwide in several early Palaeogene marine deposits (see section below). The extant osteoglossids *Osteoglossum* and *Scleropages* share a similar tooth morphology, albeit with smaller size relative to the maxilla and with less pronounced antero-posterior compression. The presence of a bony tooth base is a condition found in several fossil osteoglossids, including †*Brychaetus* and †*Phareodus*. The bony base in †*Brychaetus* and †*Phareodus* teeth is usually much deeper than in the Greenland specimen (between one third and half of the total tooth length for maxillary teeth). A shallow bony tooth base has been described in the marine osteoglossid †*Ridewoodichthys* (Taverne 2009*a*). The size of the Greenland maxilla suggests a relatively large fish, perhaps around 1 m in length assuming body proportions comparable to †*Phareodus*. It apparently differs from the maxillae of †*Brychaetus* and †*Phareodus* in having an anterior articular process that is short and stout rather than elongated and tapering (Fig. 3.3A–B). Although this feature does not seem to be the result of incomplete preservation of the articular process, the resolution of the scan does not allow us to completely exclude that a longer process was indeed present but taphonomically damaged.

Compared to extant osteoglossomorphs, the urohyal closely resembles those of the osteoglossids *Osteoglossum* and *Scleropages* in its marked posterior deepening and laterally compressed cross-section with slight dorsal thickening (Fig. 3.4A–B). While Taverne (1977) described the urohyal of these two extant genera as having a Y-shaped cross section, we were not able to observe such a feature. The relatively large size of the urohyal is consistent with an osteoglossid attribution, whereas other osteoglossomorphs – such as *Hiodon* and *Pantodon* – present a much smaller urohyal (besides differing substantially in morphology).

The overall shape of the scapula and the relative size of the sub-circular foramen are similar to those of extant osteoglossids with the exception of *Heterotis*, whose scapular foramen is not completely enclosed by bone. While a detailed comparative assessment is prohibitive due to its poor preservation, comparison with the scapula of †*Brychaetus* (Fig. 3.5C–D) suggests significant differences in shape and proportions. In particular, what we interpret as the articular surface for the first pectoral fin-ray appears to be much smaller than in †*Brychaetus*. This feature might relate to the relative size of the first pectoral fin-ray, as observed in extant osteoglossid genera. *Osteoglossum* and *Scleropages* (as well as the extinct †*Phareodus*) are characterized by an enlarged and extremely thickened first pectoral fin-ray, and in turn present a large articular

surface on the scapula. In contrast, the first pectoral fin-ray in *Arapaima* and *Heterotis* is only slightly larger than the successive rays and is not particularly thickened, and the relative size of the corresponding articular surface on the scapula of these two genera is substantially smaller than in *Osteoglossum*, *Scleropages* and †*Brychaetus*. Thus, it is possible that the Greenland osteoglossid did not have a greatly enlarged first pectoral fin-ray.

An autogenous neural arch in precaudal vertebrae is a plesiomorphic teleost feature and is consistent with an osteoglossid identification (Brinkman & Neuman 2002). The presence of several longitudinal sulci in the lateral surface of the caudal vertebra recalls the vertebral morphology of *Arapaima*. In sum, the Greenland fossil shows clear resemblance to osteoglossids, including specific correspondences to several extinct and modern genera. However, the remains are too incomplete to propose an assignment beyond the family level.

Geographic and stratigraphic distribution of marine osteoglossids

Otoliths referred to osteoglossomorphs have been found in few Mesozoic marine deposits. The otolith genus †*Archaeglossus* from the Middle–Late Jurassic of England could represent the oldest record for this group (Schwarzhans 2018). The presence of early stem osteoglossomorphs in marine deposits would not be completely unexpected, as the osteoglossomorph lineage likely derives from marine ancestors (Betancur-R *et al*. 2015). Three otolith species originally identified as albuliforms have been recently grouped in the genus †*Kokenichthys* (Schwarzhans 2010) and interpreted as possible osteoglossomorphs (Schwarzhans 2018). These are Aptian to Maastrichtian in age and have been found in lagoonal or marine deposits of Germany, Spain and USA. In addition to †*Kokenichthys ripleyensis*, the early Maastrichtian Ripley Formation in Mississippi (USA) yielded otoliths with osteoglossid-like morphology (†"Arapaimina" *tavernei*; Nolf & Stringer 1996; Nolf 2013). These otolith occurrences should be interpreted with caution, as extant osteoglossomorph otoliths show a remarkable morphological variety (Nolf 2013), unique otolith synapomorphies have not been defined for the group, and alternative systematic placements (such as within Albuliformes) are possible. A putative Mesozoic marine osteoglossomorph known from a single articulated specimen is the Cenomanian †*Prognathoglossum kalassyi* from Lebanon, assigned by Taverne & Capasso (2012) to Pantodontidae, an osteoglossomorph lineage with a single extant species: the freshwater African

butterflyfish *Pantodon buccholzi*. This bizarre taxon presents a series of peculiar features, including shortened and bulbous braincase, extremely long lower jaw, mouth oriented nearly vertically, reduced pectoral fins and long dorsal fin extending throughout the length of the body, that challenge its identification as a pantodontid osteoglossomorph. Nonetheless, despite these putative osteoglossomorph occurrences in marine settings before the Palaeogene, definitive marine osteoglossids are found only in sediments ranging from the earliest Paleocene to the middle Eocene (Fig. 3.6). A possible exception is represented by a single incomplete jaw fragment from the Maastrichtian Ménaka Formation of Mali referred to the genus †*Brychaetus* (O'Leary *et al*. 2019), which is known from early Palaeogene marine deposits worldwide (see following subsections).

Danian. The oldest unambiguous marine osteoglossid fossil is an isolated jaw fragment (probably a premaxilla) from the Fiskeler Member of the Stevns Klint in Denmark (Schwarzhans & Milàn 2017). While this single specimen was described as unidentified osteoglossomorph, the morphology of its teeth (large and columnar in lateral view) is consistent with an osteoglossid attribution. The age of this specimen is remarkable, as the Fiskeler Member immediately overlies the Cretaceous-Palaeogene (K–Pg) boundary in a continuous succession. This demonstrates the presence of osteoglossids in the marine realm in the earliest Danian, shortly after the K–Pg extinction. An almost complete specimen of a marine osteoglossid comes from the Danian Tenejapa Formation of Palenque, Mexico (Alvarado-Ortega *et al*. 2015). It is assigned to the genus †*Phareodus*, which is otherwise known from early-middle Eocene freshwater deposits of USA, China and Australia (Li 1994; Li *et al*. 1997; Zhang 2003). Thus, this specimen would represent the oldest known †*Phareodus* by at least 5 million years (see following subsections) and the only one from marine deposits. However, the attribution of this specimen to †*Phareodus* has been based on characters (such as opercle shape and absence of teeth on the parasphenoid) that have broader distribution within osteoglossids (including the marine genus †*Brychaetus*; Forey & Hilton 2010). Further study is needed to determine the generic status of this Danian osteoglossid.

Selandian. Fragmentary osteoglossid material is known from marine deposits in the Landana section of the Cabinda enclave of Angola (Taverne 2009*a*, 2016). While these deposits are classically regarded as Danian in age ('Montian' in older literature), a recent reassessment of the

regional stratigraphy indicates that the osteoglossid material is most likely Selandian in age (Solé *et al*. 2018). These remains comprise a caudal skeleton and jaw fragments assigned to †*Ridewoodichthys caheni*, which is similar to †*Brychaetus*, and a single caudal skeleton of a slightly younger, unnamed taxon (Taverne 2009*a*, 2016).

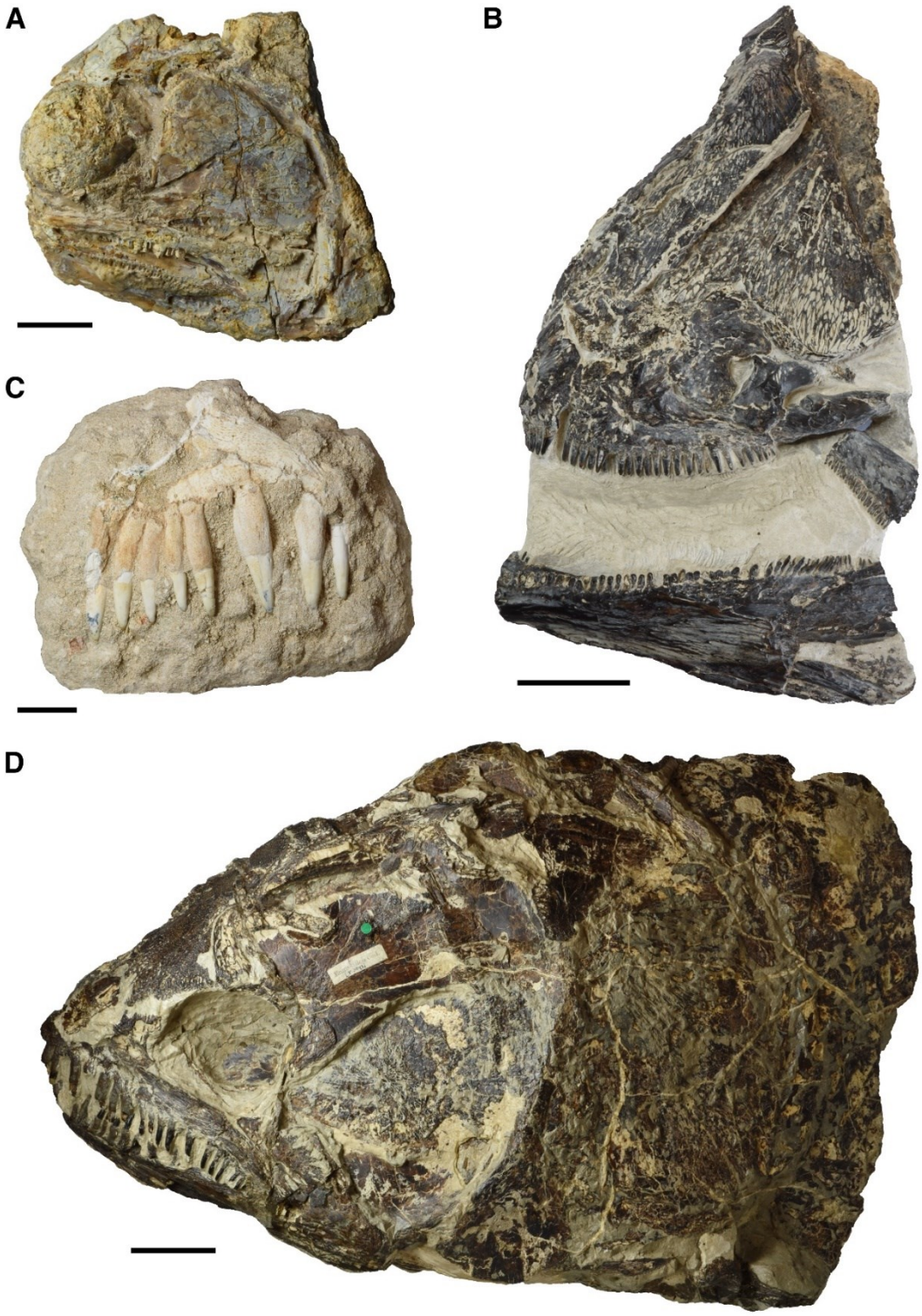

Fig. 3.6 *(previous page)*. Photographs of representative fossil osteoglossids found in marine deposits. A, †*Magnigena arabica* (holotype, NHMUK PV OE PAL 2007-1) from the Thanetian Umm Himar Formation of Saudi Arabia. B, †*Brychaetoides greenwoodi* (holotype, MGUH 28906) from the Ypresian Fur Formation of Denmark. C, †*Brychaetus*? sp. (NHMUK PV P73088) from Ypresian phosphates of

Morocco. D, †*Brychaetus muelleri* (holotype, NHMUK PV P3893) from the Ypresian London Clay Formation of England. Scale bars represent: 10 mm (A), 20 mm (B–C), 40 mm (D).

Thanetian. The marine osteoglossid †*Magnigena arabica* is known from a partial articulated skull (Fig. 3.6A) found in the late Paleocene Umm Himar Formation of Saudi Arabia (Forey & Hilton 2010). An isolated premaxilla has been found in marine deposits of similar age from Sessao, the Iullemeden Basin of Niger (Cappetta 1972). While this specimen was originally described as †*Brychaetus*, Taverne (2009*b*) assigned it to a new extinct species in the modern genus *Scleropages*, which currently occurs in fresh waters of Southeast Asia and northern Oceania. Given the scarcity of the material, we adopt a more cautious approach and consider this premaxilla to belong to an indeterminate osteoglossid. Osteoglossid teeth referred to †*Brychaetus muelleri* have been found in the Thanetian Tuscahoma Formation of Mississippi (USA), which represents either an estuarine or a marine environment (Case 1994).

Ypresian. Early Eocene deposits yield the highest diversity of marine fossil osteoglossids, which are found around the world in the Ypresian. The marine Danata Formation of Turkmenistan yielded three articulated specimens of the relatively large osteoglossid †*Opsithrissops osseus* (Danilchenko 1968). The formation straddles the Paleocene–Eocene boundary, with the fish-bearing horizon lying at the border between Thanetian and Ypresian (Bannikov & Parin 1997). Three osteoglossomorph taxa are known from the coeval Stolle Klint Clay of Denmark, which was deposited in a shallow, landlocked marine basin (Bonde 2008). The Fur Formation directly overlays the Stolle Klint Clay and increases the diversity of early Ypresian marine osteoglossomorphs from Denmark up to six distinct taxa in total (Bonde 2008). Most of these are known from articulated remains. While their systematic affinities are uncertain and require further investigation, at least some of them are recognizable as definitive osteoglossids (including †*Heterosteoglossum*, †*Xosteoglossid* and a †*Brychaetus*-like form; Bonde 2008). Contrary to Bonde (2008), we also regard †*Brychaetoides* from the Fur Formation as an osteoglossid due to shared features with other osteoglossids like †*Brychaetus* (such as the lateral expansion of the anterior portion of the frontals; Fig. 3.6B). The best known marine osteoglossid is the large predator †*Brychaetus muelleri* from the London Clay Formation of England (Fig. 3.6D), represented by several articulated specimens (mostly skulls; Casier 1966; Roellig 1974; Taverne 1978). These deposits yield an additional indeterminate osteoglossid, known from a

partial neurocranium (Forey & Hilton 2010). Isolated osteoglossid scales have been also found in the London Clay and resemble those of the Danish †*Heterosteoglossum* (Bonde 2008). Osteoglossomorphs have been also reported from the species-rich reef assemblage of the Bolca Lagerstätten in Italy (Taverne 1998), although the interpretation of some of these is questionable. Among them, †*Monopteros* could be an osteoglossid or a more basal osteoglossiform with a peculiar durophagous dentition (Taverne 1998; Bonde 2008). †*Thrissopterus* could represent an osteoglossid with broad pectoral fins and very elongated body (Taverne 1998). †*Foreyichthys* has been alternatively interpreted as closely related to osteoglossids (Taverne 1998; Bonde 2008) or as an indeterminate osteoglossomorph (Forey & Hilton 2010). However, we advise caution in any interpretation of †*Foreyichthys* as an osteoglossomorph. The only known specimen has a very poorly preserved skull and its postcranial skeleton does not present unique osteoglossomorph synapomorphies (pers. obs. of NHMUK PV P.16821). Moreover, †*Brychaetus muelleri* (or a very similar species) is also known from Monte Bolca and is represented by an undescribed articulated specimen (part and counterpart IGUP 26282 and MCSNV IG 24548) that is currently under study by some of the authors of the present paper. Jaw fragments and isolated teeth referred to †*Brychaetus* – based on the antero-posteriorly compressed tooth morphology, with conical enamel cap and hollow bony base – have been found in several Ypresian marine deposits outside of Europe, including the Indian Laki Series (Forey & Hilton 2010), the 'Couches I-0' of Oulad Abdoun, Moroccan Phosphates (Fig. 3.6C; Arambourg, 1952; Bardet *et al*. 2017), the Tamaguélelt Formation of Mali (Patterson & Longbottom 1989; O'Leary *et al*. 2019) and the Nanjemoy Formation in Maryland and Virginia, USA (Weems & Horman 1983; Weems 1999). Articulated remains of non-Eurasian marine fossil osteoglossids are rare. A three-dimensionally preserved cranial skeleton of a new osteoglossid species from the Ypresian Moroccan phosphates (UMMP 118216) is currently under study by the authors.

Lutetian. The youngest marine osteoglossid remains come from a handful of Lutetian deposits. It is uncertain whether this reflects a decline in diversity or abundance, or instead is a consequence of a more restrictive set of fish-yielding deposits in comparison to the Ypresian. Undescribed cranial material belonging to a new osteoglossid taxon has been recognised in collections from the early Lutetian Habib Rahi Formation from Punjabi Pakistan, which represent a relatively deep marine environment. This specimen (UMMP GSP-UM field no. 1981292) is currently under study by the authors. The same formation yielded another undescribed specimen (UMMP

GSP-UM field no. 1981251), preserving the impressions of skull and pectoral fin, that resembles †*Thrissopterus* from Bolca. Putative osteoglossid otoliths have been reported from middle Eocene marine deposits in Europe (Nolf & Cappetta, 1976; Stinton, 1977).

Osteoglossids completely disappear from the marine fossil record after the middle Eocene.

Paleobiological and paleobiogeographical significance of marine osteoglossids

The presence of a large-bodied marine osteoglossid in the Danian of Greenland is relevant for a variety of reasons. First, it represents the northernmost known occurrence for this clade, further expanding their geographic distribution in the Palaeogene. While the early Palaeogene was characterized by much warmer temperatures than today (Zachos *et al*. 2001), Arctic climate at that time was likely temperate with episodic cooling events (Dawson *et al.* 1976; Spielhagen & Tripati 2009; Zhang *et al*. 2019). The presence of a gadoid fish (†'*Protocodus*') in the Eqalulik Formation and of a temperate/warm temperate otolith ichthyofauna known from roughly coeval deposits of central Nuussuaq (Schwarzhans 2004) are further evidence of the environmental conditions of West Greenland in the early Paleocene. This contrasts with modern osteoglossid distribution – limited to tropical environments (Berra 2007) – and hints at a broader temperature tolerance of some early Palaeogene members of the clade.

Second, the specimen described here is also one of the oldest osteoglossid fossils known from marine deposits. This highlights that, while marine forms are best known from Ypresian deposits, osteoglossids transitioned from freshwater to the sea well before then. Due to the scarcity of osteoglossid remains in Campanian–Paleocene deposits, it remains unclear when and where this environmental transition happened. Campanian–Maastrichtian osteoglossid fossils are known only from freshwater or brackish deposits. They include †*Cretophareodus alberticus* from Campanian Oldman Formation (Li, 1996) of Canada and a possible indeterminate osteoglossid from the Maastrichtian Tremp Formation in Catalonia, Spain (Blanco *et al*. 2017), in addition to several jaw fragments, teeth and scales from Bolivia and India (Sahni & Bajpal 1988; Prasad 1989; Gayet & Meunier 1998). More complete osteoglossid remains are known from freshwater early Palaeogene deposits, roughly contemporary with the marine occurrences described in the previous section. Among these, the fossil taxa †*Phareodus*—including †*P.*

encaustus and †*P. testis* from the early–middle Eocene Green River and Bridger formations of western USA, †*P. songziensis* from the Ypresian Yangxi Formation of China and †*P. queenslandicus* from the late Paleocene – early Eocene Redbank Plains Formation of Australia—and †*Taverneichthys* from the Paleocene Palana Formation of India (Kumar *et al*. 2005; Taverne *et al*. 2009) share osteological features (including a broad supraorbital shelf formed by the frontals and aspects of tooth morphology) with at least some marine osteoglossids (Forey & Hilton 2010). The age of these taxa, their wide geographical distribution and their putative affinities with widespread marine taxa such as †*Brychaetus* raise the possibility that marine dispersal followed by multiple freshwater invasions might have played a role in the early Palaeogene biogeographic history of Osteoglossidae (Capobianco & Friedman 2018).

It has been hypothesized that the invasion of marine environments by osteoglossomorphs (and specifically osteoglossids) in the early Palaeogene could result from the opportunistic replacement of large-bodied predatory taxa that went extinct at the end of the Cretaceous (Friedman 2009; Friedman & Sallan 2012). These include some large lamniform sharks, stem teleosts such as †aspidorhynchids, †pachycormiforms and †ichthyodectiforms, and crown teleosts like †enchodontids (Friedman & Sallan 2012; Guinot & Cavin 2016). It is now clear that by the end of the Danian (around 4 million years after the extinction event) osteoglossids evolved large predatory forms in the marine realm, following a transition from freshwater to marine environments that likely occurred across the K–Pg boundary. Danian marine deposits record the presence of giant megalopids as another group of macropredatory fishes (Khalloufi *et al*. 2018). Significantly, large megalopids are known only from freshwater deposits in the latest Cretaceous (Bardack 1970) and from marine deposits during the Palaeogene (Fur Formation, London Clay Formation, and Moroccan phosphates; Bonde 1997; Friedman *et al*. 2016; Khalloufi *et al*. 2018), hinting at a comparable environmental transition before the early Paleocene. Other predatory marine fish taxa that survive up to the present day (such as scombrids, xiphioids, sphyraenids and carangoids) are currently unknown from Danian deposits and appear in the fossil record by the late Paleocene – early Eocene (Fierstine 2006; Monsch & Bannikov 2011; Carnevale *et al*. 2014). However, this pattern may result from the rarity of Danian marine deposits that have yielded fish body fossils. It is unclear whether competition with this modern suite of marine predators or other unknown factors led to the disappearance of osteoglossids from marine environments during or after the middle Eocene.

We suggest that marine osteoglossids represent a striking example of a short-lived recovery fauna, diversifying and dispersing short after the end-Cretaceous mass extinction (presumably due to ecological opportunity) and going extinct after a relatively short amount of geological time. Among marine vertebrates, an analogous pattern could have characterized the evolutionary history of †dyrosaurid crocodylians (Hastings *et al*. 2011; Martin *et al*. 2019) and †palaeophiid snakes (Rage *et al*. 2003). Similar examples of early Palaeogene recovery faunas are better known in the terrestrial realm, particularly in the context of the North American mammal fossil record (Rose 1981; Longrich *et al*. 2016).

CONCLUSIONS

NHMD 72014 A+B represents a large-bodied marine osteoglossid from the early Paleocene (Danian) of Greenland. While the specimen is too fragmentary to suggest a specific or even generic attribution, it presents both similarities (tooth morphology and size) and contrasts (morphology of maxilla and scapula) with the widespread early Palaeogene genus †*Brychaetus*. It is one of the oldest known osteoglossid occurrences in marine deposits and the most northerly occurrence for the family in any environmental setting, adding relevant temporal and biogeographic information to reconstruct a poorly known marine radiation of an ancient teleost lineage that is currently restricted to freshwater. At present, the fossil record suggests that osteoglossids invaded marine environments in close temporal proximity to the K–Pg mass extinction. Marine osteoglossids are rare and show little morphological diversity in the Paleocene, although it is possible that this might stem in part from a relatively meagre fossil record combined with the fragmentary nature of known material of this age. What is clear, however, is that early Eocene deposits yield a range of marine osteoglossids that show some diversity in body form and dentition, and are widely distributed geographically. While several of these occurrences are based upon fragmentary isolated remains, representing an obstacle to the attempt of reconstructing the osteoglossid expansion in marine environments at fine temporal and geographic scale, most named species are known from (at least partially) articulated specimens. Few of these taxa have been described in detail, and none have been subjected to µCT investigation despite relatively three-dimensional preservation in some examples (e.g. †*Brychaetus muelleri*, †*Brychaetoides greenwoodi*, †*Magnigena arabica*). The marine Paleogene

radiation of osteoglossids represents an unusual episode in the history of this otherwise freshwater group, and demands more detailed investigation than has been undertaken previously. In particular, several outstanding questions about these taxa remain unanswered, most notably whether or not they form a clade to the exclusion of other osteoglossids. Different systematic hypotheses have been offered in the literature (e.g. Taverne 1998, Bonde 2008, Forey & Hilton 2010), but we are confident that available material is of sufficient quality and quantity to make significant inroads on this problem.

Beyond the specific issues surrounding marine osteoglossids, the Greenland specimen suggests an important new source of information bearing on fish evolution in the early Cenozoic. The Danian represents a critical interval for constraining the impacts of the end-Cretaceous extinction on marine fishes, but is poorly sampled. Only a handful of informative body-fossil assemblages are known from this interval, the most prominent of which are Limhamn Quarry, Sweden (Adolfssen *et al*. 2017) and Palenque, Mexico (Alvarado-Ortega *et al*. 2015). These sites yield relatively few taxa and have received less attention than younger Paleogene assemblages. In light of the limitations of the existing Danian body-fossil record, further examination of the Eqalulik Formation fauna could make important contributions as a unit that is both paleogeographically remote and paleoenvironmentally distinct from existing sites. In addition to the yet undescribed †'*Protocodus*' and the osteoglossid reported here, several concretions from the Eqalulik Formation in the collections of the Natural History Museum of Denmark contain teleost remains (e.g. NHMD VP-7131 A-C [possible cranial material], NHMD VP-7140 A-B [parasphenoid], NHMD VP-7147 A-D [skull and pectoral fin], NHMD VP-7164 A-C [articulated postcranium covered in scales]). This study shows that such concretions are amenable to study through μCT, and that the remains they preserve—although often disrupted—are largely uncrushed. Although it is probable that known specimens will not be identifiable at the finest taxonomic levels, they still may be sufficiently diagnostic to help better illuminate a poorly understood interval with a significant bearing on the evolution of modern fish diversity.

Acknowledgments

The authors thank Bent E. K. Lindow for access to the collections of the Natural History Museum of Denmark and for loaning the fossil specimen described here; and Emma Bernard for access to the collections of the Natural History Museum, London and for specimen loans. Allison Longbottom provided μCT scans of NHMUK PV 39699. We also thank for specimen access Douglas Nelson at the University of Michigan Museum of Zoology, Bo Schultz and René L. Sylvestersen at the Fur Museum, Mariagabriella Fornasiero at the Istituto Geologico dell'Università di Padova, Anna Vaccari and Roberta Salmaso at the Museo Civico di Storia Naturale, Verona, Gaël Clement at the Muséum National d'Histoire Naturelle, Paris and Florias Mees at the Musée Royal de l'Afrique Centrale. We would like to thank Zach Randall and the Florida Museum of Natural History, University of Florida for permission to use CT data stored on Morphosource. This study includes data produced in the CTEES facility at University of Michigan, supported by the Department of Earth & Environmental Sciences and College of Literature, Science, and the Arts. This work was supported by funding from the Department of Earth and Environmental Sciences of the University of Michigan (Scott Turner Student Research Grant Award 2017, to A.C.) and by the Society of Systematic Biologists (2017 SSB Graduate Student Research Award, to A.C.).

DATA ARCHIVING STATEMENT

Data for this study, including μCT data for the specimen NHMD 72014 A+B and 3-D models of individual bones segmented from the μCT data, are available in the Dryad Digital Repository: https://datadryad.org/review?doi=doi:10.5061/dryad.nh74kg8 .

REFERENCES

ADOLFSSEN, J. S., MILAN, J. and FRIEDMAN, M. 2017. Review of the Danian vertebrate fauna of southern Scandinavia. *Bulletin of the Geological Society of Denmark*, **65**, 1–23.

ALFARO, M. E., FAIRCLOTH, B. C., HARRINGTON, R. C., SORENSON, L., FRIEDMAN, M., THACKER, C. E., OLIVEROS, C. H., ČERNÝ, D. and NEAR, T. J. 2018. Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nature Ecology & Evolution*, **2**, 688–696.

ALVARADO-ORTEGA, J., CUEVAS-GARCÍA, M., DEL PILAR MELGAREJO-DAMIÁN, M., CANTALICE, K. M., ALANIZ-GALVAN, A., SOLANO-TEMPLOS, G. and THAN-MARCHESE, B. A. 2015. Paleocene fishes from Palenque, Chiapas, southeastern Mexico. *Palaeontologia Electronica*, **18**, 1–22.

ANTHONISSEN, D. E. and OGG, J. G. 2012. Cenozoic and Cretaceous biochronology of planktonic foraminifera and calcareous nannofossils. 1083–1126. *In* GRADSTEIN, F. M., OGG, J. G., SCHMITZ, M. D. and OGG, G. M. (eds). *The Geologic Time Scale 2012*. Elsevier BV, Amsterdam, 1176 pp.

ARAMBOURG, C. 1952. Les vertébrés fossiles des gisements de phosphates (Maroc-Algérie-Tunisie). *Notes et Mémoires du Service géologique du Maroc*, **92**, 1–372.

BANNIKOV, A. F. and PARIN, N. N. 1997. The list of marine fishes from Cenozoic (Upper Paleocene-Middle Miocene) localities in southern European Russia and adjacent countries. *Journal of Ichthyology*, **37**, 133–146.

BARDACK, D. 1970. A new teleost from the Oldman Formation (Cretaceous) of Alberta. *National Museums of Canada, Publications in Palaeontology*, **3**, 1–8.

BARDET, N., GHEERBRANT, E., NOUBHANI, A., CAPPETTA, H., JOUVE, S., BOURDON, E., PEREDA SUBERBIOLA, X., JALIL, N.-E., VINCENT, P., HOUSSAYE, A., SOLE, F., ELHOUSSAINI DARIF, K., ADNET, S., RAGE, J.-C., de LAPPARENT de BROIN, F., SUDRE, J., BOUYA, B., AMAGHZAZ, M. and MESLOUH, S. 2017. The vertebrates from the Cretaceous-Palaeogene (72.1-47.8 Ma) phosphates of Morocco. 351–452. *In* ZOUHRI, S. (ed). *Vertebrate Paleontology of Morocco: The state of knowledge*. Mémoires de la Société Géologique de France, **180**, 624 pp.

BENDIX-ALMGREEN, S.E. 1969. Notes on the Upper Cretaceous and Lower Tertiary fish faunas of northern West Greenland. *Meddelelser fra Dansk Geologisk Forening*, **19**, 204–217.

BERRA, T. M. 2007. *Freshwater Fish Distribution*. The University of Chicago Press, Chicago, 615 pp.

BERG L. 1940. *Classification of fishes both recent and fossil*. Travaux de l'Institut Zoologique de l'Académie des Sciences de l'URSS, Moscow, 517 pp. [Translated and reprinted in English, 1947, Ann Arbor.]

BETANCUR-R, R., ORTÍ, G. and PYRON, R. A. 2015. Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. *Ecology Letters*, **18**, 441–450.

BLANCO, A., SZABÓ, M., BLANCO–LAPAZ, À. and MARMI, J. 2017. Late Cretaceous (Maastrichtian) Chondrichthyes and Osteichthyes from northeastern Iberia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **465**, 278–294.

BONDE, N. C. 1997. A distinctive fish fauna in the basal ash-series of the Fur/Ølst Formation (U. Paleocene, Denmark). *Aarhus Geoscience*, **6**, 33–48.

BONDE, N. 2008. Osteoglossomorphs of the marine Lower Eocene of Denmark–with remarks on other Eocene taxa and their importance for palaeobiogeography. *Geological Society of London, Special Publications*, **295**, 253–310.

BRINKMAN, D. B. and NEUMAN, A. G. 2002. Teleost centra from uppermost Judith River Group (Dinosaur Park Formation, Campanian) of Alberta, Canada. *Journal of Paleontology*, **76**, 138–155.

CAPOBIANCO, A. and FRIEDMAN, M. 2018. Vicariance and dispersal in southern hemisphere freshwater fish clades: a palaeontological perspective. *Biological Reviews*, **94**.

CAPOBIANCO, A., FOREMAN, E. and FRIEDMAN, M. 2019. Data from: A Paleocene (Danian) marine osteoglossid (Teleostei: Osteoglossomorpha) from the Nuussuaq Basin of Greenland, with a brief review of Palaeogene marine bonytongue fishes. *Dryad Digital Repository*. https://datadryad.org/review?doi=doi:10.5061/dryad.nh74kg8

CAPPETTA, H. 1972. Les poissons crétacés et tertiaires du bassin des Iullemmeden (République du Niger). *Palaeovertebrata*, **5**, 179–251.

CARNEVALE, G., BANNIKOV, A. F., MARRAMÀ, G., TYLER, J. C. and ZORZIN, R. 2014. The Pesciara-Monte Postale fossil-lagerstätte: 2. Fishes and other vertebrates. *Rendiconti della Società Paleontologica Italiana*, **4**, 37–63.

CASE, G. R. 1994. Fossil fish remains from the late Paleocene Tuscahoma and early Eocene Bashi formations of Meridian, Lauderdale County, Mississippi. Part II. Teleosteans. *Palaeontographica Abteilung A*, **230**, 139–153.

CASIER, E. 1966. *Faune ichthyologique du London Clay: Texte & Atlas*. British Museum (Natural History), London, 496 pp.

CAVIN, L. 2002. Effects of the Cretaceous-Tertiary boundary event on bony fishes. 141–158. *In* BUFFETAUT, E. and KOEBERL, C. (eds). *Geological and Biological Effects of Impact Events*. Springer-Verlag, Berlin Heidelberg, 295 pp.

CAVIN, L. and MARTIN, M. 1995. Les Actinoptérygiens et la limite Crétacé-Tertiaire. *Geobios*, **28**, 183–188.

COHEN, D.M. 1984. Gadiformes: overview. *In* MOSER, H.G. (ed). *Ontogeny and systematics of fishes*. American Society of Ichthyologists and Herpetologists, Special Publication, **1**, 259–265.

DAM, G., PEDERSEN, G. K., SØNDERHOLM, M., MIDTGAARD, H. H., LARSEN, L. M., NØHR-HANSEN, H. and PEDERSEN, A. K. 2009. Lithostratigraphy of the Cretaceous–Paleocene Nuussuaq Group, Nuussuaq Basin, West Greenland. *Geological Survey of Denmark and Greenland Bulletin*, **19**, 1–171.

DANILCHENKO, P. G. 1968. Fishes of the upper Paleocene of Turkmenia. 113–156. *In* OBRUCHEV, D. V. (ed). *Essays on the Phylogeny and Systematics of Fossil Fishes and Agnathans*. Nauka Press, Moscow.

DAVIES, T. G., RAHMAN, I. A., LAUTENSCHLAGER, S., CUNNINGHAM, J. A., ASHER, R. J., BARRETT, P. M., BATES, K. T., BENGTSON, S., BENSON, R. B., BOYER, D. M., BRAGA, J., BRIGHT, J. A., CLAESSENS, L. P. A. M., COX, P. G., DONG, X.-P., EVANS, A. R., FALKINGHAM, P. L., FRIEDMAN, M., GARWOOD, R. J., GOSWAMI, A., HUTCHINSON, J. R., JEFFERY, N. S., JOHANSON, Z., LEBRUN, R., MARTINEZ-PÉREZ, C., MARUGÁN-LOBÓN, J., O'HIGGINS, P. M., METSCHER, B., ORLIAC, M., ROWE, T.

B., RÜCKLIN, M., SÁNCHEZ-VILLAGRA, M. R., SHUBIN, N. H., SMITH, S. Y., STARCK, J. M., STRINGER, C., SUMMERS, A. P., SUTTON, M. D., WALSH, S. A., WEISBECKER, V., WITMER, L. M., WROE, S., YIN, Z., RAYFIELD, E. J. and DONOGHUE, P. C. J. 2017. Open data and digital morphology. *Proceedings of the Royal Society B*, **284**, 20170194.

DAWSON, M. R., WEST, R. M., LANGSTON, W. and HUTCHISON, J. H. 1976. Paleogene terrestrial vertebrates: northernmost occurrence, Ellesmere Island, Canada. *Science*, **192**, 781–782.

FOREY, P. L. and HILTON, E. J. 2010. Two new Tertiary osteoglossid fishes (Teleostei: Osteoglossomorpha) with notes on the history of the family. 215–246. *In* ELLIOTT, D. K., MAISEY, J. G., YU, X. and MIAO, D. (eds). *Morphology, Phylogeny and Paleobiogeography of Fossil Fishes*. Verlag Dr. F. Pfeil, München, 472 pp.

FIERSTINE, H. L. 2006. Fossil history of billfishes (Xiphioidei). *Bulletin of Marine Science*, **79**, 433–453.

FRIEDMAN, M. 2009. Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. *Proceedings of the National Academy of Sciences*, **106**, 5218–5223.

FRIEDMAN, M. 2010. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 1675–1683.

FRIEDMAN, M. and CARNEVALE, G. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. *Journal of the Geological Society*, **175**, 569–579.

FRIEDMAN, M. and SALLAN, L. C. 2012. Five hundred million years of extinction and recovery: a Phanerozoic survey of large-scale diversity patterns in fishes. *Palaeontology*, **55**, 707–742.

FRIEDMAN, M., BECKETT, H. T., CLOSE, R. A. and JOHANSON, Z. 2016. The English chalk and London clay: Two remarkable British bony fish Lagerstätten. *Geological Society of London, Special Publications*, **430**, 165–200.

KERSHAW, D. R. 1970. The cranial osteology of the 'butterfly fish', *Pantodon buchholzi* Peters. *Zoological Journal of the Linnean Society*, **49**, 5–19.

KERSHAW, D. R. 1976. A structural and functional interpretation of the cranial anatomy in relation to the feeding of osteoglossoid fishes and a consideration of their phylogeny. *The Transactions of the Zoological Society of London*, **33** (3), 173–252.

KHALLOUFI, B., EL HOUSSAINI DARIF, K., JOURANI, E., KHALDOUNE, F. and JALIL, N. E. 2018. A new Palaeocene Megalopidae (Teleostei, Elopomorpha) from the phosphate basins of Morocco. Historical Biology, 1–10.

KUMAR, K., RANA, R. S. and PALIWAL, B. S. 2005. Osteoglossid and lepisosteid fish remains from the Paleocene Palana Formation, Rajasthan, India. *Palaeontology*, **48**, 1187–1210.

GAYET, M. and MEUNIER, F. J. 1998. Maastrichtian to early late Paleocene freshwater Osteichthyes of Bolivia: additions and comments. 85–110. *In* MALABARBA, L. R., REIS, R.

E., VARI, R. P., LUCENA, Z. M. and LUCENA, C. A. S. (eds). *Phylogeny and Classification of Neotropical Fishes*. EdiPUCRS, Porto Alegre, 603 pp.

GREENWOOD P. H., ROSEN D. E., WEITZMAN S. H. and MYERS G. S. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History*, **131**, 339–455.

GUINOT, G. and CAVIN, L. 2016. 'Fish' (Actinopterygii and Elasmobranchii) diversification patterns through deep time. *Biological Reviews*, **91**, 950–981.

GUINOT, G. and CAVIN, L. 2018. Body size evolution and habitat colonization across 100 million years (Late Jurassic–Paleocene) of the actinopterygian evolutionary history. *Fish and Fisheries*, **19**, 577–597.

HASTINGS, A. K., BLOCH, J. I. and JARAMILLO, C. A. 2011. A new longirostrine dyrosaurid (Crocodylomorpha, Mesoeucrocodylia) from the Paleocene of north-eastern Colombia: biogeographic and behavioural implications for New-World Dyrosauridae. *Palaeontology*, **54**, 1095–1116.

HILTON, E. J. 2003. Comparative osteology and phylogenetic systematics of fossil and living bony-tongue fishes (Actinopterygii, Teleostei, Osteoglossomorpha). *Zoological Journal of the Linnean Society*, **137**, 1–100.

HILTON, E. J. and LAVOUé, S. 2018. A review of the systematic biology of fossil and living bony-tongue fishes, Osteoglossomorpha (Actinopterygii: Teleostei). *Neotropical Ichthyology*, **16**, e180031.

LI, G.-Q. 1994. Systematic position of the Australian fossil osteoglossid fish †*Phareodus* (= *Phareoides*) *queenslandicus* Hills. *Memoirs of the Queensland Museum*, **37**, 287–300.

LI, G.-Q. 1996. A new species of Late Cretaceous osteoglossid (Teleostei) from the Oldman Formation of Alberta, Canada, and its phylogenetic relationships. 285–298. *In* ARRATIA, G. and VIOHL, G. (eds). *Mesozoic Fishes. Systematics and Paleoecology*. Verlag Dr. F. Pfeil, München, 576 pp.

LI, G.-Q., GRANDE, L. and WILSON, M. V. H. 1997. The species of †*Phareodus* (Teleostei: Osteoglossidae) from the Eocene of North America and their phylogenetic relationships. *Journal of Vertebrate Paleontology*, **17**, 487–505.

LONGRICH, N. R., SCRIBERAS, J. and WILLS, M. A. 2016. Severe extinction and rapid recovery of mammals across the Cretaceous–Palaeogene boundary, and the effects of rarity on patterns of extinction and recovery. *Journal of Evolutionary Biology*, **29**, 1495–1512.

MARTIN, J. E., SARR, R. and HAUTIER, L. 2019. A dyrosaurid from the Paleocene of Senegal. *Journal of Paleontology*, **93**, 343–358.

MIYA, M., FRIEDMAN, M., SATOH, T. P., TAKESHIMA, H., SADO, T., IWASAKI, W., YAMANOUE, Y., NAKATANI, M., MABUCHI, K., INOUE, J. G., POULSEN, J. Y., FUKUNAGA, T., SATO, Y. and NISHIDA, M. 2013. Evolutionary origin of the Scombridae (tunas and mackerels): members of a Paleogene adaptive radiation with 14 other pelagic fish families. *PLoS ONE*, **8**, e73535.

MONSCH, K. A. and BANNIKOV, A. F. 2011. New taxonomic synopses and revision of the scombroid fishes (Scombroidei, Perciformes), including billfishes, from the Cenozoic of territories of the former USSR. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, **102**, 253–300.

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.

NØHR-HANSEN, H. and SHELDON, E. 2000. Palyno- and nannostratigraphic dating of the marine Paleocene succession in the Nuussuaq Basin, West Greenland. *GFF*, **122**, 115–116.

NOLF, D. 2013. *The Diversity of Fish Otoliths, Past and Present*. Royal Belgian Institute of Natural Sciences, Brussel, 581 pp.

NOLF, D. and CAPPETTA, H. 1976. Observations nouvelles sur les otolithes des téléostéens du calcaire Grossier (Eocène du Bassin de Paris). *Geobios*, **9**, 251–277.

NOLF, D. and STRINGER, G. L. 1996. Cretaceous fish otoliths–a synthesis of the North American record. 433–459. *In* ARRATIA, G. and VIOHL, G. (eds). *Mesozoic Fishes. Systematics and Paleoecology*. Verlag Dr. F. Pfeil, München, 576 pp.

o'leary, M. A., Bouaré, M. L., Claeson, K. M., Heilbronn, K., hill, r. v., mccartney, j., sessa, j. a., sissoko, f., tapanila, l., wheeler, e. and ROBERTS, E. M. 2019. Stratigraphy and paleobiology of the Upper Cretaceous–Lower Paleogene sediments from the Trans-Saharan Seaway in Mali. *Bulletin of the American Museum of Natural History*, **436**, 1–177.

PATTERSON, C. 1993. An overview of the early fossil record of acanthomorphs. *Bulletin of Marine Science*, **52**, 29–59.

PATTERSON, C. & LONGBOTTOM, A. E. 1989. An Eocene amiid fish from Mali, West Africa. *Copeia*, **1989**, 827–836.

PRASAD, G. V. R. 1989. Vertebrate fauna from the infra- and intertrappean beds of Andhra Pradesh: age implications. *Journal of the Geological Society of India*, **34**, 161–173.

RAGE, J. C., BAJPAI, S. M. T. J. G., THEWISSEN, J. G. and TIWARI, B. N. 2003. Early Eocene snakes from Kutch, Western India, with a review of the Palaeophiidae. *Geodiversitas*, **25**, 695–716.

RIBEIRO, E., DAVIS, A. M., RIVERO-VEGA, R. A., ORTÍ, G. and BETANCUR-R, R. 2018. Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. *Proceedings of the Royal Society B: Biological Sciences*, **285**, 20182010.

ROELLIG, H. F. 1974. The cranial osteology of *Brychaetus muelleri* (Pisces Osteoglossidae) Eocene, Isle of Sheppey. *Journal of Paleontology*, **48**, 947–951.

ROSE, K. D. 1981. Composition and species diversity in Paleocene and Eocene mammal assemblages: an empirical study. *Journal of Vertebrate Paleontology*, **1**, 367–388.

ROSEN, D. E. and PATTERSON, C. 1969. The structure and relationships of the paracanthopterygian fishes. *Bulletin of the American Museum of Natural History*, **141**, 357–454.

ROSENKRANTZ, A. 1970. Marine Upper Cretaceous and lowermost Tertiary deposits in West Greenland. *Bulletin of the Geological Society of Denmark*, **19**, 406–453.

SAHNI, A. and BAJPAI, S. 1988. Cretaceous-Tertiary boundary events: The fossil vertebrate, palaeomagnetic and radiometric evidence from peninsular India. *Journal of Geological Society of India*, **32**, 382–396.

SCHWARZHANS, W. 2004. Fish otoliths from the Paleocene (Selandian) of West Greenland. *Meddelelser om Grønland, Geoscience*, **42**, 1–32.

SCHWARZHANS, W. 2010. Otolithen aus den Gerhartsreiter Schichten (Oberkreide: Maastricht) des Gerhartsreiter Grabens (Oberbayern). *Palaeo Ichthyologica*, **4**, 1–100.

SCHWARZHANS, W. 2018. A review of Jurassic and Early Cretaceous otoliths and the development of early morphological diversity in otoliths. *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen*, **287**, 75–121.

SCHWARZHANS, W. and MILÀN, J. 2017. After the disaster: Bony fish remains (mostly otoliths) from the K/Pg boundary section at Stevns Klint, Denmark, reveal consistency with teleost faunas from later Danian and Selandian strata. *Bulletin of the Geological Society of Denmark*, **65**, 59–74.

SIBERT, E., FRIEDMAN, M., HULL, P., HUNT, G. and NORRIS, R. 2018. Two pulses of morphological diversification in Pacific pelagic fishes following the Cretaceous–Palaeogene mass extinction. *Proceedings of the Royal Society B: Biological Sciences*, **285**, 20181194.

SPIELHAGEN, R. F. and TRIPATI, A. 2009. Evidence from Svalbard for near-freezing temperatures and climate oscillations in the Arctic during the Paleocene and Eocene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **278**, 48–56.

STINTON, F. C. 1977. Fish otoliths from the English Eocene, II. *Palaeontographical Society Monographs*, **548**, 57–126.

TAVERNE L. 1977. Ostéologie, phylogénèse et systématique des Téléostéens fossiles et actuels du super-ordre des Ostéoglossomorphes, Première partie. Ostéologie des genres *Hiodon*, *Eohiodon*, *Lycoptera*, *Osteoglossum*, *Scleropages*, *Heterotis* et *Arapaima*. *Académie Royale de Belgique, Mémoires de la Classe des Sciences, Collection in-8° - 2e série*, **42**, 1–235.

TAVERNE L. 1978. Ostéologie, phylogénèse et systématique des Téléostéens fossiles et actuels du super-ordre des Ostéoglossomorphes, Deuxième partie. Ostéologie des genres *Phareodus*, *Phareoides*, *Brychaetus*, *Musperia*, *Pantodon*, *Singida*, *Notopterus*, *Xenomystus* et *Papyrocranus*. *Académie Royale de Belgique, Mémoires de la Classe des Sciences, Collection in-8° - 2e série*, **42**, 1–213.

TAVERNE, L. 1998. Les ostéoglossomorphes marins de l'Éocène du Monte Bolca (Italie): *Monopteros* Volta 1796, *Thrissopterus* Heckel, 1856 et *Foreyichthys* Taverne, 1979. Considérations sur la phylogénie des téléostéens ostéoglossomorphes. *Studi e Ricerche sui Giacimenti Terziari di Bolca, Miscellanea Paleontologica*, **7**, 67–158.

TAVERNE L. 2009*a*. *Ridewoodichthys*, a new genus for *Brychaetus caheni* from the marine Paleocene of Cabinda (Africa): re-description and comments on its relationships within the

Osteoglossidae (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre*, **79**, 147–153.

TAVERNE, L. 2009*b*. On the presence of the osteoglossid genus *Scleropages* in the Paleocene of Niger, Africa (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre*, **79**, 161–167.

TAVERNE, L. 2016. New data on the osteoglossid fishes (Teleostei, Osteoglossiformes) from the marine Danian (Paleocene) of Landana (Cabinda Enclave, Angola). *Geo-Eco-Trop*, **40**, 297–304.

TAVERNE, L. and CAPASSO, L. 2012. Osteology and relationships of *Prognathoglossum kalassyi* gen. & sp. nov. (Teleostei, Osteoglossiformes, Pantodontidae) from the marine Cenomanian (Upper Cretaceous) of En Nammoura (Lebanon). *Cybium*, **36**, 563–574.

TAVERNE, L., KUMAR, K. and RANA, R. S. 2009. Complement to the study of the Indian Paleocene osteoglossid fish genus *Taverneichthys* (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre*, **79**, 155–160.

WEEMS, R. E. 1999. Actinopterygian fishes from the Fisher/Sullivan site. 53–100. *In* WEEMS, R. E. and GRIMSLEY, G. J. (eds). *Early Eocene Vertebrates and Plants from the Fisher/Sullivan Site (Nanjemoy Formation) Stafford County, Virginia*. Virginia Division of Mineral Resources, Publication 152, 159 pp.

WEEMS, R. E. and HORMAN, S. R. 1983. Teleost fish remains (Osteoglossidae, Blochiidae, Scombridae, Triodontidae, Diodontidae) from the Lower Eocene Nanjemoy Formation of Maryland. *Proceedings of the Biological Society of Washington*, **96**, 38–49.

WILSON, M.V.H. and MURRAY, A.M. 2008. Osteoglossomorpha: phylogeny, biogeography, and fossil record and the significance of key African and Chinese fossil taxa. *Geological Society of London, Special Publications*, **295**, 185–219.

ZACHOS, J., PAGANI, M., SLOAN, L., THOMAS, E. and BILLUPS, K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686–693.

ZHANG, J.-Y. 2003. First *Phareodus* (Osteoglossomorpha: Osteoglossidae) from China. *Vertebrata PalAsiatica*, **41**, 327–334.

ZHANG, L., HAY, W. W., WANG, C. and GU, X. 2019. The evolution of latitudinal temperature gradients from the latest Cretaceous through the Present. *Earth-Science Reviews*, 1–12.

CHAPTER 4

A Long-Snouted Marine Bonytongue (Teleostei: Osteoglossidae) from the Early Eocene of Morocco and the Phylogenetic Affinities of Marine Osteoglossids

Abstract

Osteoglossid bonytongues (arapaimas, arowanas, and relatives) are extant tropical freshwater fishes with a relatively abundant and diverse fossil record. Most osteoglossid fossils come from a 25-million-year interval in the early Palaeogene, when these fishes were distributed worldwide in both freshwater and marine environments. Despite their biogeographic and palaeoecological relevance, and a relative abundance of well-preserved material, the evolutionary relationships between these Palaeogene forms and extant bonytongues remain unclear. Here we describe a new genus of bonytongue from early Eocene marine deposits of Morocco, represented by an articulated, three-dimensionally preserved skull with associated pectoral girdle. This taxon is characterized by an elongated snout, contrasting with the short jaws usually found in marine representatives of the clade. A revision of morphological characters in bonytongues allows us to place this new genus, together with other marine and freshwater Eocene taxa, within crown osteoglossids and closely related to extant arapaimines. The discovery of the new Moroccan taxon hints at a previously underestimated eco-morphological diversity of marine bonytongues, highlighting the diverse trophic niches that these fishes occupied in early Palaeogene seas.

INTRODUCTION

Osteoglossomorpha (bonytongue fishes) is one of the earliest diverging clades of crown teleost fishes (Arratia, 1997), with a long evolutionary history that extends to at least the Middle Jurassic (Capobianco & Friedman, 2019). The low species diversity of modern osteoglossomorphs contrasts with their remarkable diversity of form (i.e., disparity), ranging from the unassuming mooneyes to the gigantic arapaima to the electrical elephantfishes. Despite

this disparity, all extant species are ecologically restricted to freshwater environments (a few species of notopterid knifefishes are occasionally found in brackish waters; Berra, 2007) in mostly tropical areas with the exception of two species of temperate-adapted mooneyes. In contrast to most groups of tropical freshwater fishes, osteoglossomorphs are known from numerous fossil species, many of which are represented by relatively well-preserved, articulated specimens. In fact, extinct bonytongue genera surpass extant ones in number (Murray & Wilson, 2008; Hilton & Lavoué, 2018).

Perhaps the most surprising feature of paleontological record of bonytongues is the presence of several fossils (including well-preserved, articulated skeletons) in marine deposits worldwide (see Capobianco *et al*., 2019 for a review of marine osteoglossomorph occurrences). The quantity and preservational quality of these specimens, as well as the range of marine environments represented by these deposits (ranging from estuarine and lagoonal to offshore pelagic), suggest that their presence in marine depositional settings is not an artifact of taphonomic processes like post-mortem transport. Remarkably, these marine occurrences are narrowly restricted to a ~25 million year interval in the early Palaeogene (with few dubious exceptions; see Capobianco *et al*., 2019). Although fossil marine bonytongues have been known for almost two centuries (Agassiz, 1845; Woodward, 1901), their taxonomic diversity and widespread geographic distribution have become apparent only in the last two decades (Taverne, 1998; Bonde, 2008; Forey & Hilton, 2010). Several of these marine forms can be confidently assigned to the osteoglossomorph sub-clade Osteoglossidae (*sensu* Forey & Hilton, 2010) due to the presence of anatomical features diagnostic of the family (Forey & Hilton, 2010; Hilton & Lavoué, 2018; Capobianco & Friedman, 2019).

Osteoglossidae currently includes only four genera and around 10 species, distributed in tropical freshwater areas worldwide. Within Osteoglossidae, two distinct clades can be recognized: Osteoglossinae and Arapaiminae. The former comprises the South American *Osteoglossum* and the Southeast Asian and northern Oceanian *Scleropages*, whereas the latter comprises the South American *Arapaima* and the African *Heterotis*. The disjunct geographic distribution of extant osteoglossids has sparked the interest of several researchers investigating underlying biogeographic processes. Time-calibrated molecular and total-evidence phylogenies suggest that the divergences between extant genera postdate major breakups of the Gondwanan

supercontinent (such as the separation between West Gondwana and East Gondwana during the Late Jurassic – Early Cretaceous, or the South America–Africa breakup in the Early Cretaceous; Blakey, 2008), implying that continental vicariance is an unlikely explanation for the current distribution of osteoglossids (Lavoué, 2015; 2016). The fossil record of Osteoglossinae and Arapaiminae is consistent with these results, but caution should be applied when interpreting it at face value due to incompleteness. Fossil osteoglossines belonging to the genus *Scleropages* are known from complete articulated specimens from the early Eocene of China (Zhang & Wilson, 2017; Zhang, 2020), whereas *Osteoglossum* is unknown from the fossil record. Remains of fossil arapaimines are mostly fragmentary, with fragments of *Heterotis* found in Afro-Arabian deposits of at most Oligocene age (Otero & Gayet, 2001; Otero *et al*., 2017), and specimens of *Arapaima* known from the Miocene of Brazil (Lundberg & Chernoff, 1992). The earliest putative occurrence of arapaimines consists of jaw fragments and squamules from the latest Cretaceous (Maastrichtian) El Molino Formation of Bolivia (Gayet & Meunier, 1998; Gayet *et al*., 2001). This material was attributed to indeterminate arapaimines, but its identity remains uncertain. These incomplete remains are potentially joined by articulated fossils of *Sinoglossus* from the late Eocene-Oligocene Lushan Formation of China (Su, 1986). This taxon has been surprisingly recovered as an arapaimine in phylogenetic analyses of bonytongues (Murray & Wilson, 2008). The presence of a freshwater arapaimine in continental Asia adds complexity to the biogeographic history of this clade, and it is difficult to interpret from a purely vicariant perspective.

In addition to the scarce record of osteoglossines and arapaimines, the fossil record of Osteoglossidae contains several taxa (including the marine forms) that cannot be easily placed in either of the two extant sub-clades. The most well-studied of these is the freshwater †*Phareodus*, known from hundreds of complete specimens from the early middle Eocene (Wasatchian-Bridgerian North American Land Mammal Ages, overlapping the Ypresian and Lutetian of the global timescale) Green River Formation of Wyoming and Utah, USA, where it is represented by two distinct species, †*P. encaustus* and †*P. testis* (Li *et al*., 1997a). Other species referred to †*Phareodus* are found in the Ypresian Yangxi Formation of China (†*P. songziensis*; Zhang, 2003) and in the late Paleocene-early Eocene Redbank Plains Formation of Australia (†*P. queenslandicus*; Li, 1994). Alvarado-Ortega *et al*. (2015) report a potential marine representative of †*Phareodus* in the Danian Tenejapa Formation of Mexico, but further study is needed to

confirm its generic status. Several taxa similar to †*Phareodus* have been described from both freshwater and marine deposits around the world. Among these, the marine †*Brychaetus* (Ypresian) and the freshwater †*Musperia* (Eocene, age indeterminate) were included, together with the aforementioned †*Phareodus*, in an osteoglossid subclade coined †Phareodontinae by Taverne (1979). The diagnosis of this taxon includes several osteological features, such as a relatively short skull and jaws, a lateral expansion of the frontal, an elongation of the occipital region of the neurocranium, the presence of a paired dorso-occipital depression or fossa, an autogenous articular and a third infraorbital smaller than the fourth one (Taverne, 1979). Other fossil bonytongues have been subsequently proposed to belong to †Phareodontinae: the freshwater †*Cretophareodus* (Campanian), †*Phareodusichthys* (Maastrichtian–Danian) and †*Taverneichthys* (Paleocene, age indeterminate), and the marine †*Ridewoodichthys* (Selandian) (Li, 1996; Kumar *et al*., 2005; Taverne, 2009; Forey & Hilton, 2010). These are joined by a variety of marine taxa that might be included in this clade, or be closely related to it: †*Magnigena* (Thanetian), †*Brychaetoides* (earliest Ypresian), †*Xosteoglossid* (earliest Ypresian), †*Monopteros* (Ypresian), †*Opsithrissops* (Paleocene–Eocene thermal maximum) and a few unnamed taxa (Bonde, 2008; Forey & Hilton, 2010; Hilton & Lavoué, 2018). However, most species included in or referred to †Phareodontinae have never been added to a formal phylogenetic analysis, and their systematic placement remains dubious. Additionally, there are other marine bonytongues that might not be related to †phareodontines or even to osteoglossids (Hilton & Lavoué, 2018). Among these, †*Furichthys* from the earliest Eocene Fur Formation of Denmark was described as a basal osteoglossiform by Bonde (2008) and is unique among marine bonytongues for its elongated preorbital region of the skull (or snout), contrasting with the short-snouted condition seen in †phareodontines as described by Taverne (1979).

The existence of several different forms of marine bonytongues in the early Paleogene has been linked to the hypothesis of marine dispersal as main driver of the current disjunct distribution of osteoglossids (Patterson, 1975; Bonde, 2008; Hilton & Lavoué, 2018; Capobianco & Friedman, 2019). However, the lack of a robust phylogenetic framework for marine bonytongues has precluded a test of this hypothesis. Phylogenetic relationships are uncertain even for fossil osteoglossids known from numerous well-preserved specimens, such as †*Phareodus* and †*Brychaetus*. Past studies recovered †*Phareodus* as sister-taxon to osteoglossines (Lavoué, 2016), sister-taxon to arapaimines (Wilson & Murray, 2008), as a stem osteoglossid (Murray *et*

al., 2018), or closely related to the butterflyfish *Pantodon* (Hilton, 2003). This uncertainty stems mainly from the peculiar mix of osteoglossine-like and arapaimine-like characters of †*Phareodus*, and the difficulty of inferring which of these are plesiomorphic for osteoglossids. Selection of taxa and characters employed in phylogenetic analyses has also been shown to have a strong influence on the position of fossil taxa in the osteoglossomorph tree (Murray *et al*., 2018). †*Brychaetus* has been interpreted as very closely related to †*Phareodus*, to the point of the former being considered a junior synonym of the latter (Li *et al*., 1997a). However, recent works on bonytongue systematics do not include †*Brychaetus* (Hilton, 2003; Wilson & Murray, 2008; Murray *et al*., 2018).

Here we describe a new genus and species of osteoglossid from early Eocene marine deposits of Morocco, based on a three-dimensionally preserved and articulated skull with pectoral girdle. This taxon bears some similarity with the Danish †*Furichthys* in having an elongated preorbital region and a long lower jaw, probably indicative of a feeding ecology very distinct from that of other short-faced marine bonytongues. New anatomical observations, as well as the reexamination of key taxa such as †*Phareodus*, †*Brychaetus*, and †*Furichthys*, strongly suggest that the new species clusters with other fossil bonytongues (both marine and freshwater) as sister-group to the arapaimines. This widely distributed and ecomorphologically diverse clade of bonytongues points to an unexpected radiation of these fishes in the early Palaeogene, possibly as a consequence of ecological opportunity and release after the Cretaceous/Palaeogene (K/Pg) mass extinction.

MATERIALS AND METHODS

Micro-computed tomography

The holotype of †*Macroprosopon hiltoni*, as well as comparative material of extinct and extant osteoglossomorphs, was imaged using a Nikon XT H 225ST industrial µCT scanner at the University of Michigan CTEES facility (Computed Tomography in Earth & Environmental Sciences). Individual scanning parameters are given below:

†*Macroprosopon hiltoni* UMMP 118216. Voltage, 215 kV, current, 109 µA; filter, 2.5 mm copper; reflection target, tungsten; effective pixel size, 122.7 µm. Additional scans were

performed on smaller regions of interest of the specimen, with effective pixel sizes ranging from 40.3 to 64.0 μm.

†*Brychaetus muelleri* NHMUK PV P641. Voltage, 190 kV; current, 305 μA; filter, 2.7 mm copper; reflection target, tungsten; effective pixel size, 62.9 μm.

cf. †*Brychaetus* sp. NHMUK PV P26758. Voltage, 200 kV; current, 205 μA; filter, 2.1 mm copper; reflection target, tungsten; effective pixel size, 42.7 μm.

†*Phareodus encaustus* FMNH PF 11947. Voltage, 210 kV; current, 115 μA; filter, 1.0 mm copper; reflection target, tungsten; effective pixel size, 92.2 μm.

†*Phareodus encaustus* FMNH PF 11949. Voltage, 200 kV; current, 108 μA; filter, 1.5 mm copper; reflection target, tungsten; effective pixel size, 73.9 μm. Both scans of †*Phareodus encaustus* specimens were not particularly informative, as these fossils from Green River Formation are extremely flattened and their depth is not sufficient to distinguish relevant anatomical features in tomograms.

Hiodon tergisus UMMZ 247425. Voltage, 110 kV; current, 165 μA; filter, none; reflection target, tungsten; effective pixel size, 40.9 μm.

Chitala blanci UMMZ 232272. Voltage, 180 kV; current, 170 μA; filter, none; reflection target, tungsten; effective pixel size, 47.2 μm.

Petrocephalus simus UMMZ 200167. Voltage, 55 kV; current, 195 μA; filter, none; reflection target, tungsten; effective pixel size, 12.5 μm.

Pantodon buchholzi UMMZ 249782. Voltage, 65 kV; current, 195 μA; filter, none; reflection target, tungsten; effective pixel size, 14.9 μm.

Heterotis niloticus UMMZ 195004. Voltage, 105 kV; current, 155 μA; filter, 0.1 mm copper; reflection target, tungsten; effective pixel size, 35.8 μm.

Scans were acquired using Inspect-X and reconstructed using CT Pro 3-D (Nikon Metrology, USA). Additionally, reconstructed tomograms for *Arapaima gigas* UF 33107 (Morphosource media M51346) and *Osteoglossum bicirrhosum* UF 189007 (Morphosource media M26520) were downloaded from Morphosource.

Reconstructed datasets were visualized and segmented using Mimics v. 19.0 (Materialise, Belgium). Models of segmented skeletal elements were exported as surface files (.ply) and rendered as high-quality images in Blender v. 2.79 (blender.org).

Fossil preparation

Mechanical preparation of UMMP 118216 was conducted by Dr. William Sanders (chief vertebrate preparator, UMMP), using mounted carbide needles under a binocular microscope.

Specimens examined

In addition to the material listed above, the following skeletonized specimens of extant taxa and fossil specimens of extinct taxa belonging to Osteoglossomorpha were examined as comparative material:

Arapaima gigas UMMZ 177540, UMMZ 203831; †*Brychaetus muelleri* NHMUK PV 39448, NHMUK PV 39699, NHMUK PV P638, NHMUK PV P641, NHMUK PV P1748, NHMUK PV P3893, NHMUK PV P66889, SM C 21208, SM C 21209; cf. †*Brychaetus* sp. NHMD 28907, NHMUK PV P26758, NHMUK PV P66355; †*Brychaetus*? sp. NHMUK PV P73087, NHMUK PV P73088; †*Furichthys fieldsoei* FUM-N 1440, FUM-N 1848A; *Heterotis niloticus* UMMZ 213845; *Hiodon tergisus* UMMZ 180315; *Marcusenius macrolepidotus* UMMZ 200066; *Mormyrus lacerda* UMMZ 200084; *Osteoglossum bicirrhosum* UF 189007, UMMZ 203832; †*Phareodus encaustus* AMNH 4587, AMNH 19441, FMNH PF 10237, FMNH PF 10255, FMNH PF 10256, FMNH PF 10257, FMNH PF 10285, FMNH PF 11946, FMNH PF 12683, FMNH PF 13321, FMNH PF 16527, FMNH PF 16528, FMNH PF 16529, FMNH PF 16538, NHMUK PV P64636I-II; †*Phareodus testis* FMNH PF 11942, FMNH PF 16535, FMNH PF 16536, FMNH PF 16540, FMNH PF 17493, FMNH PF 17496, FMNH PF 17500, NHMUK PV P61230; *Scleropages formosus* UMMZ 203833, UMMZ 213853.

Phylogenetic analysis

The phylogenetic analysis performed in this study draws on the morphological character dataset by Murray *et al.* (2018), with modifications listed below. This dataset is itself the latest iteration of a character matrix first assembled by Wilson & Murray (2008) by combining the matrices of

Li *et al.* (1997b) and Hilton (2003); it has been modified subsequently in several descriptive studies on fossil osteoglossomorphs (Murray *et al.*, 2010; 2016; 2018). The taxa included in this analysis are mostly the same as those included by Murray *et al*. (2018), with the following exceptions. †*Tanolepis* was excluded because of its potential synonymy with †*Paralycoptera* and because of our inability to verify the scoring of its characters in the matrix (see "*Rescoring of* †*Paralycoptera and exclusion of* †*Tanolepis*" in Results). †*Ostariostoma* was also excluded from the analysis because its bonytongue affinities are questionable, as some of its anatomical features would be unique among osteoglossomorphs and its vertebral morphology is more concordant with a basal ostariophysan identification (Murray *et al*., 2018). A broader taxonomic sampling outside Osteoglossomorpha would be needed to test the phylogenetic affinities of †*Ostariostoma*, which is beyond the scope of this paper. We added to the Murray *et al*. (2018) character matrix the recently redescribed †*Laeliichthys* from the Early Cretaceous of Brazil, using the scoring of Brito *et al*. (2020) plus the additional characters added for this study. This taxon was originally described as a close relative to arapaimines (Taverne, 1979), but it has been recently reinterpreted as the sister taxon to notopterid knifefishes (Brito *et al*., 2020). Apart from the new taxon described here, we included two additional marine bonytongues, †*Brychaetus* and †*Furichthys*, and based their character scoring on direct observation of specimens, literature, and μCT data (for †*Brychaetus*). Whereas Wilson & Murray (2008) and most subsequent iterations of their character matrix lumped three mormyrid genera (*Petrocephalus*, *Gnathonemus* and *Campylomormyrus*) into the Operational Taxonomic Unit (OTU) 'Mormyroidea' and three notopterid genera (*Chitala*, *Xenomystus* and *Papyrocranus*) into the OTU 'Notopteridae', we decided to keep these taxa distinct at genus level in our analysis. As in Murray *et al*. (2018), we included *Amia*, †Ellimmichthyiformes, Clupeiformes, and *Elops* as our sample of non-osteoglossomorphs. Because the character scoring for Clupeiformes in Murray *et al*. (2018) is based exclusively on the anatomy of *Dorosoma cepedianum* (as described and figured in Grande, 1985), we changed the name of that OTU from 'Clupeiformes' to '*Dorosoma*'. The extant holostean *Amia calva* was selected as the outgroup to all other taxa included in the analysis (that is, all trees were rooted *a posteriori* on *Amia*). The data matrix, which ultimately comprised 96 characters for 34 taxa, was assembled and edited in Mesquite v. 3.61 (Maddison & Maddison, 2019).

For phylogenetic reconstruction, the character matrix was analyzed through maximum parsimony (MP), maximum likelihood (ML) and Bayesian approaches. The MP analysis was performed in PAUP* v. 4.0a169 (Swofford, 2002). All characters were designated as unweighted and unordered, except for character 96 (number of branchiostegal rays), which was ordered along a numerical morphocline. Multiple states of a character in a single taxon were treated as polymorphisms. MP trees were found with a heuristic search, using random stepwise addition (100 replicates, 10 trees held at each step) and tree-bisection-reconnection branch swapping algorithm. Support for the results of the MP analysis was evaluated by calculating Bremer decay indices for every node. Additionally, 1000 bootstrap replicates were run and visualized with a bootstrap consensus tree including all groups compatible with the 50% majority-rule consensus tree.

The ML analysis was performed in IQ-TREE, using its dedicated web server (Trifinopoulos *et al*., 2016). The Mkv model (Markov k model with only variable characters) was used as model of character evolution. A gamma-distributed rate model with four rate categories was used to account for rate variability across characters. Node support was evaluated with 1000 ultrafast bootstrap (Hoang *et al*., 2018) replicates.

The Bayesian phylogenetic analysis was conducted in MrBayes v. 3.2.7 (Ronquist *et al*., 2012). As in the ML analysis, an Mkv model with gamma-distributed rates (four rate categories) was chosen for the analysis. Like in the MP analysis, character 96 was set as ordered. Two simultaneous analyses were run for 10 million generations, sampling every 1000 generations. Maximum standard deviation of split frequencies between the two runs reached <0.02 after 2 million generations, indicating good convergence. The first 25% of sampled trees and parameters were discarded as burn-in. Posterior probabilities were visualized on a consensus majority-rule tree showing all compatible partitions.

All phylogenetic trees were visualized using FigTree v. 1.4.4 (Rambaut, 2012).

Institutional abbreviations

FMNH, The Field Museum, Chicago, IL, USA; FUM, Fur Museum, Fur, Denmark; NHMD, Natural History Museum of Denmark, Copenhagen, Demark; NHMUK, The Natural History Museum, London, UK; SM, Sedgwick Museum of Earth Sciences, Cambridge, UK; UF, Florida

Museum, Gainesville, FL, USA; UMMP, University of Michigan Museum of Paleontology, Ann Arbor, MI, USA; UMMZ, University of Michigan Museum of Zoology, Ann Arbor, MI, USA.

Dagger symbols

Following the convention of Patterson & Rosen (1977), the dagger symbol (†) precedes extinct taxa.

SYSTEMATIC PALAEONTOLOGY

TELEOSTEI Müller, 1845

OSTEOGLOSSOMORPHA Greenwood, Rosen, Weitzman & Myers, 1966

OSTEOGLOSSIFORMES Berg, 1940

OSTEOGLOSSIDAE Berg, 1940

†PHAREODONTINAE Taverne, 1979

†***Macroprosopon*** Capobianco *et al.*, GEN. NOV.

Type species: †*Macroprosopon hiltoni* (monotypic)

Etymology: Generic name from the combination of the Ancient Greek *makrós* (‘long’) and *prósōpon* (‘face’), referring to the elongated snout.

Diagnosis: As for the type species.

†***Macroprosopon hiltoni*** Capobianco *et al.*, SP. NOV.

Holotype: UMMP 118216, an almost complete and three-dimensionally preserved skull articulated with part of pectoral girdle and axial skeleton.

Etymology: Specific name in honour of Eric J. Hilton (Virginia Institute of Marine Science and College of William and Mary), in recognition of his fundamental contributions on bonytongue comparative anatomy and systematics, and ichthyology in general.

Type locality/horizon: Due to non-specialist private collection of the specimen, information on the locality for UMMP 118216 is limited to the early Eocene (Ypresian) phosphates of the Ouled Abdoun Basin, Morocco (Fig. 4.1). The surrounding matrix provides two lines of corroborative evidence. First, a slightly deformed and damaged shark tooth embedded in the matrix was tentatively identified as a posterior tooth of †*Brachycarcharias atlasi* (C. Underwood, Birkbeck College, pers. comm. 2020), which occurs in Thanetian–Ypresian strata of the Ouled Abdoun phosphates (Arambourg, 1952). Secondly, the matrix includes poorly sorted peloids, and is thus lithologically consistent with Ypresian phosphates in the basin (Beds I and 0; Yans *et al*., 2014, Zouhri, S., 2017).

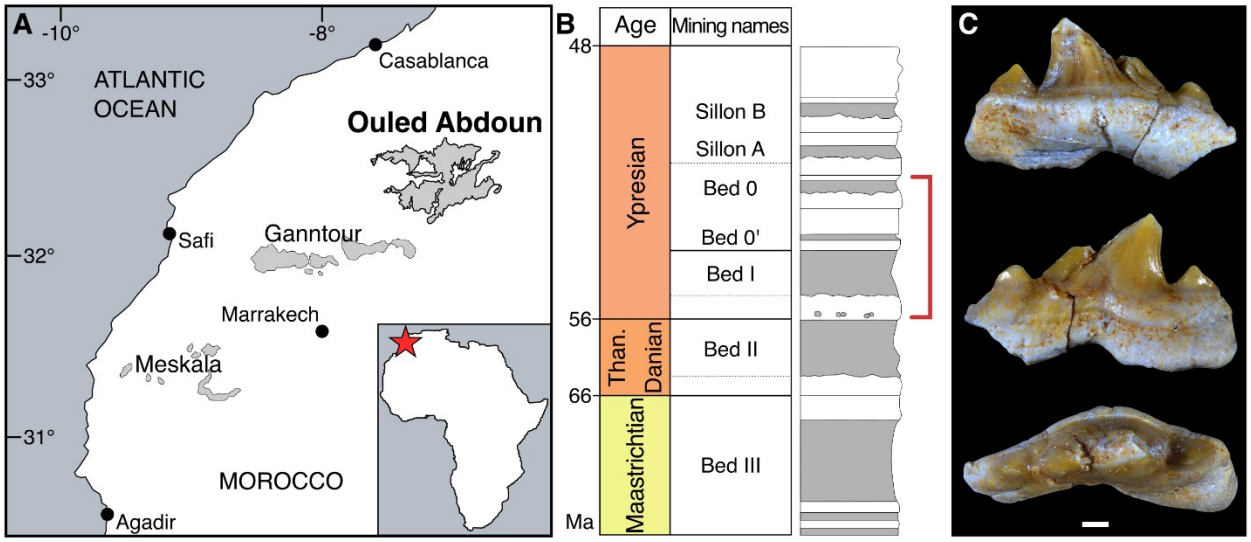

Fig. 4.1. Geographic and stratigraphic context for the holotype of †*Macroprosopon hiltoni* (UMMP 118216). A, geographic distribution of phosphorite deposits (in grey) in Morocco (modified from Yans *et al*., 2014); B, simplified stratigraphic chart of the Ouled Abdoun Basin deposits, with phosphorite sands in grey (modified from Yans *et al*., 2014). Red bracket indicates possible range of UMMP 118216, based on lithology and ichthyoliths associated with the specimen. C, posterior tooth of ?*Brachycarcharias atlasi* found embedded in the matrix of UMMP 118216, in labial, lingual and occlusal views (from top to bottom). Scale bar: 0.5 mm.

Diagnosis: Osteoglossiform with roughly triangular skull profile, relatively long jaws and terminal mouth; bulbous antorbital with strong ornamentation; two semicircular scleral rings;

approximately 23 maxillary teeth; >26 dentary teeth; bony collars at tooth base less than half the tooth height; lower jaw more than three times longer than deep; very long posterior process of the hyomandibula; opercle with dorsally-oriented concavity above the articular facet; 17 branchiostegals. †*Macroprosopon hiltoni* differs from †*Furichthys* in having the retroarticular included in (instead of excluded from) the articulation between lower jaw and quadrate; and the posterior process of the hyomandibula longer (rather than shorter) than the dorsal articulating surface of the hyomandibula. †*Macroprosopon hiltoni* differs from both †*Phareodus* and †*Brychaetus* in having proportionally much longer lower jaws; supraorbital shelf of the frontal not extending to the anterior margin of the frontal; posterior toothless portion of maxilla not substantially deeper than toothed portion; cleithrum extending anteriorly to the level of the angular (rather than extending to just below the preopercle).

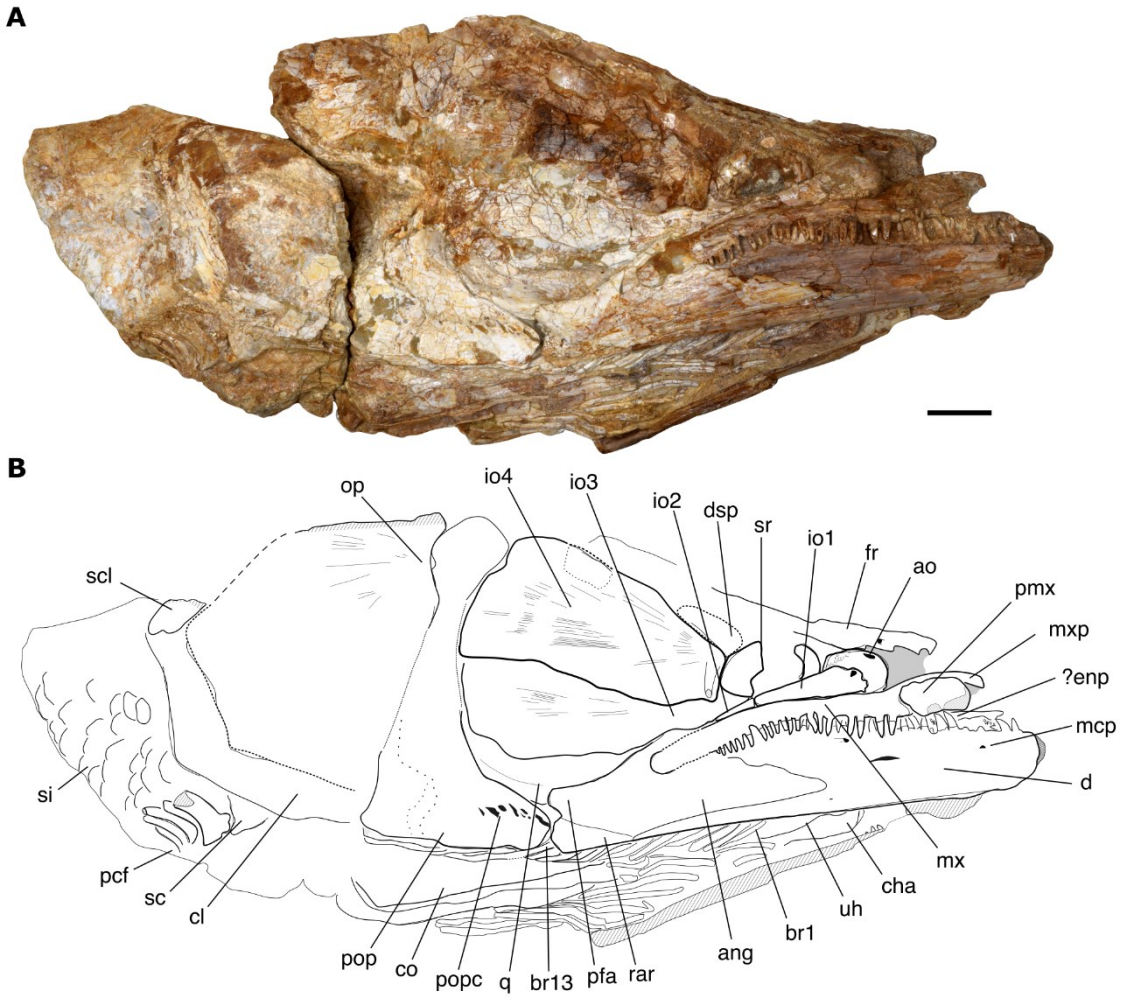

Fig. 4.2. †*Macroprosopon hiltoni* (holotype, UMMP 118216). Photograph (A) and interpretative line drawing (B) in right lateral view. Abbreviations: ang, angular; ao, antorbital; br1-br13, branchiostegal rays 1-13; cha, anterior ceratohyal; cl, cleithrum; co, coracoid; d, dentary; dsp, dermosphenotic; enp, endopterygoid; fr, frontal; io1-io4, infraorbitals 1-4; mcp, pore opening to the mandibular canal; mx, maxilla; mxp, anterior articular process of maxilla; op, opercle; pcf, pectoral fin; pfa, posterodorsal flange of the angular; pmx, premaxilla; pop, preopercle; popc, openings to the preopercular lateral line canal; q, quadrate; rar, retroarticular; sc, scapula; scl, supracleithrum; si, scale imprints; sr, scleral ring; uh, urohyal. Scale bar: 20 mm.

Description:

The holotype is broken transversely into two blocks that meet at the level of the opercle. It is strongly medio-laterally compressed; this compression caused the collapse of the skull roof on the left side of the specimen, which in turn resulted in several bones on the left side being crushed or completely missing. The anteriormost part of the skull is completely missing. Bones in the specimen have a widely varying state of preservation: some are heavily damaged and/or delaminated such that their surface is often missing (e.g., infraorbitals, opercular series, left angular); others are almost pristine (e.g., parts of the skull roof, branchiostegal rays).

Neurocranium. The anteriormost portion of the neurocranium (including nasals, vomer and part of the ethmoid region) is missing from the specimen. The frontal is very long, accounting for approximately two thirds of the skull roof length when excluding the nasals. It possesses a broad supraorbital shelf overlying the orbit and articulating antero-ventrally with the antorbital. The shelf bears a radial pattern of ornamentation on its dorsal surface, consisting of furrows and shallow pits. The anterior margin of the frontal is likely as broad as its posterior margin, as the supraorbital shelf does not seem to extend anteriorly to the articulation with the nasal. At the level of the orbit, the frontal is around 1.5 times broader than its posterior margin. The supraorbital canal ends posteriorly at about two thirds of the length of the frontal. The suture between the two frontals is not visible, because the right frontal partially overlaps the left one due to taphonomic distortion of the specimen. The suture between frontal and parietal is at least partially interdigitated. The parietal is short and bears a transverse crest dividing it in two portions: the anterior one is ornamented, while the posterior one is depressed with respect to the rest of the skull roof and forms part of a dorso-occipital fossa ("dépression dorso-occipitale" or "fosse dorso-occipitale" of Taverne, 1978). This fossa is bounded antero-ventrally by the parietal, medially by the supraoccipital, and postero-laterally by the epioccipital. The external

surface of the left parietal is partially broken, revealing a transverse canal-like structure that is likely the supratemporal commissure extending through the parietal. Although the temporal fossa is not exposed in the specimen and it is not possible to determine the bones that border it, the parietal clearly does not contribute to its margins. The epioccipital is a large bone forming the postero-lateral corner of the skull roof; it bears a strong ridge in continuity with the dorsal ridge of the pterotic that terminates posteriorly with a marked thickening. The antero-lateral margin of the epioccipital sutures with the pterotic. The supraoccipital bears a crest that is partially broken in the specimen, such that its full extent cannot be determined. A broken and flattened piece of tubular, canal-bearing bone overlying the medial part of epioccipital and dorso-occipital fossa is interpreted as the extrascapular. The sphenotic is relatively short and has a marked lateral projection (partially broken in the specimen) perpendicular to the antero-posterior axis of the skull. The pterotic is very long, overlies the sphenotic anteriorly and sutures with frontal, parietal and epioccipital medially. It bears a strong dorsal ridge on its posterior half. The lateral surface of the pterotic is smooth and lacks large pits or foramina. Ventro-medial to the sphenotic and pterotic, the prootic forms at least part of the articular surface for the anterior head of the hyomandibula. Ventral to pterotic and epioccipital, the intercalar bears a triangle-shaped posterior projection.

Orbital region. There is no identifiable supraorbital. The antorbital is bulbous and presents a heavily ornamented surface, with two different ornamentation fields: a postero-dorsal one with chevron-like patterns, and an antero-lateral one with radial furrows and shallow pits. The first infraorbital is slender and tapers posteriorly. It defines most of the ventral margin of the orbit and contributes partially to its anterior margin. The anterior portion of the first infraorbital is ventro-lateral to the antorbital. Posterior to the first infraorbital and lining the remaining portion of the ventral margin of the orbit there is a short and thin second infraorbital. The third and fourth infraorbitals are very large, covering most of the lateral postorbital area (‘cheek’) of the skull. The third and fourth infraorbitals are at least twice as long as they are deep. Their surface is ornamented with thin radial ridges. The fourth infraorbital is deeper than the third one and partially overlaps it. The infraorbital sensory canal is completely enclosed in a bony canal that extends through all the infraorbitals. Although the dermosphenotic is absent in the specimen, we interpret an elongated and roughly triangular surface postero-dorsal to the orbit as an impression

left by that bone. Two semicircular ossified scleral rings (anterior and posterior) surround the eye.

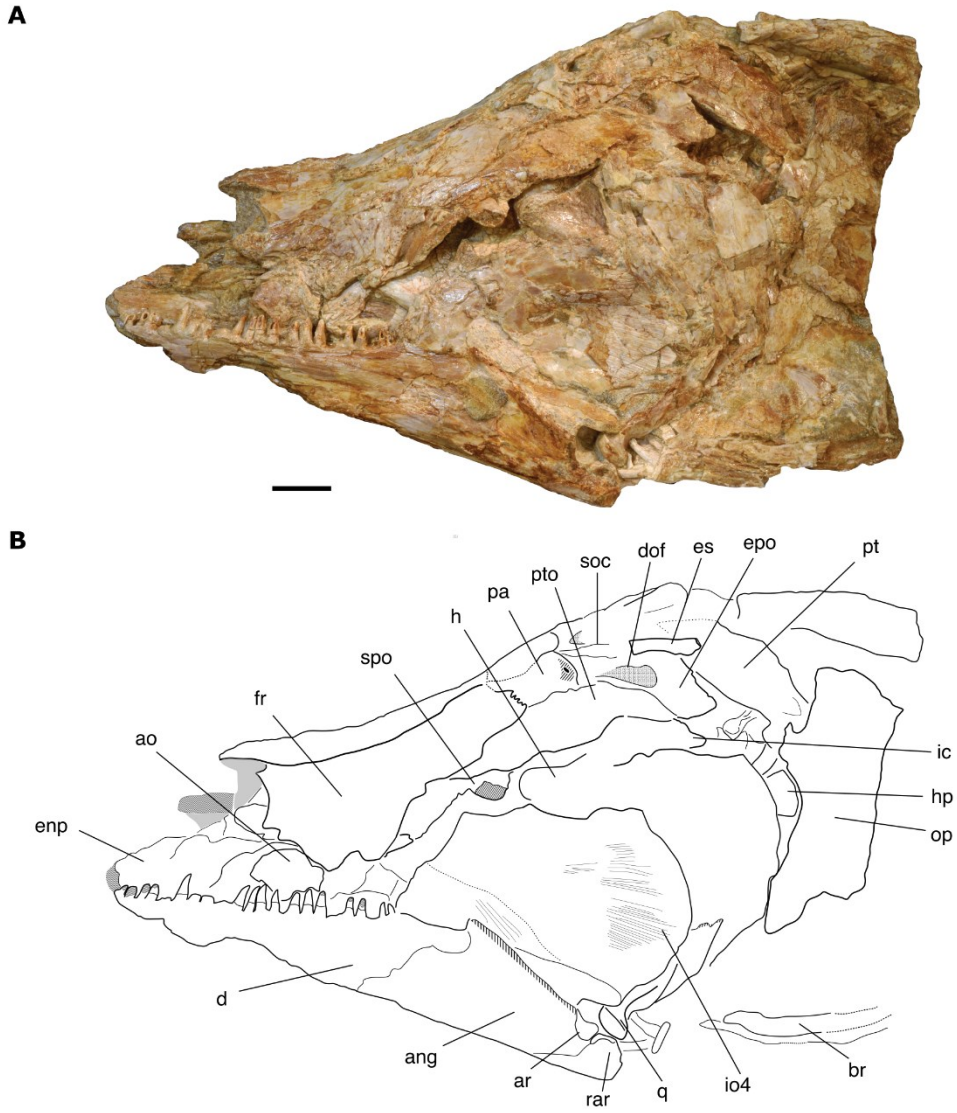

Fig. 4.3. †*Macroprosopon hiltoni* (holotype, UMMP 118216). Photograph (A) and interpretative line drawing (B) in left lateral view. Abbreviations: ang, angular; ao, antorbital; d, dentary; dof, dorso-occipital fossa; enp, endopterygoid; epo, epioccipital; es, extrascapular; fr, frontal; h, hyomandibula; hp, posterior (=opercular) process of the hyomandibula; ic, intercalar; io4, infraorbital 4; op, opercle; pa, parietal; pt, posttemporal; pto, pterotic; q, quadrate; soc, supraoccipital; spo, sphenotic; rar, retroarticular. Scale bar: 20 mm.

Jaws. The premaxillae are not preserved in the specimen, except for a broken splinter of bone antero-lateral to the right maxilla. The maxilla is long and slightly curved with ventral concavity.

It tapers anteriorly into an elongated, narrow, and arched anteromedial process. This process, which would articulate with the premaxilla, is missing its anterior tip. The length of the process—coupled with the length and proportions of the lower jaw—suggests that the premaxilla was a relatively long bone, especially when compared with other osteoglossids. There is no distinct dorsal swelling in the maxilla behind the anteromedial process. There are 19 maxillary teeth arranged in a single row that are visible on the right maxilla, with a complete maxillary set consisting of approximately 23 teeth when accounting for empty spaces left by tooth replacement. The teeth decrease in size from the anterior to the posterior portion of the maxilla. They are hollow, sub-conical in shape, with a short (less than a third of the tooth height) bony collar at the base and a small conical acrodin cap at the tip. The posterior, toothless portion of the maxilla, which overlies the angular, is rounded and not substantially deeper than the rest of the bone. Based on tomograms, broken pieces of bone dorsal to the posterior portion of the maxilla likely belong to the third infraorbital and potentially the maxilla itself, rather than to a supramaxilla. The lower jaws are incomplete, missing their anteriormost portions. They are straight and elongated, with a low coronoid process and a relatively long post-coronoid region. The dentary is very lightly ornamented with parallel lines running along its length. Two large pores of the mandibular canal are visible on the external surface of the dentary. A complete dentary would include more than 26 teeth. Dentary teeth are larger on average than the maxillary ones, and they are markedly compressed antero-posteriorly. The relative size of the bony collar at the tooth base varies from a third to half the tooth height. Several replacement tooth crowns are visible on the right lower jaw. The angular is very long and extends anteriorly at the antorbital level. It presents a large postero-dorsal flange that laterally covers the quadrate articular condyle and the articular surface of the lower jaw. Articular and retroarticular are not fused with the angular. They both contribute to the surface of the jaw joint.

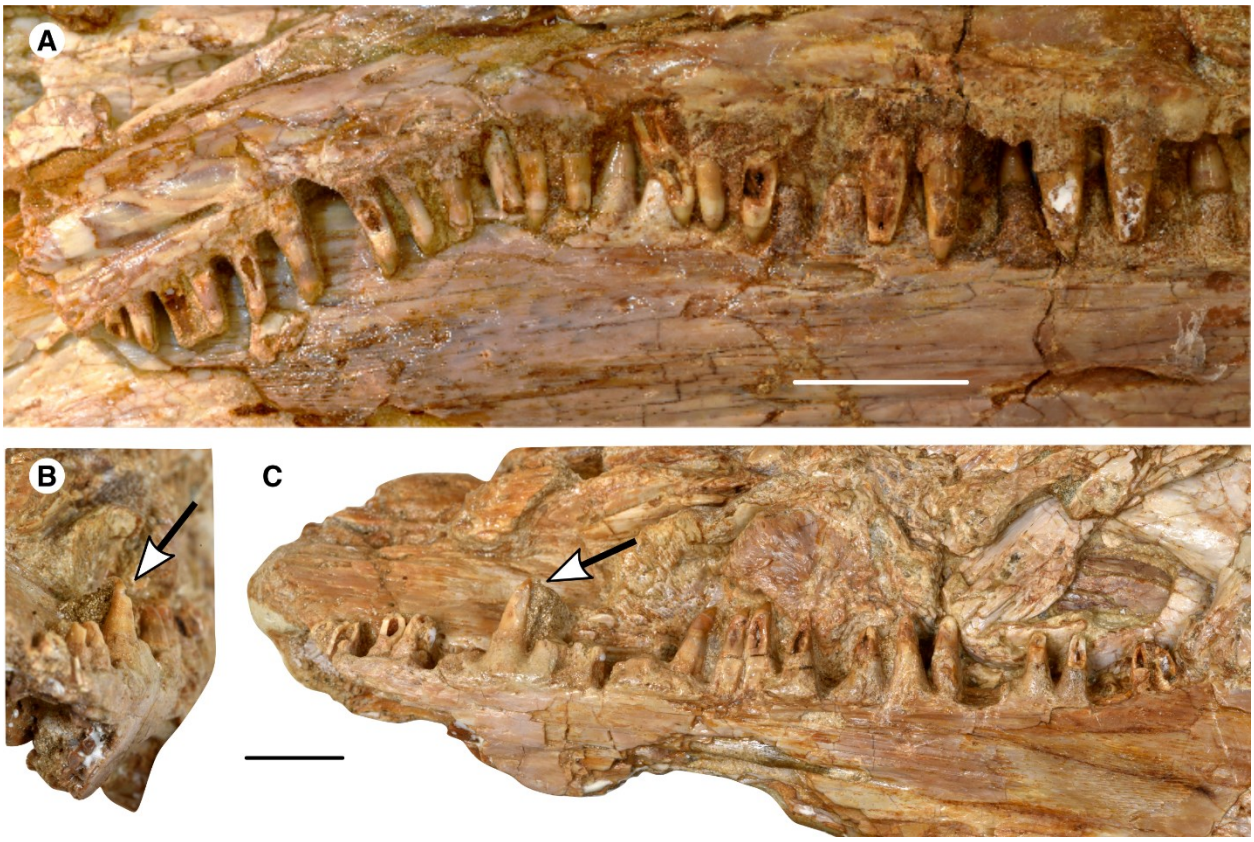

Fig. 4.4. †*Macroprosopon hiltoni* (holotype, UMMP 118216), dentition. A, close-up of right maxillary teeth in lateral view; B, C, left dentary teeth in anterior (B) and lateral (C) views. Arrows point to the same tooth in different views. Scale bars: 10 mm.

Palate and suspensorium. Only a small part of the palate can be seen in the specimen. Anteriorly, an exposed plate-like bone with a multitude of small teeth is interpreted as the palatine-ectopterygoid or the endopterygoid. The quadrate is approximately triangular in lateral view and likely longer than deep. The 'peg-like' head of the quadrate articulates with the articular and retroarticular of the lower jaw. Posterodorsal to its head, a strong ridge marks a portion of the posterior edge of the quadrate. The symplectic cannot be easily identified on the specimen. The hyomandibula is mostly covered by the third and fourth infraorbital on both sides of the specimen. The anterior hyomandibular head is clearly distinct from the posterior one on the partially exposed left hyomandibula. The articular surface with the opercle can also be seen on the left side of the specimen. Its posterior position, distant from the heads articulating with the braincase, suggest a very long posterior (=opercular) process. This is confirmed by examination

of the tomograms, which show the posterior process being slightly longer than the dorsal articulating surface of the hyomandibula.

Opercular series. The ventral half of the right preopercle is well preserved. It presents a curved anterior margin, with an angle larger than 90º between its vertical arm and a very short horizontal arm that does not anteriorly reach the level of the orbit. The preopercular sensory canal opens in the horizontal arm through 6 large, antero-ventrally directed pores, arranged in a straight horizontal line. It is unclear whether these were originally covered by a thin lamina of bone that might have broken off post-mortem. The opercle is incompletely preserved and fractured in several pieces on both sides. Its anterior margin has a distinct dorsally-oriented concavity just above the articular facet for the hyomandibular process. The dorsal margin of the opercle is almost flattened, with only a moderate amount of curvature. Based on its imprint on the right side of the specimen, the ventral margin of the opercle was likely straight or very slightly curved.

Branchial skeleton. The left anterior ceratohyal and the urohyal are the only exposed bones of the ventral hyoid arch. The anterior ceratohyal has a broad anterior head. The urohyal has a distinct head and a narrow ventral margin. An isolated tooth crown embedded in the sediment anterior to the ceratohyal might be part of the basibranchial toothplate dentition. There are 17 branchiostegals, with the posterior ones notably deeper than the anterior ones.

Pectoral girdle and fin. Part of the posttemporal can be seen on the left side of the specimen. It bears a broad and flattened dorsal arm that articulates with the back of the neurocranium. The ventral arm of the posttemporal can be identified in the tomograms; it is laterally compressed and relatively short, reaching around half the length of the dorsal arm. The scapula and part of the cleithrum are exposed on the right side of the specimen. However, anatomical details of these bones cannot be discerned. The coracoid extends anteriorly beyond the lower jaw joint, as inferred by its exposed ventral margin. Six pectoral rays are partially preserved. The first one is greatly enlarged and thickened.

Vertebral column. The first fourteen vertebrae (or at least their centra) are preserved in the specimen, as evaluated from the tomograms. All the vertebrae are amphicoelous and much deeper than long. Examination of the tomograms reveals the presence of a paired autogenous structure ventral to the first vertebra that extends anteriorly few centimeters below the occipital

region of the neurocranium. We interpret this structure as a greatly expanded first parapophysis that is wedge-shaped in lateral view (see 'Modified coding and scoring of characters' on Character 88 for further discussion of this feature). The second and third vertebrae are partially exposed on the left side of the specimen, and clearly illustrate the autogenous nature of the neural arches.

Scales. Scales are poorly preserved in the specimen, with small scale fragments including their surface texture found in the anterior block, and whole scales (often fractured and delaminated) in the right side of the posterior block, posterior to the pectoral girdle. Scales appear to be subcycloid, few centimeters in diameter and partially overlapping. They seem to lack reticulate furrows; instead, small tubercles ornament their surface.

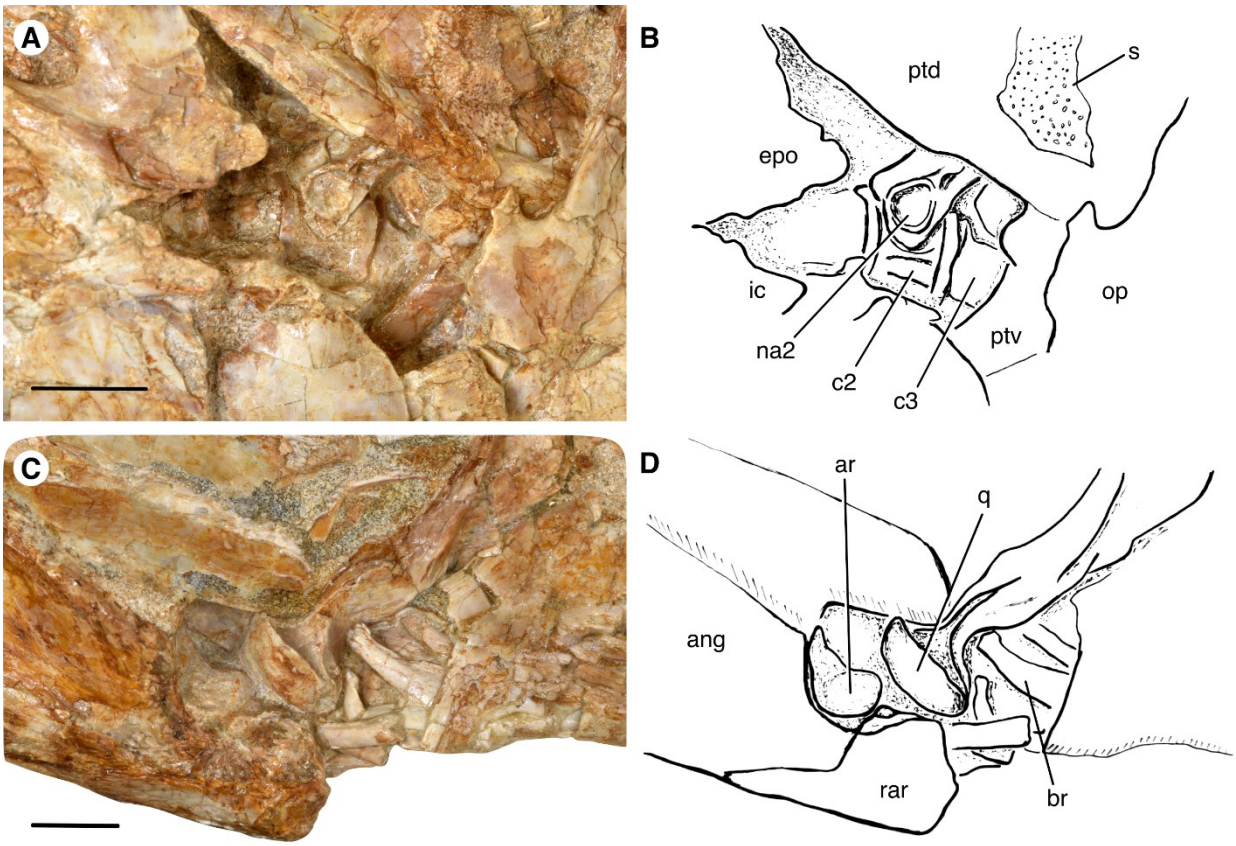

Fig. 4.5. †*Macroprosopon hiltoni* (holotype, UMMP 118216). A, B, close-up of exposed second and third vertebrae in left lateral view, with interpretative line drawing (B); C, D, close-up of the articulation between quadrate and lower jaw in left lateral view, with interpretative line drawing (D). ang, angular; ar, articular; br, branchiostegal ray; c2, second vertebral centrum; c3, third vertebral centrum; epo, epioccipital; ic, intercalar; na2, neural arch of the second vertebra; op, opercle; ptd, dorsal arm of the posttemporal; ptv, ventral arm of the posttemporal; q, quadrate; rar, retroarticular; s, scale fragment. Scale bars: 10 mm.

RESULTS

Modified coding and scoring of characters.

Character (2): Shape of extrascapular. This character has been scored for †*Sinoglossus* as '0' (expanded) by Wilson & Murray (2008), who adapted the scoring by Li *et al*. (1997b) to the coding of this character by Hilton (2003). However, many of the taxa scored by Li *et al*. (1997b) as possessing an expanded extrascapular have a reduced extrascapular according to the coding of Hilton (2003; see Hilton 2003:30 for an in-depth discussion of this character). Although we were not able to examine any specimen of †*Sinoglossus* first-hand, the original description of this taxon (Su, 1986) does not suggest the presence of an extrascapular expanded in a similar way to the extrascapulars of hiodontids and mormyrids. Given the uncertainty on the state of this character for this taxon, we changed the scoring of †*Sinoglossus* from '0' (expanded) to '?'.

Character (8): Parasphenoid teeth. The ventral surface of the parasphenoid of †*Phareodus* is almost devoid of teeth, with the exception of one large conical tooth (sometimes joined by two smaller ones) between the basipterygoid processes (Li *et al*., 1997a). We changed the scoring of †*Phareodus* for this character from '?' to '3' (large and restricted to the basal portion of the parasphenoid).

Character (9): Basipterygoid process. Hilton (2003) coded this character with two possible scores, absent (state 0) and present (state 1). However, the scores in Hilton's (2003) character matrix are inverted, with taxa possessing a basipterygoid process (such as *Osteoglossum* and *Pantodon*) scored as '0' and taxa lacking a basipterygoid process (such as *Hiodon* and *Chitala*) scored as '1'. This mis-scoring has been repeated in all successive versions of the character matrix. We maintain the original definition and coding of the character, but we fixed the scoring such that '0' indicates absence and '1' indicates presence of the basipterygoid process.

Character (13): Basisphenoid. The basisphenoid has been identified in †*Phareodus* by Li *et al.* (1997a) as being one of the cartilaginous bones forming the dorso-medial wall of the orbit, located ventromedial to the orbitosphenoid. However, it is difficult to establish whether the ventral portion of the orbital wall in †*Phareodus* (as seen, for example, in FMNH PF 10237, FMNH PF 10285, FMNH PF 16536) represents a basisphenoid or rather a medial vertical lamina

of the parasphenoid. Because of the uncertainty in interpreting this feature in †*Phareodus*, we changed the scoring for this character from '0' (present) to '?'.

Character (20): Supraorbital bone. Several non-osteoglossomorph teleosts possess a supraorbital bone anterodorsal to the orbit. Among these, *Elops* (Forey, 1973) and *Dorosoma* (Grande, 1985) are included in the character matrix and were previously scored as lacking the supraorbital (state 1) by Murray *et al*. (2018). Thus, we changed the scoring for *Elops* and *Dorosoma* from '1' (absent) to '0' (present). Additionally, †Ellimmichthyiformes was also scored as lacking the supraorbital (state 1) by Murray *et al*. (2018). Because the supraorbital is present in several †ellimmichthyiforms and secondarily lost in the sub-clade †Paraclupeinae (Murray & Wilson, 2013), we changed the scoring for †Ellimmichthyiformes to '0' (present), reflecting the likely ancestral state of this character within this clade.

Character (22): Number of bones in the infraorbital series, not including the dermosphenotic or the antorbital if present. We changed the scoring for *Pantodon*, which is unique among osteoglossomorphs in having five (instead of four) infraorbitals (Hilton, 2003), from '1' (four) to '0' (five). The previous scoring was likely an accidental error in the Wilson & Murray (2008) matrix, which carried on to the Murray *et al*. (2018) matrix.

Character (23): First infraorbital. This character, as defined by Hilton (2003), distinguishes a condition in which the first infraorbital does not contribute or only partially contributes to the anterior margin of the orbit (state 0) from a condition in which the first infraorbital is the only bone that contributes to the anterior margin of the orbit (state 1). Based on this definition, we changed the scoring of this character for *Dorosoma* from '1' to '0' (Grande, 1985); for †*Xixiaichthys* from '1' to '0' (Zhang, 2004); for †*Joffrichthys tanyourus* from '?' to '0' (Murray *et al*., 2018); and for †*Paralycoptera* from '?' to '0' (Xu & Chang, 2009). It should be noted that redefining this character by including more states that distinguish between a condition in which the first infraorbital does not contribute at all to the anterior margin of the orbit and another one in which the first infraorbital contributes to the ventral portion of the anterior margin of the orbit might better capture the range of morphologies and topological relationships observed for the first infraorbital of osteoglossomorphs.

Character (26): Dermosphenotic. Li & Wilson (1996) defined this character to distinguish the triradiate condition found exclusively in Hiodontiformes from other osteoglossomorphs, which

were assigned the plesiomorphic state (defined as 'irregularly triangular'). Hilton (2003) added a third state (tubular) to describe the condition seen in some notopterids and mormyrids, and changed the definition of the plesiomorphic state to simply 'triangular'. Several taxa scored as having a 'triangular' dermosphenotic have a quadrangular or irregularly shaped dermosphenotic (e.g., *Heterotis*, *Notopterus*, †*Lycoptera*; Hilton, 2003). To avoid future ambiguities in scoring and highlight the distinction from a tubular or triradiate state, we changed the definition of the plesiomorphic state (state 0) to 'flattened, plate-like.'

Character (28): Neurocranial heads of the hyomandibula. †*Phareodus* has been previously described as having one continuous hyomandibular head, corresponding to state 0 (one head or two heads but continuous) of this character (Wilson & Murray, 2008). However, we observed two clearly distinct hyomandibular heads in †*Phareodus encaustus* (e.g., FMNH PF 10237, FMNH PF 10285) and †*Phareodus testis* (e.g., FMNH PF 11942, FMNH PF 17493). We changed the scoring for †*Phareodus* from '0' to '1' (two heads, separate). We also changed the scoring for †*Wilsonichthys* from '1' (two heads, separate) to '?', because Murray *et al*. (2016: 7) report that the hyomandibula "[...] has two articular heads, but the bone is not well preserved, and whether or not the heads might have had a bony connection cannot be determined". As the condition of having two hyomandibular heads bridged by a bony connection would correspond to state 2 of this character, it is more conservative to score †*Wilsonichthys* as uncertain for this character.

Character (31): Autopalatine bone. A bony autopalatine is absent in osteoglossomorphs, with the only reported exceptions in *Heterotis* and possibly *Scleropages leichardti* (Arratia & Schultze, 1991; Hilton, 2003). Reexamination of †*Phareodus encaustus* (e.g., FMNH PF 10237, FMNH PF 16529) reveals the presence of a bony autopalatine in this species as well. We changed the scoring of this character for †*Phareodus* from '?' to '0' (present). Some non-osteoglossid osteoglossomorphs (†*Joffrichthys tanyourus*, †*Shuleichthys* and †*Xixiaichthys*) have been scored as '0' (present) in previous versions of the data matrix (Wilson & Murray, 2008; Murray *et al*., 2010, 2018). Since there is no mention of an autopalatine in the descriptions of those taxa (Zhang, 2004; Murray *et al*., 2010, 2018) and an autopalatine cannot be identified in specimen photographs and interpretative drawings, we changed the scoring of †*Joffrichthys tanyourus*, †*Shuleichthys* and †*Xixiaichthys* for this character from '0' (present) to '?'.

Character (40): Supramaxillae. †*Joffrichthys tanyourus* is scored as state 0 (present) in the Murray *et al*. (2018) character matrix. However, the description clearly states that there are no supramaxillae in this taxon (Murray *et al*., 2018). Hence, we changed the scoring for †*Joffrichthys tanyourus* from '0' to '1' (absent).

Character (42): Posterior bones of the lower jaw. †*Lopadichthys* is scored as state 2 (all separate) in the Murray *et al*. (2018) character matrix. However, the description and figures clearly indicate the angular and the articular as indistinguishably fused with each other (Murray *et al*., 2018). Hence, we changed the scoring for †*Lopadichthys* to '1' (angular and articular fused).

Character (75): Intestine. One of the few synapomorphies shared by all extant osteoglossomorphs is having an intestine that passes to the left of the stomach, instead of passing to the right like in the vast majority of ray-finned fishes (Nelson, 1972). Although this character is particularly difficult -- if not impossible -- to evaluate in fossil taxa, due to the low preservation potential of soft tissues like the gastrointestinal tract, it is nonetheless valuable in supporting the monophyly of Osteoglossomorpha on the basis of morphological characters. We changed the scoring of several taxa that were listed as uncertain ('?') according to whether their intestine coils to the right of the stomach ('0') or to the left of the stomach ('1'), based on relevant literature (Nelson, 1972; Banan Khojasteh, 2012): *Amia* ('?' ⟶ '0'), *Elops* ('?' ⟶ '0'), *Gnathonemus* ('?' ⟶ '1'), *Chitala* ('?' ⟶ '1'), *Xenomystus* ('?' ⟶ '1'), and *Papyrocranus* ('?' ⟶ '1').

Character (78): Second infraorbital shape and size. †*Joffrichthys tanyourus* was scored as state 1 (triangular or rectangular and smaller than third infraorbital) in the Murray *et al*. (2018) character matrix. However, in this taxon the second infraorbital is probably fused to the third (Murray *et al*., 2018). Thus, we consider this character to be not applicable in †*J. tanyourus.* The second infraorbital of †*Sinoglossus* is approximately rectangular in shape and relatively deep, yet substantially smaller than the third infraorbital (Su, 1986; Li & Wilson, 1996). Accordingly, we changed the scoring of †*Sinoglossus* for this character from '2' (expanded and equivalent in size to or larger than third infraorbital) to '1' (triangular or rectangular and smaller than third infraorbital). *Pantodon* has five infraorbitals instead of the usual condition of four infraorbitals seen in extant osteoglossomorphs, complicating the assessment of their homology. However, the neuromast pattern of the infraorbitals of *Pantodon* suggests that, in this taxon, the first two

infraorbital bones correspond to the first infraorbital of other osteoglossomorphs (Nelson, 1969; Hilton, 2003). Thus, the third infraorbital of *Pantodon*, which is substantially deeper than the first two and smaller than the fourth, is homologous to the second infraorbital of other bonytongues. Following this identification, we changed the scoring of *Pantodon* for this character from '0' (more or less slender or tubular and small in size) to '1' (triangular or rectangular and smaller than third infraorbital).

Character (86): Anal fin sexual dimorphism. We changed the scoring for *Pantodon*, which is characterized by extreme sexual dimorphism in the anal fin (Lastein & Van Deurs, 1973), from '0' (absent) to '1' (present).

Character (88): Parapophysis on the first centrum. The parapophysis of the first vertebral centrum is expanded or hypertrophied in several osteoglossids, as first noted by Forey & Hilton (2010) in *Arapaima*, *Osteoglossum*, and †*Phareodus*. Murray *et al*. (2018) included this information in their character matrix by adding a two-state character where one state indicates a non-expanded or hypertrophied first parapophysis, and the other state indicates an expanded or hypertrophied first parapophysis that reaches under the occiput. However, the enlarged first parapophysis in osteoglossids can exist in two very different conditions. In *Osteoglossum* and *Scleropages* (and maybe †*Singida*; Murray *et al*., 2018), the first parapophysis is rounded in lateral and ventral views, it touches the basioccipital but does not contact the parasphenoid. In *Arapaima* and several fossil taxa (including †*Phareodus*, †*Brychaetus* and †*Macroprosopon*), the first parapophysis is greatly hypertrophied, it appears wedge-shaped in lateral view, reaches anteriorly below the occipital region of the neurocranium and contacts (or even interdigitates with) the parasphenoid. Thus, we changed the coding of this character to encompass its observed variability among bonytongues: parapophysis on the first centrum not expanded or hypertrophied (state 0); expanded and rounded, barely reaching below the occiput and not touching the parasphenoid (state 1); greatly hypertrophied and extending anteriorly to touch the parasphenoid, wedge-shaped in lateral view (state 2).

This character is difficult to score for *Heterotis*, because in this taxon the first vertebra is completely fused to the occipital region of the neurocranium (Taverne, 1977; Forey & Hilton, 2010). Ontogenetic studies suggest that *Heterotis* has completely lost the parapophysis on the first centrum, as it cannot be identified in young specimens where the first vertebra is not yet

fused to the basioccipital (Britz & Johnson, 2010). Hence, we considered the scoring of character 88 to be not applicable for *Heterotis*.

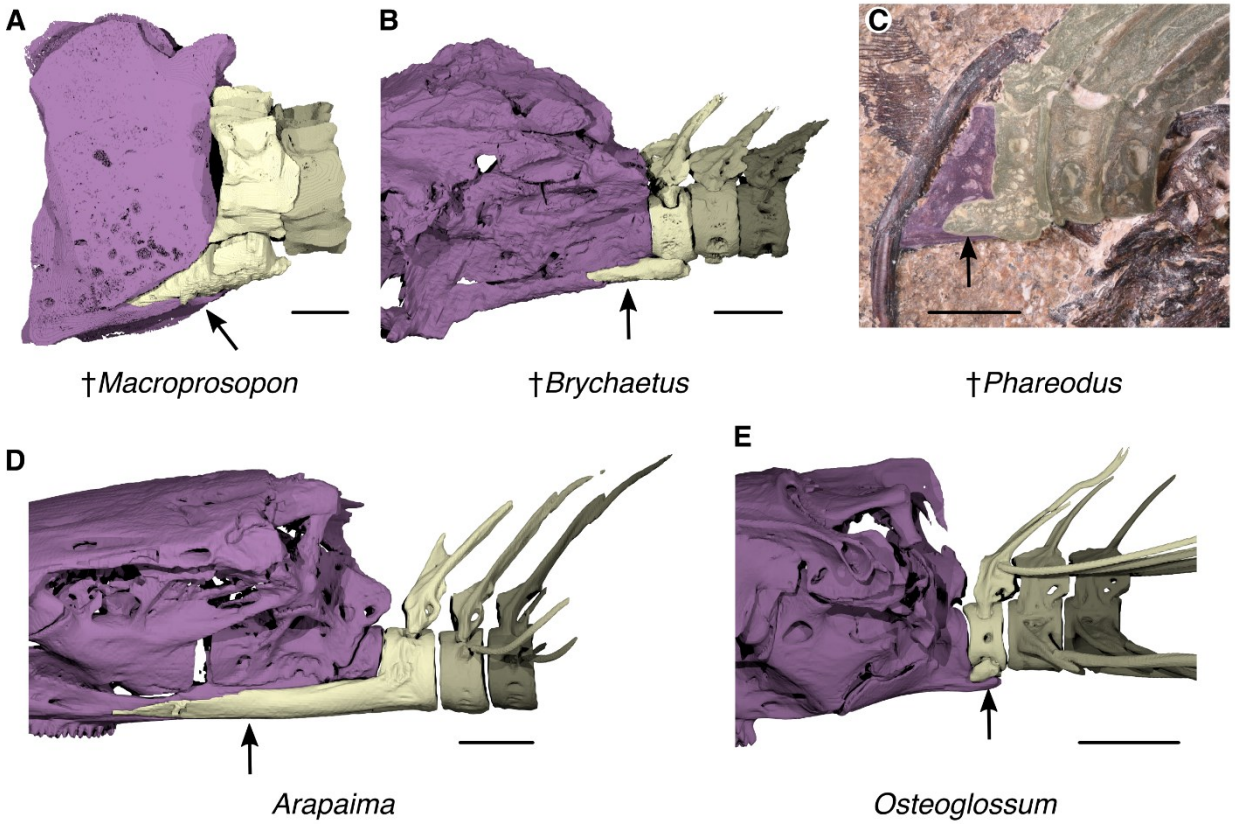

Fig. 4.6. Left lateral view of the occipital region of the braincase and the first few vertebrae in Osteoglossidae. Arrows point to the enlarged parapophysis ventral to the first vertebral centrum (character 88). Notice that in *Arapaima* and in the fossil taxa shown here, the first parapophysis is greatly hypertrophied and extends anteriorly below the braincase (state 2). A, †*Macroprosopon hiltoni* (UMMP 118216; braincase model approximate due to insufficient contrast in tomography data); B, †*Brychaetus muelleri* (NHMUK PV P641); C, †*Phareodus encaustus* (FMNH PF 10257); D, *Arapaima gigas* (UF 33107); E, *Osteoglossum bicirrhosum* (UF 189007). Colours indicate different anatomical regions: neurocranium (purple); first, second and third vertebrae (shades of beige). Scale bars: 10 mm (A, B, C, D); 5 mm (E).

Rescoring of †Paralycoptera and exclusion of †Tanolepis

We rescored 33 characters for the Early Cretaceous †*Paralycoptera changi* based on the redescription of this genus by Xu & Chang (2009). Most of these changes replace previously missing data. The following list indicates the number of the updated character, the state of that character in the Murray *et al*. (2018) matrix and the new state scored in this study: (2) ?⟶0; (5)

?⟶0&1; (6) ?⟶2; (7) ?⟶3; (8) ?⟶2; (12) ?⟶0; (20) ?⟶1; (21) ?⟶0; (24) ?⟶0; (25) ?⟶1; (29) ?⟶0; (32) 0⟶2; (33) 0⟶1; (35) 1⟶0; (36) ?⟶1; (38) ?⟶0; (42) ?⟶1; (43) ?⟶1; (44) ?⟶1; (47) 0⟶?; (51) ?⟶1; (55) ?⟶0; (56) ?⟶0; (60) ?⟶0; (61) ?⟶1; (63) ?⟶0; (64) ?⟶0; (67) 0⟶1; (68) ?⟶2; (78) ?⟶0; (82) 1⟶0; (83) ?⟶0; (87) ?⟶1.

†*Tanolepis ninjagouensis* from the Late Jurassic–?Early Cretaceous Fenshuiling Formation has been included in several phylogenetic analyses of Osteoglossomorpha alongside †*Paralycoptera*, always falling as sister taxa to each other (Li *et al*., 1997*b*; Wilson & Murray, 2008; Murray *et al*., 2018). †*Tanolepis* has been synonymized with †*Paralycoptera* by Jin *et al*. (1995). This decision is supported by Xu & Chang (2009) in their redescription of †*Paralycoptera*, whereas Li *et al*. (1997*b*) listed a few characters differentiating the two taxa and rejected the synonymization. Because we were not able to visually examine any specimens of †*Tanolepis* and check in this taxon the state of the characters we rescored for †*Paralycoptera*, we decided to exclude †*Tanolepis* from our character matrix and phylogenetic analysis.

Newly defined characters.

Character (89): Dorso-occipital fossa: absent [0]; present [1]. The dorso-occipital fossa, a large paired depression on the posterior portion of the skull roof bordered by parietal, supraoccipital, and epioccipital, was proposed by Taverne (1979) as a diagnostic characteristic of †Phareodontinae. Among the taxa included in this study, only †*Phareodus*, †*Brychaetus*, and †*Macroprosopon* present a dorso-occipital fossa. The presence or absence of this depression cannot be determined from available specimens of †*Furichthys*.

Character (90): Contact between dermosphenotic and anteriormost bone of the infraorbital series: absent [0]; present [1]. A contact between the dermosphenotic and the anteriormost bone of the infraorbital series (either the antorbital or the first infraorbital, in cases where the antorbital is absent) is seen only in *Osteoglossum*, *Scleropages*, and *Arapaima* among osteoglossomorphs. It should be noted that this character is partially correlated to some extent with character 4 (supraorbital shelf of frontal bone), as taxa that have a supraorbital shelf of the frontal will likely lack a contact between the dermosphenotic and the antorbital (or first infraorbital). However,

several taxa in which the dermosphenotic does not contact the antorbital (or first infraorbital) lack a supraorbital shelf of the frontal.

Character (91): Depth of dorsal posterior infraorbital compared to ventral posterior infraorbital: shallower [0]; equal [1]; deeper [2]. The two most posterior infraorbitals in osteoglossomorphs are usually identified (*sensu* Hilton, 2003) as infraorbitals 3 (ventral) and 4 (dorsal). Exceptions are found in *Pantodon* (infraorbitals 4 and 5, as there is one more element in the infraorbital series), some species of †*Lycoptera* (e.g., †*L. middendorffi* and †*L. davidi*, with three posterior elements of the infraorbital series identified as infraorbitals 3, 4 and 5; Greenwood, 1970; Ma, 1987), *Gymnarchus* (with more than 10 small tubular elements in the infraorbital series; Taverne, 1972), and potentially †*Phareodus testis* and †*Brychaetus*, in which there are apparently only three elements of the infraorbital series (excluding antorbital and dermosphenotic; Roellig, 1974; Li *et al.*, 1997a). This character captures the relative depth proportions of the two posteriormost infraorbitals. †*Lycoptera* was scored as '?', due to the difficulty in defining the identity and homology of the three posteriormost elements of its infraorbital series.

Character (92): Scleral ossicles: absent [0]; present [1]. Scleral ossicles are supportive bony structures found in the eyes of some teleost fishes. The phylogenetic distribution of scleral ossicles within teleosts is complex, with multiple clades losing or gaining scleral ossicles independently (Franz-Odendaal, 2008; 2020). However, presence or absence of scleral ossicles tends to be conserved within family-level taxa (Mok & Liu, 2012). Among extant osteoglossomorphs, two scleral ossicles forming a thin ring around the eye are only found in large specimens of *Hiodon* (Hilton, 2002; contrary to Taverne, 1977). Scleral ossicles seem to be absent in the majority of fossil osteoglossomorphs. However, at least two extinct taxa (†*Brychaetus* and †*Macroprosopon*) have a robust, well-developed scleral ring. Remarkably, the scleral ring of †*Brychaetus* appears to be made of one single circular ossicle (Casier, 1966; Roellig, 1974), whereas the scleral ring of †*Macroprosopon* is made of two ossicles (anterior and posterior). The variability in the number of ossicles making up the scleral ring could be included in future phylogenetic analyses, especially if this feature is discovered in more fossil taxa.

Character (93): Postero-dorsal flange of the angular: absent [0]; present [1]. This character captures the presence (or absence) of a raised flange in the postero-dorsal portion of the angular (or angulo-articular, or angulo-retroarticular, when this bone is fused to other bones of the lower

jaw), which covers the articular surface of the quadrate in lateral view. Most osteoglossomorphs do not present this anatomical feature, and in these taxa the articular condyle of the quadrate can be clearly seen in lateral view (when it is not covered by infraorbitals or by the preopercular). A postero-dorsal flange of the angular covering the articular condyle of the quadrate is present in most osteoglossids (*Arapaima*, *Heterotis*, *Scleropages*, †*Macroprosopon*, †*Phareodus* and †*Brychaetus*), but not in *Osteoglossum*.

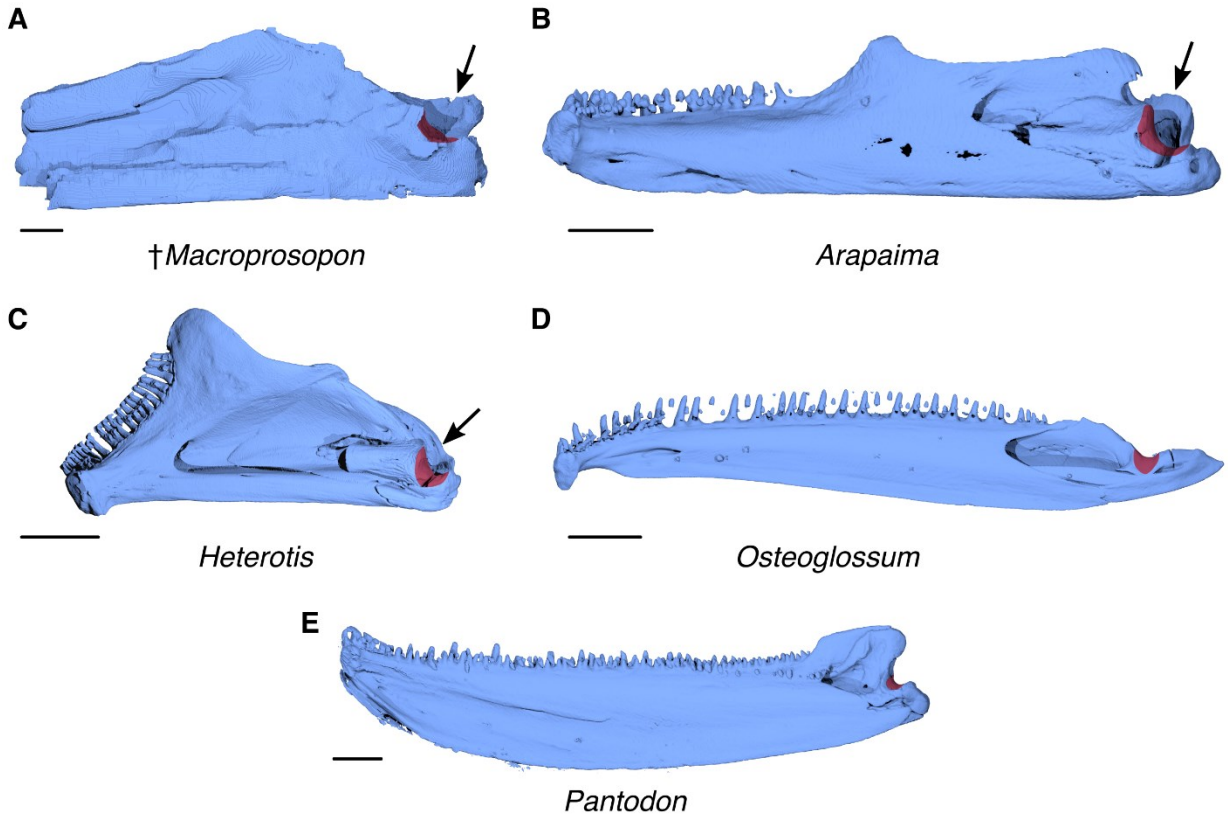

Fig. 4.7. Right lower jaws of osteoglossomorphs in medial view (digital renderings). The red shade indicates the articular surface of the quadrate-lower jaw articulation. The arrows point to the postero-dorsal flange of the angular, when present (character 93, state 1). A, †*Macroprosopon hiltoni* (UMMP 118216; partial lower jaw); B, *Arapaima gigas* (UF 33107); C, *Heterotis niloticus* (UMMZ 195004); D, *Osteoglossum bicirrhosum* (UF 189007); E, *Pantodon buchholzi* (UMMZ 249782). Scale bars: 10 mm (A, B); 5 mm (C, D); 1 mm (E).

Character (94): Posterior process of the hyomandibula: short (less than half the length of the dorsal articulating surface of the hyomandibula) [0]; long (more than half the length of the dorsal articulating surface of the hyomandibula) [1]; absent or extremely reduced [2]. The posterior [=

opercular] process of the hyomandibula is short in extant and fossil holosteans (Grande & Bemis, 1998; Grande, 2010), stem teleosts such as †pholidophorids (Arratia, 2013) and †ichthyodectiforms (Cavin *et al*., 2013), and several crown teleost clades (e.g. elopomorphs, clupeomorphs, osmeriforms, galaxiids, salmonids; Forey, 1973; Grande, 1985; Sanford, 2000; McDowall & Burridge, 2011), suggesting that a short posterior process is the ancestral state for crown teleosts and for osteoglossomorphs. Within Osteoglossomorpha, there is great variability in the relative length of the posterior process of the hyomandibula. In †*Lycoptera*, the posterior process is more strongly developed than in most teleosts, but it reaches at most half the length of the dorsal articulating surface of the hyomandibula (Greenwood, 1970; Jin *et al*., 1995; Zhang, 2002). Hiodontids are characterized by a long and deep opercular process, with dorsomedial and ventrolateral flanges (Hilton, 2002; Hilton & Grande, 2008). Mormyroids present an extremely modified hyomandibula that lacks a distinct posterior process and articulates with the opercle through a deep condyle (Taverne, 1972). The opercular process varies considerably within notopterids, with *Notopterus*, *Chitala*, and *Xenomystus* having strongly developed but relatively short processes, and *Papyrocranus* showing a very long posterior process connected with the dorsal articulating surface of the hyomandibula by a bony wing (Taverne, 1978). Most of the non-osteoglossid fossil taxa examined here, as well as the extant *Pantodon*, display the plesiomorphic condition of having a short posterior process of the hyomandibula. An exception to this is represented by †*Paralycoptera*, which has a strongly developed posterior process. Among extant Osteoglossidae, the hyomandibulae of *Osteoglossum*, *Scleropages* and *Heterotis* have a short posterior process, whereas the hyomandibula of *Arapaima* has a very long and pillar-like posterior process. †Phareodontines also possess a very long posterior process of the hyomandibula, being even longer than its dorsal surface in †*Phareodus encaustus*. However, in contrast to *Arapaima*, the posterior process in †phareodontines points slightly ventrally rather than just extending posteriorly.

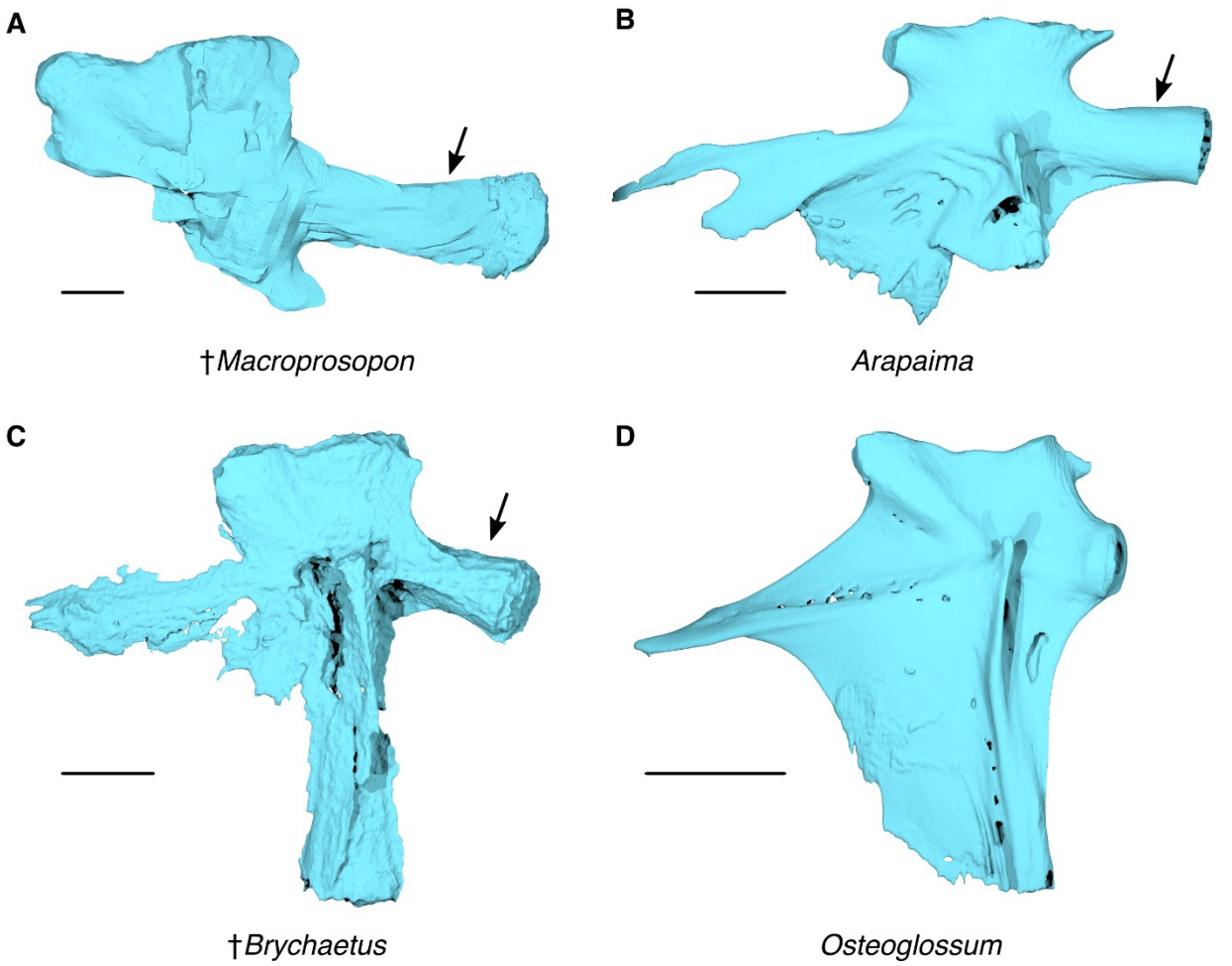

Fig. 4.8: Posterior process of the hyomandibula in osteoglossomorphs. Lateral views of left hyomandibulae (digital renderings). Arrows point to a long (more than half the length of the dorsal articulating surface of the hyomandibula) posterior process (character 94, state 1). A, †*Macroprosopon hiltoni* (UMMP 118216; partial hyomandibula); B, *Arapaima gigas* (UF 33107); C, cf. †*Brychaetus* sp. (NHMUK PV P26758); D, *Osteoglossum bicirrhosum* (UF 189007). Scale bars: 10 mm (A, B); 5 mm (C, D).

Character (95): Endopterygoid dentition: patch of shagreen-like fine teeth or small conical teeth [0]; few rows of large conical teeth [1]; one or more medio-dorsal rows of large conical teeth, bordered laterally by a patch of shagreen-like fine teeth [2]; teeth absent or extremely reduced [3]. The oral surface of the endopterygoid bears a patch of shagreen-like fine teeth in extant and fossil holosteans (Grande & Bemis, 1998; Grande, 2010) and in several stem and crown teleosts (Forey, 1973; Arratia, 2013, 2016; Cavin *et al*., 2013). The dentition of the endopterygoid varies greatly within osteoglossomorphs. Among extant bonytongues, an endopterygoid with a patch of

fine teeth can be seen in *Hiodon*, *Arapaima* and most notopterids. The extension of this tooth patch on the endopterygoid varies greatly between these taxa, from the small tooth patch of *Hiodon* restricted to the ventrolateral area near the ectopterygoid (Hilton, 2002), to the tooth patch of *Arapaima* that covers the whole medial surface of the bone. The endopterygoids of *Pantodon* and *Heterotis* bear instead few rows of relatively large conical teeth dorso-mesially. Extant osteoglossines (*Osteoglossum* and *Scleropages*) have one or two rows of large teeth on the dorso-mesial margin of the endopterygoid, bordered laterally by a patch of fine teeth that covers the whole medial surface of the bone. Teeth are completely absent from the endopterygoid in mormyrids, *Gymnarchus*, and *Xenomystus* (in mormyrids and *Gymnarchus,* the endopterygoid is fused with the ectopterygoid in a single bone; Taverne, 1998). Among fossil osteoglossomorphs for which the medial side of the endopterygoid can be observed, a patch of shagreen-like fine teeth is present in †*Lycoptera* (Ma, 1987), †*Paralycoptera*, †*Shuleichthys*, †*Laeliichthys*, †*Chauliopareion*, and †*Brychaetus*. †*Phareodus* has a single row of very large conical teeth in the dorso-mesial margin of the bone, bordered laterally by a few rows of teeth decreasing progressively in size until they become a shagreen-like tooth patch that covers the rest of the bone (a condition very similar to extant osteoglossines). Teeth are apparently absent from the endopterygoid of †*Singida* (Murray & Wilson, 2005).

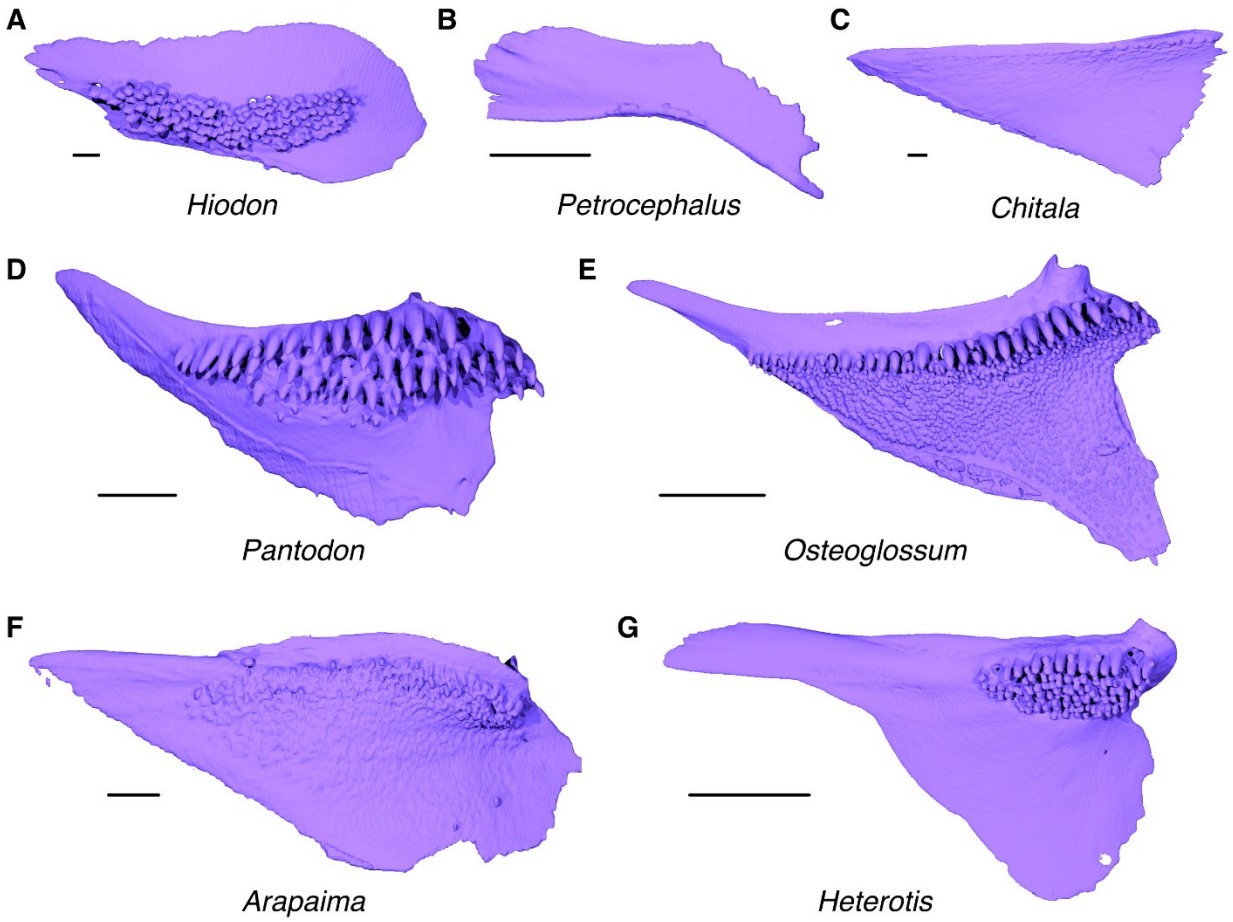

Fig. 4.9: Endopterygoid dentition (character 95) in extant osteoglossomorphs. Medial views of right endopterygoids (digital renderings). State 0 (patch of shagreen-like fine teeth or small conical teeth) is represented by *Hiodon*, *Chitala*, and *Arapaima*; state 1 (few rows of large conical teeth) by *Pantodon* and *Heterotis*; state 2 (one or more medio-dorsal rows of large conical teeth, bordered laterally by a patch of shagreen-like fine teeth) by *Osteoglossum*; state 3 (teeth absent or extremely reduced) by *Petrocephalus*. A, *Hiodon tergisus* (UMMZ 247425); B, *Petrocephalus simus* (UMMZ 200167); C, *Chitala blanci* (UMMZ 232272); D, *Pantodon buchholzi* (UMMZ 249782); E, *Osteoglossum bicirrhosum* (UF 189007); F, *Arapaima gigas* (UF 33107); G, *Heterotis niloticus* (UMMZ 195004). Scale bars: 1 mm (A, B, C, D); 5 mm (E, F, G).

Character (96): Number of branchiostegal rays: 8 or fewer [0]; between 9 and 13 [1]; 14 or more [2]. Branchiostegal rays are long paired struts of dermal bone that form the floor of the gill chamber and are involved in ventilatory functions by being part of the buccal pump (Hughes, 1960; Farina *et al*., 2015). The number of branchiostegal rays varies widely within teleosts, but tends to not vary much within species and is conserved among closely related species (McAllister, 1968; Ascarrunz *et al*., 2019). Extant *Hiodon* species have 7–9 branchiostegal rays

(8 as modal value; Hilton, 2002). The butterflyfish *Pantodon* has 10 branchiostegals (Taverne, 1978). Despite their great taxonomic and morphological diversity, all mormyrids have either 7 or 8 branchiostegal rays (Taverne, 1968; 1969; 1971; 1972), while their sister taxon *Gymnarchus* is characterized by a reduced set of 4 branchiostegals. All notopterids have 8 or less branchiostegals (extremely reduced to only 3 rays in *Xenomystus*; Taverne, 1978). Extant osteoglossids show a large variance in number of branchiostegal rays, with *Heterotis* on the lower end of the spectrum (7–8 rays) and *Scleropages* on the higher end (14–16). In fossil specimens, the count of branchiostegal rays (when preserved) is relatively straightforward, but we acknowledge that this might not always reflect the true number of branchiostegals in extinct taxa, due to lack of preservation of loosely attached rays or to difficulty in distinguishing left and right branchiostegal series in two-dimensionally preserved specimens. Most fossil taxa considered in this analysis have between 8 and 13 branchiostegal rays. Notable outliers include †*Wilsonichthys* with only 5 branchiostegals (Murray *et al*., 2016), and †*Brychaetus* and †*Macroprosopon* with 15–18 and 17 branchiostegals, respectively. After discretizing the number of branchiostegal rays into three states designed to minimize the number of taxa scored as polymorphic, this character was designated as ordered to reflect its underlying meristic nature and its relative phylogenetic conservatism.

Phylogenetic analysis

The maximum parsimony (MP) phylogenetic analysis recovered six most parsimonious trees, with tree length = 375, consistency index (CI) = 0.4080, retention index (RI) = 0.6487, and rescaled consistency index (RC) = 0.2647. The most parsimonious trees differ in the relative positions of †*Shuleichthys*, †*Wilsonichthys*, and the hiodontid clade (*Eohiodon* + *Hiodon*), in the positions of †*Xixiaichthys*, †*Paralycoptera*, †*Chauliopareion* and the †*Joffrichthys* clade (†*J. symmetropterus* + †*J. tanyourus*) with respect to the rest of Osteoglossiformes, and in the relationships of the three notopterid genera included in the analysis (*Chitala*, *Xenomystus* and *Papyrocranus*). †*Macroprosopon* is consistently recovered as an osteoglossid most closely related to †*Brychaetus*. †*Macroprosopon*, †*Brychaetus* and †*Phareodus* form a phareodontine clade to the exclusion of other osteoglossids. †*Furichthys* is reconstructed as sister taxon to a crown Arapaiminae clade formed by *Arapaima*, *Heterotis* and †*Sinoglossus*. A close relationship

between phareodontines and arapaimines to the exclusion of Osteoglossinae (*Osteoglossum* + *Scleropages*) is strongly supported (Bremer index = 4). A crown Osteoglossidae clade that excludes *Pantodon* and †*Singida* is also strongly supported by the MP analysis (Bremer index = 4).

Maximum likelihood (ML) and Bayesian phylogenetic analyses recovered identical tree topologies. They differ from the MP strict consensus tree in having *Pantodon* (rather than †*Singida*) as sister taxon to crown Osteoglossidae, and in recovering †*Furichthys* as sister taxon to †*Macroprosopon* (instead of being more closely related to crown arapaimines). Statistical support in both analyses is relatively high for the node uniting arapaimines and phareodontines (posterior probability = .97, ML bootstrap = 96%), for crown Arapaiminae (posterior probability = 1, ML bootstrap = 99%) and for crown Osteoglossidae (posterior probability = 1, ML bootstrap = 98%).

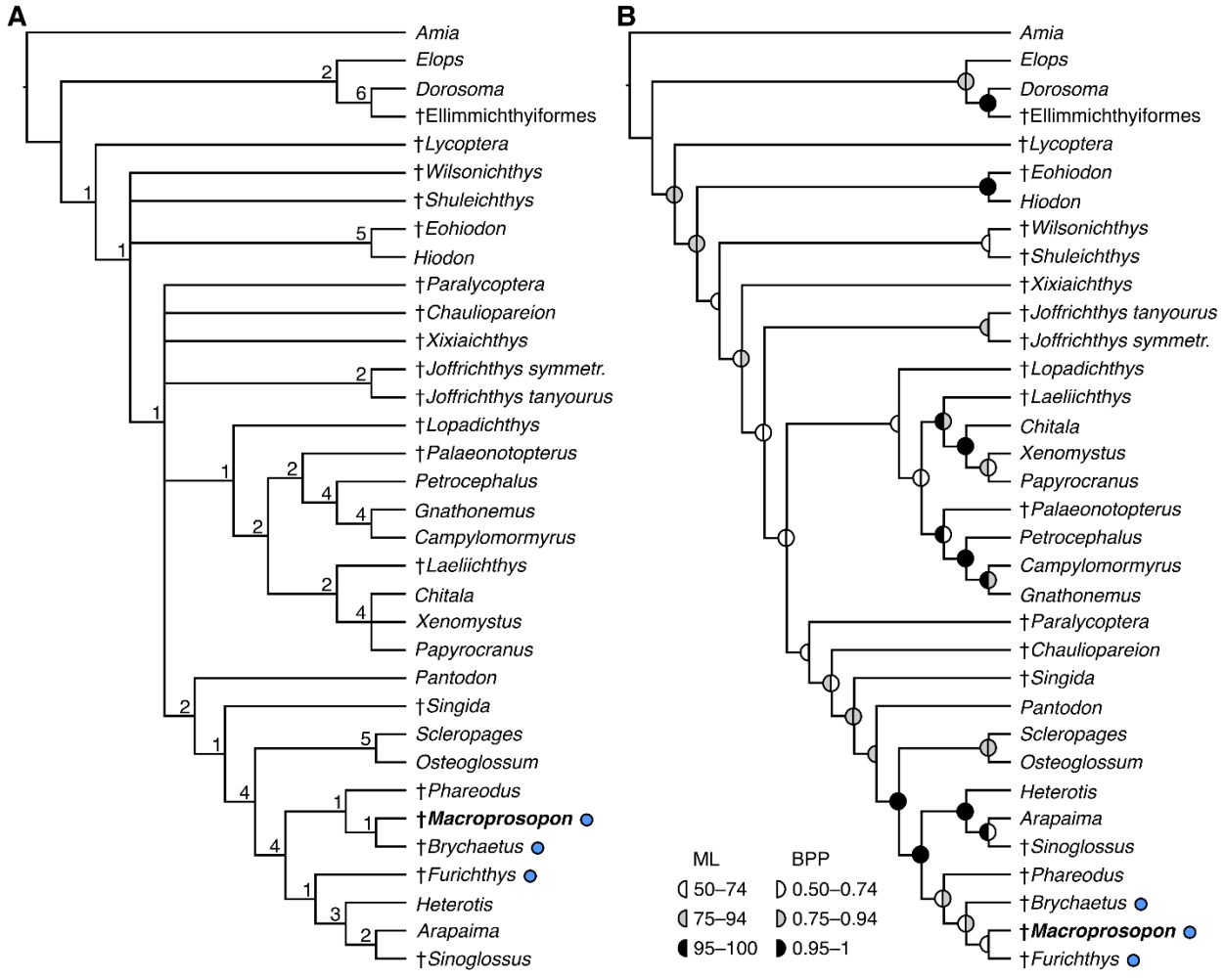

Fig. 4.10 (*previous page*)**.** Phylogenetic relationships of Osteoglossomorpha. The position of †*Macroprosopon hiltoni* is highlighted in bold. Marine taxa are marked with a blue dot. A, strict consensus tree of 20 maximum parsimony phylogenies, with Bremer decay indices above nodes; B, tree topology shared by the maximum likelihood and Bayesian consensus trees. Statistical support values for nodes are indicated by shaded semicircles. Maximum likelihood bootstrap percentages (ML) and Bayesian posterior probabilities (BPP) are represented by the left and right semicircles, respectively. Nodes without a right semicircle have less than 0.50 BPP.

DISCUSSION

Phylogenetic position of Macroprosopon with remarks on osteoglossomorph phylogeny

In every phylogenetic analysis performed here, †*Macroprosopon* is a member of Osteoglossidae closely related to the freshwater †*Phareodus* and to the marine †*Brychaetus*. Additionally, in the maximum likelihood topology it is recovered as sister taxon to †*Furichthys*, another long-snouted marine bonytongue. †*Macroprosopon*, †*Phareodus* and †*Brychaetus* (possibly together with †*Furichthys*) form a distinct osteoglossid clade that corresponds to the †Phareodontinae coined by Taverne (1979; see Hilton & Lavoué (2018) for the taxonomic history of †Phareodontinae). Besides being characterized by a series of unique anatomical features such as the presence of a dorso-occipital fossa and of a lateral expansion of the frontal (see "Comparison between other extinct and extant osteoglossids" below), †Phareodontinae can be defined operationally as the clade including every taxon most closely related to †*Phareodus* than either *Arapaima* or *Osteoglossum*. Contrary to most previous phylogenetic analyses (e.g., Li *et al*., 1997a; Hilton, 2003; Lavoué, 2016), †*Phareodus* (together with other †phareodontines) is here found to be more closely related to Arapaiminae than to Osteoglossinae. In fact, every extinct osteoglossid included in this analysis is either an arapaimine or lies on the arapaimine stem. Remarkably, phylogenetic relationships within Osteoglossidae (except for the position of the incompletely known †*Furichthys*) seem robust to the different methods of phylogenetic analysis used here, and nodes within this group have stronger statistical support than most other nodes in osteoglossomorph phylogeny as indicated by Bremer decay indices, ML bootstrap values, and Bayesian posterior probabilities.

Outside of Osteoglossidae, the phylogenetic hypotheses supported in this study are topologically compatible with previously published osteoglossomorph phylogenies. The only exception is presented by †*Paralycoptera*, which is recovered as an osteoglossiform (possibly a crown-

member of the group) instead of a stem osteoglossomorph as in previous studies (e.g., Murray & Wilson, 2008; Lavoué, 2016). This is due to the rescoring of this taxon based on a recent redescription (Xu & Chang, 2009). Whereas all major extant clades (Osteoglossidae, Mormyridae and Notopteridae) are relatively well-supported, the backbone of osteoglossomorph phylogeny is highly unstable. Previous work has shown that there are few or no characters supporting several relationships found in the more basal portions of the osteoglossomorph tree, and that exclusion of just one character or one taxon from the morphological matrix has the potential to substantially change the position of some fossil taxa (Murray *et al*., 2018). Although a more uncertain placement of extinct taxa in comparison to extant ones should be generally expected based on loss of information in fossils, this problem is exacerbated by the state of preservation of most Cretaceous and Paleocene osteoglossomorphs. These are known from flattened and approximately two-dimensional specimens often lacking any information about highly informative anatomical regions, such as hyoid and branchial skeletons and some parts of the neurocranium. Exceptions are represented by the Paleocene osteoglossids †*Taverneichthys* and †*Magnigena*, which are known from three-dimensional articulated cranial material, thus being key taxa for future investigation of the evolutionary history of bonytongues. Inclusion of molecular data in a total-evidence (morphology + molecules) approach might at least help to stabilize deep nodes subtended by extant taxa, and possibly increase the statistical support for the placement of some fossil taxa as well.

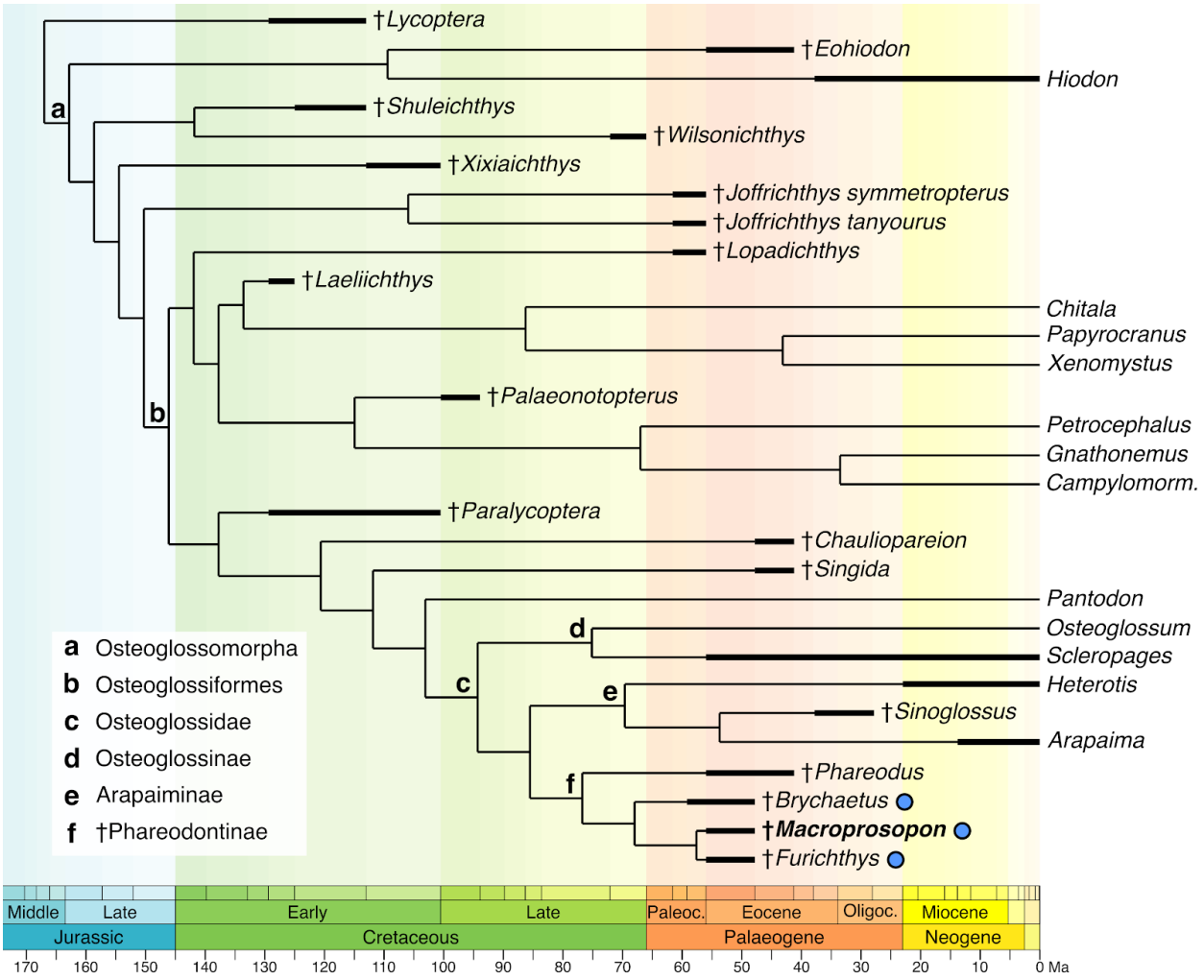

Fig. 4.11. Time-scaled phylogenetic tree of Osteoglossomorpha, based on the topology of the maximum likelihood phylogeny. The position of †*Macroprosopon hiltoni* is highlighted in bold. Marine taxa are marked with a blue dot. Stage-level temporal ranges of taxa with fossil record are indicated by solid black bars. Node ages are plotted using the 'equal' method of *a posteriori* time-scaling with user-defined root age (Lloyd *et al*., 2012); these are not analytical estimates under an evolutionary model. Root age is fixed to the minimum fossil-based age for total-group Osteoglossomorpha estimated in Capobianco & Friedman (2019).

Comparison with other extinct and extant osteoglossids

The phylogenetic position of †*Macroprosopon* as a member of Osteoglossidae closely related to arapaimines is supported by several characters. †*Macroprosopon* shares with other osteoglossids the presence of a postero-dorsal flange of the angular, a character that is uniquely present in osteoglossids among Osteoglossomorpha and has been secondarily lost in *Osteoglossum*.

Additionally, in †*Macroprosopon*, like in other osteoglossids (except †*Phareodus*), the supratemporal canal passes through the parietals. This feature evolved independently several times within Osteoglossomorpha, as it is present also in †*Joffrichthys symmetropterus*, †*Chauliopareion*, and notopterids. We were not able to determine the presence or absence of other synapomorphies of Osteoglossidae in †*Macroprosopon*, including: a foramen for cranial nerve V + anteroventral lateral line nerve in the prootic (also observed in *Hiodon* and secondarily modified in *Arapaima*); an anterior process of the hyomandibula contacting the endopterygoid; a small subopercle anterior to the opercle (also seen in †*Chauliopareion* and secondarily modified in †*Furichthys*); the lateral line piercing the supracleithrum (independently evolved in notopterids as well); and a dorsal fin with long base and rounded outline.

†*Macroprosopon* shares exclusively with arapaimines and with other †phareodontines an autogenous articular (thus having angular, articular, and retroarticular all separate from each other) and a greatly hypertrophied parapophysis on the first vertebra. Additional characters link †*Macroprosopon* and other †phareodontines to arapaimines, but they also evolved independently in more distantly related osteoglossomorphs. These are the presence of two separate neurocranial heads of the hyomandibula (also seen in †*Lopadichthys*), and a dorsal posterior infraorbital deeper than the ventral one (also seen in †*Paralycoptera*, *Xenomystus* and some specimens of †*Eohiodon*). A long posterior process of the hyomandibula also appears to be a synapomorphy of the Arapaiminae + †Phareodontinae clade, as it is present in *Arapaima*, †*Macroprosopon*, †*Brychaetus*, †*Phareodus* and †*Furichthys* (*Heterotis* possesses a short posterior process of the hyomandibula instead). This state evolved independently in hiodontids and †*Paralycoptera*. Similarly to arapaimines but unlike †*Phareodus* and †*Furichthys*, the retroarticular of †*Macroprosopon* takes part in the articulation between lower jaw and quadrate. Other synapomorphies of arapaimines, such as a third infrapharyngobranchial divided into two elements (one of which entirely cartilaginous), are very unlikely to be observed in fossil taxa because they involve non-mineralized or poorly mineralized tissues.

†*Macroprosopon*, †*Phareodus*, and †*Brychaetus* uniquely share the presence of a dorso-occipital fossa and of a supraorbital shelf of the frontal. We were not able to ascertain the state of these characters in †*Furichthys*. †*Macroprosopon* also shares with †*Brychaetus* the presence of a scleral ring formed by scleral ossicles, which is absent in †*Phareodus* and apparently in

†*Furichthys* – although for the latter it is difficult to establish whether the absence of a scleral ring in the only specimen with preserved orbital region is due to true absence or to an artifact of preservation. The presence of a well-developed scleral ring in both †*Macroprosopon* and †*Brychaetus* is intriguing from an ecomorphological perspective, as scleral ossicles tend to be more common in fish clades with an active lifestyle (Franz-Odendaal, 2008) and they are more robust and forming a complete ring in fast pelagic predators such as scombrids and istiophorids (Nakamura & Yamaguchi, 1991; Franz-Odendaal, 2008). †*Macroprosopon* and †*Brychaetus* are also characterized by a high branchiostegal ray count (17 and 15–18, respectively). Among all other osteoglossomorphs, only *Scleropages* bears a similarly high count (between 14 and 16). The number of branchiostegal rays is unknown in †*Furichthys*. While it has been historically suggested that higher numbers of branchiostegals are found in marine fishes (Hubbs, 1919), recent phylogenetic comparative studies do not find statistical support for this proposed pattern (Ascarrunz *et al*., 2019).

The scales of †*Macroprosopon* seem to lack the reticulate furrows that are characteristic of osteoglossid scales. Instead, they present an ornamentation of small tubercles, similar to the scales of †*Furichthys* and of †*Monopteros*—a marine bonytongue with uncertain phylogenetic affinities from the early Eocene Bolca Lagerstätte (Taverne, 1998). Interestingly, these tubercles resemble the ones found on the external surface of posterior squamules (scale fragments) of extant osteoglossids, and in particular of *Heterotis* (pers. obs. of UMMZ 213845), except for their larger size.

Diversity of early Paleogene marine osteoglossids and biogeographic remarks

The discovery of †*Macroprosopon* not only increases the known taxonomic diversity of marine bonytongues, but highlights how these fishes diversified in a wide range of morphologies that likely reflect a previously underappreciated ecological variety. Unlike other coeval marine bonytongues characterized by short and robust snouts, such as †*Brychaetus* and several other †*Phareodus*-like forms, *Macroprosopon* and the potentially closely related †*Furichthys* are long-snouted bonytongues, with elongated straight lower jaws and an extensive preorbital region of the skull. The long snout and sub-conical tooth tips of †*Macroprosopon* indicate a predatory—and likely piscivorous—feeding ecology for this taxon. The aspect ratio of its lower jaw

(relatively moderate height compared to length) suggests that †*Macroprosopon* had fast-closing jaws and probably a weaker bite relative to †*Brychaetus* and †*Phareodus* (Westneat, 2004). Although elongated lower jaws capable of fast-bite strikes are also found in some extant osteoglossids (*Osteoglossum* and *Scleropages*), these taxa have a short preorbital region of the skull and their mouth is superior (strongly oblique), suggesting that modern bonytongues are not direct trophic analogues for the early Paleogene long-snouted forms. Strikingly, long-snouted and short-snouted predatory bonytongues likely co-existed in the same marine habitats. This is implied by the occurrence of several osteoglossid jaw fragments and isolated teeth found in Ypresian Moroccan phosphates that do not correspond to the morphology of †*Macroprosopon* and are instead referable to †*Brychaetus* or a †*Brychaetus*-like form—characterized by more robust jaw bones and teeth and by a taller bony collar at the tooth base (Arambourg, 1952; Forey & Hilton, 2010). Similarly, in the Danish Fur Formation †*Furichthys* is found alongside †*Brychaetus*-like jaw fragments (plus several other bonytongue species; Bonde, 2008). Thus, at least in the western Tethys and North Sea, during the earliest Eocene osteoglossids occupied different trophic niches within the guild of large carnivorous fishes.

†*Macroprosopon* joins a great diversity of early Paleogene marine bonytongues, which are now known from more than 10 different genera—a striking contrast with the depauperate taxonomic richness of extant osteoglossids. Several of these marine taxa, such as †*Monopteros*, †*Heterosteoglossum*, †*Brychaetoides*, †*Xosteoglossid*, and †*Thrissopterus*, remain poorly known with uncertain phylogenetic affinities (Taverne, 1998; Bonde, 2008; Hilton & Lavoué, 2018; Capobianco *et al.*, 2021). Although some of them are known from fragmentary or poorly preserved remains (thus complicating efforts to elucidate their relationships), they hint at a much higher degree of ecomorphological disparity of Palaeogene osteoglossids relative to extant representatives of this clade, including forms with durophagous dentition and others with broad pectoral fins and very elongated bodies (Capobianco *et al.*, 2021).

The phylogenetic placement of the marine †*Macroprosopon*, †*Furichthys* and †*Brychaetus* as stem arapaimines suggests that marine dispersal might have played a role in the present-day disjunct geographic distribution of this clade, with *Arapaima* endemic to South America and *Heterotis* endemic to Africa. Model-based ancestral state reconstruction and biogeographic inference might provide a way to test whether arapaimines were ancestrally marine, thus making

extant *Arapaima* and *Heterotis* secondarily freshwater taxa. We point out that the phylogenetic hypotheses derived in this study are also compatible with one or two freshwater-to-marine transitions not involving the lineage leading directly to crown arapaimines. However, the presence of †*Phareodus*-like and †*Brychaetus*-like fossils in freshwater and marine deposits worldwide (Capobianco *et al*., 2021) lends some credibility to the scenario of long-distance marine dispersals followed by multiple marine-to-freshwater transitions. Inclusion of additional fossil osteoglossids within the phylogenetic framework laid out in this study will be paramount to paint a clearer picture of the biogeographic history of bonytongue fishes.

The diversity of early Paleogene bonytongues and their worldwide presence in a variety of depositional environments hints at an unexpected evolutionary radiation of this group of fishes. This radiation might have been triggered or facilitated by the K/Pg mass extinction, which decimated large predatory fishes and other trophic specialists in marine settings, opening up new ecological opportunities for surviving taxa (Cavin, 2002; Friedman, 2009; Capobianco et al., 2021). The inclusion of fossil bonytongues in a 'tip-dated' phylogenetic analysis could help to better constrain the time of origin of the early Paleogene bonytongue radiation and would be an important step towards clarifying the role of the K/Pg mass extinction in the evolutionary history of osteoglossomorphs.

Future efforts to resolve the phylogenetic affinities of Paleogene marine bonytongues will be instrumental in elucidating some of the outstanding questions about osteoglossomorph evolution that remain unanswered. These include when and how many times bonytongues transitioned from freshwater to marine environments (and vice-versa); how these environmental transitions affected the biogeographic history of the group; and the role of the K/Pg mass extinction in the diversification dynamics of bonytongues. We believe that the well-supported close relationship between arapaimines and †phareodontines found in this study is a promising first step towards answering these questions.

ACKNOWLEDGMENTS

The authors thank William Sanders (University of Michigan Museum of Paleontology) for the mechanical preparation of the specimen described in this paper. We thank Kyle Kramer

(Undergraduate Research Opportunity Program at the University of Michigan) for great help with segmentation of CT data. We also thank for specimen access Douglas Nelson and Randy Singer at the University of Michigan Museum of Zoology, William F. Simpson at the Field Museum of Natural History, Chicago, Bo Schultz and René L. Sylvestersen at the Fur Museum, Bent E. K. Lindow at the Natural History Museum of Denmark, Emma Bernard at the Natural History Museum, London, Mariagabriella Fornasiero at the Istituto Geologico dell'Università di Padova, Anna Vaccari and Roberta Salmaso at the Museo Civico di Storia Naturale, Verona, Gaël Clement at the Muséum National d'Histoire Naturelle, Paris and Florias Mees at the Musée Royal de l'Afrique Centrale. We would like to thank Zach Randall and the Florida Museum of Natural History, University of Florida for permission to use CT data stored on Morphosource. We thank Charlie Underwood for his identification of the embedded shark tooth. We finally would like to thank the Friedman Lab students and postdocs for helpful discussion and comments on this manuscript. This study includes data produced in the CTEES facility at University of Michigan, supported by the Department of Earth & Environmental Sciences and College of Literature, Science, and the Arts. This work was supported by funding from the Department of Earth and Environmental Sciences of the University of Michigan (Scott Turner Student Research Grant Award 2017, to A.C.), by the Rackham Graduate School of the University of Michigan (Rackham Predoctoral Fellowship Award 2020-2021, to A.C.) and by the Society of Systematic Biologists (2017 SSB Graduate Student Research Award, to A.C.).

REFERENCES

Agassiz, L. (1845). Report on the fossil fishes of the London Clay. *Report of the British Association for the Advancement of Science*, 14, 279–310.

Alvarado-Ortega, J., Cuevas-García, M., del Pilar Melgarejo-Damián, M., Cantalice, K. M., Alaniz-Galvan, A., Solano-Templos, G., & Than-Marchese, B. A. (2015). Paleocene fishes from Palenque, Chiapas, southeastern Mexico. *Palaeontologia Electronica*, 18, 1–22.

Arambourg, C. (1952). Les vertébrés fossiles des gisements de phosphates (Maroc-Algérie-Tunisie). *Notes et Mémoires du Service Géologique du Maroc*, 92, 1–372.

Arratia, G. (1997). Basal teleosts and teleostean phylogeny. *Palaeo Ichthyologica,* 7, 5–168.

Arratia, G. (2013). Morphology, taxonomy, and phylogeny of Triassic pholidophorid fishes (Actinopterygii, Teleostei). *Journal of Vertebrate Paleontology*, 33 (supplement 1), 1–138.

Arratia, G. (2016). New remarkable Late Jurassic teleosts from southern Germany: Ascalaboidae n. fam., its content, morphology, and phylogenetic relationships. *Fossil Record*, 19, 31–59.

Arratia, G., & Schultze, H. P. (1991). Palatoquadrate and its ossifications: development and homology within osteichthyans. *Journal of morphology*, 208, 1–81.

Ascarrunz, E., Sánchez-Villagra, M. R., Betancur-R, R., & Laurin, M. (2019). On trends and patterns in macroevolution: Williston's law and the branchiostegal series of extant and extinct osteichthyans. *BMC evolutionary biology*, 19, 1–11.

Banan Khojasteh, S. M. (2012). The morphology of the post-gastric alimentary canal in teleost fishes: a brief review. *International Journal of Aquatic Science*, 3, 71–88.

Berg, L. (1940). *Classification of fishes both recent and fossil*. Travaux de l'Institut Zoologique de l'Académie des Sciences de l'URSS, Moscow, 517 pp. [Translated and reprinted in English, 1947, Ann Arbor]

Berra, T. M. (2007). *Freshwater Fish Distribution*. The University of Chicago Press, Chicago.

Blakey, R. C. (2008). Gondwana paleogeography from assembly to breakup—A 500 my odyssey. *Geological Society of America Special Papers*, 441, 1-28.

Bonde, N. (2008). Osteoglossomorphs of the marine Lower Eocene of Denmark – with remarks on other Eocene taxa and their importance for palaeobiogeography. *Geological Society of London, Special Publications*, 295, 253– 310.

Brito, P. M., Figueiredo, F. J., & Leal, M. E. C. (2020). A revision of *Laeliichthys ancestralis* Santos, 1985 (Teleostei: Osteoglossomorpha) from the Lower Cretaceous of Brazil: Phylogenetic relationships and biogeographical implications. *Plos one*, 15, e0241009.

Britz, R., & Johnson, G. D. (2010). Occipito-vertebral fusion in actinopterygians: conjecture, myth and reality. Part 1: non-teleosts. In: Nelson, J. S., Schultze, H. P., & Wilson, M. V. H. (eds.) *Origin and Phylogenetic Interrelationships of Teleosts. Honoring Gloria Arratia*, pp. 77–95. Verlag Dr. F. Pfeil, München.

Capobianco, A., & Friedman, M. (2019). Vicariance and dispersal in southern hemisphere freshwater fish clades: a palaeontological perspective. *Biological Reviews*, 94, 662–699.

Capobianco, A., Beckett, H. T., Steurbaut, E., Gingerich, P. D., Carnevale, G., & Friedman, M. (2020). Large-bodied sabre-toothed anchovies reveal unanticipated ecological diversity in early Palaeogene teleosts. *Royal Society Open Science*, 7, 192260.

Capobianco, A., Foreman, E., & Friedman, M. (2021). A Paleocene (Danian) marine osteoglossid (Teleostei, Osteoglossomorpha) from the Nuussuaq Basin of Greenland, with a brief review of Palaeogene marine bonytongue fishes. *Papers in Palaeontology*, 7, 625–640.

Casier, E. (1966). *Faune icthyologique du London Clay: Atlas*, Edition (Volume **2**). British Museum (Natural History), London.

Cavin, L. (2002). Effects of the Cretaceous-Tertiary boundary event on bony fishes. In: Buffetaut, E., & Koeberl, C. (eds.) *Geological and biological effects of impact events*, pp. 141-158. Springer, Berlin, Heidelberg.

Cavin, L., Forey, P. L., & Giersch, S. (2013). Osteology of *Eubiodectes libanicus* (Pictet & Humbert, 1866) and some other ichthyodectiformes (Teleostei): phylogenetic implications. *Journal of Systematic Palaeontology*, 11, 115–177.

Farina, S. C., Near, T. J., & Bemis, W. E. (2015). Evolution of the branchiostegal membrane and restricted gill openings in Actinopterygian fishes. *Journal of morphology*, 276, 681–694.

Forey, P. L. (1973). A revision of the Elopiformes fishes, fossil and recent. *Bulletin of the British Museum of Natural History (Geology)*, 10, 1–222.

Forey, P. L. & Hilton, E. J. (2010). Two new Tertiary osteoglossid fishes (Teleostei: Osteoglossomorpha) with notes on the history of the family. *In* D. K. Elliott, J. G. Maisey, X. Yu & D. Miao (eds). *Morphology, phylogeny and paleobiogeography of fossil fishes*, pp. 215–246. Verlag Dr. F. Pfeil, München.

Franz-Odendaal, T. A. (2008). Scleral ossicles of teleostei: evolutionary and developmental trends. *The Anatomical Record*, 291, 161–168.

Franz-Odendaal, T. A. (2020). Skeletons of the eye: An evolutionary and developmental perspective. *The Anatomical Record*, 303, 100–109.

Friedman, M. (2009). Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. *Proceedings of the National Academy of Sciences*, 106, 5218–5223.

Gayet, M. & Meunier, F. J. (1998). Maastrichtian to early late Paleocene freshwater Osteichthyes of Bolivia: additions and comments. *In* L. R. Malabarba, R. E. Reis, R. P. Vari, Z. M. Lucena & C. A. S. Lucena (eds). *Phylogeny and classification of neotropical fishes*, pp. 85–110. EdiPUCRS, Porto Alegre.

Gayet, M., Marshall, L. G., Sempere, T., Meunier, F. J., Cappetta, H., & Rage, J. C. (2001). Middle Maastrichtian vertebrates (fishes, amphibians, dinosaurs and other reptiles, mammals) from Pajcha Pata (Bolivia). Biostratigraphic, palaeoecologic and palaeobiogeographic implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 169, 39–68.

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. (2010). An empirical synthetic pattern study of gars (Lepisosteiformes) and closely related species, based mostly on skeletal anatomy. The resurrection of Holostei. *Copeia,* 2010 (Suppl. 2A), 1– 871.

Grande, L. & Bemis, W. E. (1998). A comprehensive phylogenetic study of amiid fishes (Amiidae) based on comparative skeletal anatomy. An empirical search for interconnected patterns of natural history. *Journal of Vertebrate Paleontology,* 18, (Suppl. 1), 1– 696.

Greenwood, P. H. (1970). On the genus *Lycoptera* and its relationship with the family Hiodontidae (Pisces, Osteoglossomorpha). *Bulletin of the British Museum of Natural History (Zoology)*, 19, 259–285.

Greenwood, P. H., Rosen, D. E., Weitzman, S. H., & Myers, G. S. (1966). Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History*, 131, 339–456.

Hilton, E. J. (2002). Osteology of the extant north American fishes of the genus *Hiodon* Lesueur, 1818 (Teleostei: Osteoglossomorpha: Hiodontiformes). *Fieldiana Zoology*, 100, 1–142.

Hilton, E. J. (2003). Comparative osteology and phylogenetic systematics of fossil and living bony-tongue fishes (Actinopterygii, Teleostei, Osteoglossomorpha). *Zoological Journal of the Linnean Society*, 137, 1–100.

Hilton, E. J., & Grande, L. (2008). Fossil mooneyes (Teleostei: Hiodontiformes, Hiodontidae) from the Eocene of western North America, with a reassessment of their taxonomy. *Geological Society, London, Special Publications*, 295, 221–251.

Hilton, E. J., & Lavoué, S. (2018). A review of the systematic biology of fossil and living bony-tongue fishes, Osteoglossomorpha (Actinopterygii: Teleostei). *Neotropical Ichthyology*, 16, e180031.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, 35, 518–522.

Hubbs, C. L. (1919). A comparative study of the bones forming the opercular series of fishes. *Journal of Morphology*, 33, 60–71.

Hughes, G. M. (1960). A comparative study of gill ventilation in marine teleosts. *Journal of Experimental Biology*, 37, 28–45.

Jin, F., Zhang, J. & Zhou, Z. (1995). Late Mesozoic fish fauna from western Liaoning, China. *Vertebrata PalAsiatica,* 33, 169–193.

Kumar, K., Rana, R. S. & Paliwal, B. S. (2005). Osteoglossid and lepisosteid fish remains from the Paleocene Palana Formation, Rajasthan, India. *Palaeontology,* 48, 1187–1210.

Lastein, U., & Van Deurs, B. (1973). The copulatory organ of *Pantodon buchholzi* Peters (Teleostei). *Acta Zoologica*, 54, 153–160.

Lavoué, S. (2015). Testing a time hypothesis in the biogeography of the arowana genus *Scleropages* (Osteoglossidae). *Journal of Biogeography*, 42, 2427–2439.

Lavoué, S. (2016). Was Gondwanan breakup the cause of the intercontinental distribution of Osteoglossiformes? A time-calibrated phylogenetic test combining molecular, morphological, and paleontological evidence. *Molecular phylogenetics and evolution*, 99, 34–43.

Li, G. Q. (1994). Systematic position of the Australian fossil osteoglossid fish †*Phareodus* (= *Phareoides*) *queenslandicus* Hills. *Memoirs of the Queensland Museum*, 37, 287–300.

Li, G. Q. (1996). A new species of Late Cretaceous osteoglossid (Teleostei) from the Oldman Formation of Alberta, Canada, and its phylogenetic relationships. In *Mesozoic Fishes. Systematics and Paleoecology* (eds G. Arratia and G. Viohl), pp. 285–298. Verlag Dr. F. Pfeil, München.

Li, G.Q., & Wilson, M. V. (1996). Phylogeny of Osteoglossomorpha. In Stiassny, M. L. J., Parenti, L. R., & Johnson, G. D. (eds.) *Interrelationships of Fishes*, pp.163–174 . Academic Press, San Diego.

Li, G. Q., Grande, L., & Wilson, M. V. (1997a). The species of †*Phareodus* (Teleostei: Osteoglossidae) from the Eocene of North America and their phylogenetic relationships. *Journal of Vertebrate Paleontology*, 17, 487–505.

Li, G. Q., Wilson, M. V., & Grande, L. (1997b). Review of *Eohiodon* (Teleostei: Osteoglossomorpha) from western North America, with a phylogenetic reassessment of Hiodontidae. *Journal of Paleontology*, 71, 1109–1124.

Lundberg, J. G., & Chernoff, B. (1992). A Miocene fossil of the Amazonian fish *Arapaima* (Teleostei, Arapaimidae) from the Magdalena River region of Colombia--Biogeographic and evolutionary implications. *Biotropica*, 24, 2–14.

Ma, F. Z. (1987). Review of *Lycoptera davidi*. *Vertebrata PalAsiatica*, 25, 8–19.

Maddison, W. P. & Maddison, D.R. (2019). Mesquite: a modular system for evolutionary analysis. Version 3.61. http://www.mesquiteproject.org

McAllister, D. E. (1968). Evolution of branchiostegals and classification of teleostome fishes. *Bulletin of the National Museum of Canada,* 221, 1–237.

McDowall, R. M., & Burridge, C. P. (2011). Osteology and relationships of the southern freshwater lower euteleostean fishes. *Zoosystematics and Evolution*, 87, 7–185.

Mok, H. K., & Liu, S. H. (2012). Morphological variations in the scleral ossicles of 172 families of actinopterygian fishes with notes on their phylogenetic implications. *Zoological Studies*, 51, 1490–1506.

Müller, J. (1845). Uber den Bau und die Grenzen der Ganoiden, und uber das naturliche System der Fische. *Physikalisch-Mathematische Abhandlungen der koniglichen Akademie der Wissenschaften zu Berlin* 1845, 117–216.

Murray, A. M. & Wilson, M. V. H. (2005). Description of a new Eocene osteoglossid fish and additional information on †*Singida jacksonoides* Greenwood and Patterson, 1967 (Osteoglossomorpha), with an assessment of their phylogenetic relationships. *Zoological Journal of the Linnean Society,* 144, 213–228.

Murray, A. M., & Wilson, M. V. (2013). Two new paraclupeid fishes (Clupeomorpha: Ellimmichthyiformes) from the Upper Cretaceous of Morocco. *Mesozoic fishes 5*, 267–290.

Murray, A. M., You, H. L., & Peng, C. (2010). A new Cretaceous osteoglossomorph fish from Gansu Province, China. *Journal of Vertebrate Paleontology*, 30, 322–332.

Murray, A. M., Newbrey, M. G., Neuman, A. G., & Brinkman, D. B. (2016). New articulated osteoglossomorph from Late Cretaceous freshwater deposits (Maastrichtian, Scollard Formation) of Alberta, Canada. *Journal of Vertebrate Paleontology*, 36, e1120737.

Murray, A. M., Zelenitsky, D. K., Brinkman, D. B., & Neuman, A. G. (2018). Two new Palaeocene osteoglossomorphs from Canada, with a reassessment of the relationships of the genus †*Joffrichthys*, and analysis of diversity from articulated versus microfossil material. *Zoological Journal of the Linnean Society*, 183, 907–944.

Nakamura, K., & Yamaguchi, H. (1991). Distribution of scleral ossicles in teleost fishes. *Memoirs of Faculty of Fisheries, Kagoshima University*, 40, 1–20.

Nelson, G. J. (1972). Observations on the gut of the Osteoglossomorpha. *Copeia*, 17, 325–329.

Otero, O., & Gayet, M. (2001). Palaeoichthyofaunas from the Lower Oligocene and Miocene of the Arabian Plate: palaeoecological and palaeobiogeographical implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, , 141–169.

Otero, O., Garcia, G., Valentin, X., Lihoreau, F., Manthi, F. K., & Ducrocq, S. (2017). A glimpse at the ectotherms of the earliest fauna from the East African Rift (Lokone, Late Oligocene of Kenya). *Journal of Vertebrate Paleontology*, 37, e1312691.

Patterson. C. 1975. The distribution of Mesozoic freshwater fishes. *Memoires du Museum national d'Histoire naturelle de Paris, serie A*, 88, 156–174.

Patterson, C., & Rosen, D. E. (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.

Rambaut A. (2012). Figtree v 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/

Roellig, H. F. (1974). The cranial osteology of *Brychaetus muelleri* (Pisces Osteoglossidae) Eocene, Isle of Sheppey. *Journal of Paleontology*, 48, 947–951.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61, 539–542.

Sanford, C. P. J. (2000). Salmonoid fish osteology and phylogeny (Teleostei: Salmonoidei). *Theses Zoologicae*, 33, 1–264.

Su, D. (1986). The discovery of a fossil osteoglossid fish in China. *Vertebrata PalAsiatica*, 24, 10–19.

Swofford, D.L. (2002) PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4.0 Beta 10. Sinauer Associates, Sunderland.

Taverne, L. (1968). Ostéologie du genres *Gnathonemus* Gill sensu stricto, *Gnathonemus petersii* (Gthr.) et espèces voisines. *Annales du Musée Royal de l'Afrique Centrale, Sciences Zoologiques*, 170, 1-91.

Taverne, L. (1969). Etude ostéologique des genres *Boulengeromyrus* Taverne et Géry, *Genyomyrus* Boulenger, *Petrocephalus* Marcusen (Pisces Mormyriformes). *Annales du Musée Royal de l'Afrique Centrale, Sciences Zoologiques*, 174, 1–85.

Taverne, L. (1971). Ostéologie des genres *Marcusenius* Gill, *Hippopotamyrus* Pappenheim, *Cyphomyrus* Myers, *Pollimyrus* Taverne et *Brienomyrus* Taverne (Pisces, Mormyriformes). *Annales du Musée Royal de l'Afrique Centrale, Série In-8°, Sciences Zoologiques*, 188, 1-143.

Taverne, L. (1972). Ostéologie des genres *Mormyrus* Linné, *Mormyrops* Müller, *Hyperopisus* Gill, *Isichthys* Gill, *Myomyrus* Boulenger, *Stomatorhinus* Boulenger et *Gymnarchus* Cuvier:

considérations générales sur la systématique des poissons de l'ordre des Mormyriformes. *Musée royal de l'Afrique centrale—Tervuren, Belgique Annales—Serie in-8°—Sciences Zoologiques,* 200, 1– 198.

Taverne, L. (1977). Ostéologie, phylogénèse et systématique des Téléostéens fossiles et actuels du super-ordre des Ostéoglossomorphes, Première partie. Ostéologie des genres *Hiodon*, *Eohiodon*, *Lycoptera*, *Osteoglossum*, *Scleropages*, *Heterotis* et *Arapaima*. *Acadé mie Royale de Belgique, Mémoires de la Classe des Sciences, Collection in-8° - 2e série*, 42, 1– 235.

Taverne, L. (1978). Osteologie, phylogénèse et systématique des Téléostéens fossiles et actuels de super–ordre des Ostéoglossomorphes. Deuxième partie. Ostéologie des genres *Phareodus*, *Phareoides*, *Brychaetus*, *Musperia*, *Pantodon*, *Singida*, *Notopterus*, *Xenomyst us* et *Papyrocranus*. *Mémoires de la Classe des Sciences, Académie Royale de Belgique, Collection in–8°, 2e Série,* 42, 1–213.

Taverne, L. (1979). Osteologie, phylogénèse et systématique des Téléostéens fossiles et actuels de super-ordre des Ostéoglossomorphes. Troisième partie. Evolution des structures ostéologiques et conclusions générales relatives à la phylogénèse et à la systématique du super-ordre. Addendum. *Mémoires de la Classe des Sciences, Académie Royale de Belgique, Collection in-8°, 2e Série,* 43, 1–168.

Taverne, L. (1998). Les Ostéoglossomorphes marins de l'Eocène du Monte Bolca (Italie): *Monopteros* Volta, 1796, *Thrissopterus* Heckel, 1856 et *Foreyichthys* Taverne, 1979. Considérations sur la phylogénie des Téléostéens ostéoglossomorphes. *Studi e ricerche sui giacimenti Terziari di Bolca*, *Miscellanea Paleontologica*, 7, 67–158.

Taverne, L. (2009). New insights on the osteology and taxonomy of the osteoglossid fishes *Phareodus*, *Brychaetus* and *Musperia* (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre,* 79, 175–190.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research*, 44, W232–W235.

Westneat, M. W. (2004). Evolution of levers and linkages in the feeding mechanisms of fishes. *Integrative and Comparative Biology*, 44, 378–389.

Wilson, M. V. H., & Murray, A. M. (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.

Woodward, A. S. (1901). *Catalogue of fossil fishes in the British Museum (N. H.). Part IV*. London, Longmans and Co., 636 pp.

Xu, G. H., & Chang, M. M. (2009). Redescription of †*Paralycoptera wui* Chang & Chou, 1977 (Teleostei: Osteoglossoidei) from the Early Cretaceous of eastern China. *Zoological Journal of the Linnean Society*, 157, 83–106.

Yans, J., Amaghzaz, M. B., Bouya, B., Cappetta, H., Iacumin, P., Kocsis, L., Mouflih, M., Selloum, O., Sen, S., Storme, J. Y. & Gheerbrant, E. (2014). First carbon isotope

chemostratigraphy of the Ouled Abdoun phosphate Basin, Morocco; implications for dating and evolution of earliest African placental mammals. *Gondwana Research*, 25, 257–269.

Zhang, J. Y. (2002). A new species of *Lycoptera* from Liaoning, China. *Vertebrata PalAsiatica*, 40, 257–266.

Zhang, J. Y. (2003). First *Phareodus* (Osteoglossomorpha: Osteoglossidae) from China. *Vertebrata PalAsiatica*, 41, 327–334.

Zhang, J. Y. (2004). New fossil osteoglossomorph from Ningxia, China. *Journal of Vertebrate Paleontology*, 24, 515–524.

Zhang, J. Y. (2020). A new species of *Scleropages* (Osteoglossidae, Osteoglossomorpha) from the Eocene of Guangdong, China. *Vertebrata PalAsiatica*, 58*,* 100–119.

Zhang, J. Y., & Wilson, M. V. H. (2017). First complete fossil *Scleropages* (Osteoglossomorpha). *Vertebrata PalAsiatica*, 55*,* 1–23.

Zouhri, S. (ed.) (2017). *Paléontologie des vertébrés du Maroc : état des connaissances*. Mémoires de la Société géologique de France, 180.

CHAPTER 5

Ancient Marine Dispersals Shaped the Geographic Distribution of an Extant Lineage of Freshwater Fishes

ABSTRACT

Aim We reconstruct the biogeographic history of the oldest extant lineage of freshwater teleosts, bonytongue fishes (Osteoglossomorpha), and explore how the inclusion of fossil taxa in model-based biogeographic analyses impacts estimates of ancestral ranges.

Location Worldwide

Methods We apply phylogenetic models of biogeographic evolution (including trait-dependent dispersal models) on a Bayesian total-evidence (morphology + DNA) phylogeny of extant and extinct bonytongues to infer ancestral geographic ranges. Additionally, we test how uncertainty in topology and estimated divergence times affects biogeographic inference.

Results Extant osteoglossids likely derive from marine ancestors that dispersed circumglobally and reentered freshwater systems multiple times independently. Despite considerable ambiguity in the phylogenetic placement of several fossil taxa, these results are relatively robust to topological uncertainties. Ignoring fossil taxa results in a radically different biogeographic reconstruction that mostly corresponds to a continental vicariance scenario.

Main conclusions Fossil data can completely overthrow biogeographic patterns that are apparent from the examination of extant distributions. This study provides the first known case for a marine last common ancestor of a fish clade where all its extant members and their closest living relatives are ecologically restricted to freshwater settings.

Keywords long-distance dispersal, historical biogeography, trait-dependent dispersal models, fossils, paleobiogeography, Osteoglossomorpha, bonytongue, BioGeoBEARS

INTRODUCTION

Paleontological data have been an integral part of biogeographic research since the very early days of this discipline (Darwin, 1857; Wallace, 1876). As such, biogeographic studies of fossil taxa abound in the literature (e.g., Upchurch *et al*., 2002; Lieberman, 2003; Bauer, 2021). However, the degree to which the fossil record can overthrow biogeographic patterns that are apparent from examination of extant species alone has been contentious (Grande, 1985; Humphries & Parenti, 1986). The recent development of phylogenetic models of historical biogeography (Ree & Smith, 2008; Matzke, 2014) has led to a great increase of model-based approaches towards biogeographic inference (e.g., Landis *et al*., 2018; Klaus & Matzke, 2020; Santaquiteria *et al*., 2021). However, because these models require a fully resolved time-calibrated phylogenetic tree as input data, inclusion of fossil taxa is often impractical, either because fossils are too incomplete to be included as tips in a phylogenetic tree, or simply because a matrix of morphological characters is not readily available for a clade of interest, and would require considerable time and expertise to be assembled.

Bonytongue fishes (Osteoglossomorpha) represent an excellent target for phylogenetic models of historical biogeography: not only they are a geographically widespread clade of freshwater fishes with a long history of systematic studies and well-resolved relationships among major subgroups, but their fossil record is particularly abundant and includes several taxa known from well-preserved, articulated remains (Hilton & Lavoué, 2018; Capobianco & Friedman, 2019). Moreover, contrasting hypotheses to explain their geographic distribution abound in the literature (see Hilton & Lavoué, 2018 for a review), but remain substantially untested. The fossil record of bonytongues has two peculiarities from a biogeographic perspective. First, the geographic range of fossil taxa far exceeds the range displayed by extant forms and encompasses every continent except for Antarctica. Second, although all extant bonytongues are ecologically restricted to freshwater environments, several fossil forms are known from marine deposits of Paleogene age. Marine dispersal has been proposed as a possible explanation for the current disjunct distribution of some bonytongue sub-clades (Patterson, 1975; Bonde, 2008; Forey & Hilton, 2010; Lavoué, 2016; Capobianco & Friedman, 2019; Capobianco *et al*., 2021), but this hypothesis has never been tested within a phylogenetic framework.

Here, we estimate ancestral geographic ranges for bonytongue fishes under a newly derived total-evidence (morphology + DNA) phylogenetic hypothesis that includes nearly all extant and 32 fossil bonytongue genera, some of which are included in a phylogenetic analysis for the first time. We employ two different strategies for the biogeographic analysis: one in which fossil taxa found in marine deposits are scored as occupying their own 'marine' geographic area, the other in which an ecological binary trait (freshwater/marine) is scored separately from geographic areas and can influence rates of dispersal (trait-dependent dispersal model; Klaus & Matzke, 2020). Additionally, we explore how excluding fossil taxa from biogeographic analyses impacts estimates of ancestral ranges. Our analyses provide strong evidence for a marine origin of osteoglossid bonytongues and for marine dispersal as main cause for their current disjunct distribution. Additionally, we highlight how much the inclusion of fossil data overthrows biogeographic patterns that are apparent from examination of extant species only.

MATERIALS AND METHODS

Morphological characters and taxon sampling

A modified version of the morphological character matrix from Chapter 4 of this Dissertation was used as morphological partition for a total-evidence (morphology + DNA) phylogenetic analysis. We added the following taxa to the Chapter 4 morphological matrix:

- *Notopterus notopterus* (Notopteridae; extant, Indo-Malayan realm), with characters scored on the basis of Taverne (1978) and Hilton (2003)
- *Gymnarchus niloticus* (Gymnarchidae; extant, Afrotropical realm), with characters scored on the basis of Taverne (1972) and μCT scan of UMMZ 195003
- †*Heterosteoglossum foreyi* (Osteoglossidae; Ypresian, Denmark), based on personal observations of MGUH 28904 and on Bonde (2008)
- †*Tetoriichthys kuwajimaensis* (Osteoglossiformes *incertae sedis*; Berriasian–Hauterivian, Japan), modified from Yabumoto (2008)
- †*Notopterus primaevus* (Notopteridae; Eocene, Sumatra), based on personal observations and μCT scan of NMHUK PV P47512

- †*Phareodus* (=*Phareoides*) *queenslandicus* (Osteoglossidae; Thanetian–Ypresian, Australia), based on personal observations of FMNH PF 14254 (holotype cast) and UQ F.14960 (cast at AMNH), and on Hills (1934), Li (1994) and Taverne (2009)
- †*Musperia radiata* (Osteoglossidae; Eocene, Sumatra), based on Sanders (1934) and Taverne (1978, 2009)
- †*Taverneichthys bikanericus* (Osteoglossidae; Paleocene, India), based on Kumar *et al*. (2005) and Taverne *et al*. (2009)
- †*Thrissopterus catullii* (Osteoglossidae; Ypresian, Italy), based on personal observations of IGUP 8839–8840 and MCSNV IG 91137-91138, and on Taverne (1998)
- †*Xosteoglossid rebeccae* (?Osteoglossidae; Ypresian, Denmark), based on personal observations of MGUH 28905 and on Bonde (2008)
- †*Scleropages sinensis* (Osteoglossidae; Ypresian, China), based on Zhang & Wilson (2017)
- †*Scleropages sanshuiensis* (Osteoglossidae; Ypresian, China), based on Zhang (2020)
- †*Cretophareodus alberticus* (?Osteoglossidae; Campanian, Canada), based on Li (1996)
- †*Kuntulunia longipterus* (Osteoglossomorpha *incertae sedis*; Aptian–Albian, China), based on Zhang (1998)
- †*Asiatolepis muroii* (stem Osteoglossomorpha; Barremian–early Aptian, China), based on Zhang (2010)
- †*Tongxinichthys microdus* (stem Osteoglossomorpha; Albian, China), based on Zhang & Jin (1999)
- †*Chanopsis lombardi* (?Osteoglossidae; Aptian–Albian, Democratic Republic of Congo), based on personal observations of MRAC RG 13.608a-z[6] and MRAC RG 13.636, and on Taverne (2016)
- †*Magnigena arabica* (Osteoglossidae; Thanetian, Saudi Arabia), based on personal observations and μCT scan of NHMUK PV OE PAL 2007-1, and on Forey & Hilton (2010)
- UMMP GSP-UM field no. 1981292 (undescribed Osteoglossidae; Habib Rahi Formation, Lutetian, Pakistan), based on personal observations and μCT scan

Several of these taxa are particularly relevant because of their age, paleogeographic and paleoenvironmental setting, and potential taxonomic affinities, yet they have never been included

in a phylogenetic analysis. To make the morphological matrix compatible with the molecular dataset for a total-evidence analysis, the taxonomic resolution of extant OTUs (Operational Taxonomic Units) was changed from genus-level to species-level. In the cases where an extant genus was represented by multiple species in the molecular dataset, we assigned the morphological character scoring for that genus to the species that was examined by the original scorer of those characters (e.g., Hilton, 2003) and/or to the species that we could examine through osteological specimens or μCT data. Thus, we changed OTUs from the matrix in Chapter 4 as follows: *Campylomormyrus* → *C. tamandua*; *Chitala* → *C. chitala*; *Osteoglossum* → *O. bicirrhosum*; *Papyrocranus* → *P. afer*; *Petrocephalus* → *P. simus*; *Scleropages* → *S. formosus*. Notably, the morphological characters of this matrix are mostly invariant for congeneric extant species, because they were defined to capture morphological variation across Osteoglossomorpha with the purpose of resolving relationships between major bonytongue clades. An exception is represented by *Scleropages*, in which two characters (characters 31 and 71) differ among its three extant species. Specifically, an autopalatine bone (character 31) is present in *S. leichardti* and absent in at least *S. formosus* (unknown state in *S. jardinii*; Taverne, 1978; Hilton, 2003); and the number of hypurals (character 71) is seven in *S. leichardti* and *S. jardinii*, and six in *S. formosus* (Hilton & Britz, 2010). We included *S. leichardti* and *S. jardinii* as OTUs in the morphological matrix, with character scoring identical to *S. formosus* except for the two aforementioned characters.

The morphological matrix, which ultimately comprised 96 characters for 52 OTUs (36 extinct and 17 extant), was assembled and edited in Mesquite v. 3.61 (Maddison & Maddison, 2019).

Molecular dataset assemblage and sampling of extant taxa

The molecular data matrix was assembled through semi-automated extraction of DNA sequences from Genbank (via the NCBI platform) and BOLD (Barcode Of Life Data system), using functions from the R package 'regPhylo' (Eme *et al*., 2019). A list of extant osteoglossomorph species was compiled from FishBase (Froese & Pauly, 2020) and checked through the NCBI taxonomic database to extract their NCBI taxonomic ID. Extant outgroups (*Amia calva*, *Elops saurus*, and *Dorosoma cepedianum*) were added to this list. DNA sequences belonging to the listed species were extracted from Genbank and BOLD; after removing microsatellites and

unassigned DNA, a species-by-gene matrix was assembled to identify which genetic markers maximize taxonomic coverage. Following the criterion of maximum taxonomic coverage, 12 DNA markers were selected, 2 of which are protein-coding mitochondrial markers (*coI*, *cytb*), 2 are non-protein-coding mitochondrial regions (*12S rRNA*, *16S rRNA*), and 8 are protein-coding nuclear genes (*rag2*, *rag1*, *rhod*, *glyt*, *plagl2*, *sreb2*, *zic1*, *sh3px3*). After removal of potential outlier sequences, one sequence per species and marker was selected using the *SelBestSeq* function. Multiple alignments of these sequences were performed both with MAFFT (Katoh & Standley, 2013) and MUSCLE (Edgar, 2004). Alignment quality was assessed in MUMSA (Lassmann & Sonnhammer, 2005) by calculating the multiple overlap score of each alignment (MOS); alignments with the highest MOS for each marker were picked for the following steps. Poorly aligned positions were trimmed with GBLOCKS (Castresana, 2000). Trimmed alignments were visually inspected with AliView (Larsson, 2014) and manually edited when needed. Finally, alignments for the 12 different markers were concatenated into a single supermatrix. Species for which fewer than 4 molecular markers have been retrieved were pruned from the supermatrix and excluded from subsequent phylogenetic analyses. The final molecular dataset comprised 10080 nucleotide characters for 64 OTUs, including 30 out of 31 extant osteoglossomorph genera (with the only exception being the monotypic *Heteromormyrus*).

Total-evidence phylogenetic analysis

We combined the morphological and molecular matrices to generate a total-evidence dataset including 100 OTUs (36 extinct and 64 extant). A partitioning scheme for the molecular portion of the dataset was determined using PartitionFinder 2 (Lanfear *et al*., 2017), with greedy search algorithm and allowing for codon position-based partitions in the 10 protein-coding markers. As a result, the best partitioning scheme included 11 molecular partitions (Appendix B). The morphological portion of the dataset was treated as a separate additional partition.

Several fossil taxa included in the dataset could only be scored for a small percentage of morphological characters. These species can often behave as ‘rogue taxa’ in a phylogenetic analysis, substantially decreasing the resolution and support values of a consensus tree (Wilkinson, 1996). In order to assess the presence of rogue taxa in our dataset, we conducted a maximum likelihood (ML) phylogenetic analysis using IQ-TREE2 (Minh *et al*., 2020). A

topological constraint was employed to fix the position of outgroups, with *Amia*, *Elops*, and (*Dorosoma* + †Ellimmichthyiformes) as progressively closer outgroups to total-group Osteoglossomorpha. Best-fit models for the 12 partitions were estimated using ModelFinder Plus (Kalyaanamoorthy *et al*., 2017). Node support was evaluated with 1000 ultrafast bootstrap (UFBoot; Hoang *et al*., 2018) replicates, optimized by nearest neighbor interchanged based directly on bootstrap alignment (option -bnni; this reduces the risk of overestimating branch supports in case of severe model violations). A consensus tree from this preliminary phylogenetic analysis can be seen in Appendix D. We used RogueNaRok (Aberer *et al*., 2013) on the UFBoot replicates to identify rogue taxa that, when dropped, significantly increase the UFBoot support values in an extended majority-rule (MRE) consensus tree. The set of taxa that increased support values most included †*Chanopsis lombardi*, †*Musperia radiata* and †*Tetoriichthys kuwajimaensis*. These taxa had, respectively, 75%, 72% and 80% of missing data in the morphological partition. We considered †*Chanopsis*, †*Musperia* and †*Tetoriichthys* rogue taxa for the purpose of this study, and dropped them from subsequent analyses.

The total-evidence Bayesian phylogenetic analysis was conducted in MrBayes v. 3.2.7 (Ronquist *et al*., 2012). An Mkv model with gamma-distributed rates (four rate categories) was chosen for the morphological partition, while unlinked GTR models with gamma-distributed rates (four rate categories) were applied to each molecular partition. We forced total-group Osteoglossomorpha as a monophyletic constraint to have the remaining OTUs (*Amia*, *Elops*, *Dorosoma* and †Ellimmichthyiformes) as outgroups and to facilitate convergence. We also constrained two extant genera, *Chitala* (with 4 species included in the analysis) and *Papyrocranus* (with 2 species), to be monophyletic. These constraints serve to avoid the possibility of the fossil taxon †*Notopterus primaevus* nesting within *Chitala* or *Papyrocranus* simply as a consequence of spurious attraction between it and the only species of *Chitala* or *Papyrocranus* with morphological data. Two simultaneous analyses were run for 50 million generations, sampling every 10000 generations. Maximum standard deviation of split frequencies between the two runs reached <0.05 after 25 million generations, indicating reasonably good convergence. Effective Sample Sizes (ESS) for all estimated parameters were >500. The first 50% of sampled trees and parameters were discarded as burn-in. Posterior probabilities were visualized on a consensus majority-rule tree showing all compatible partitions. The MrBayes analysis was conducted on the CIPRES Science Gateway web server (Miller *et al*., 2011). Uncertainty in the phylogenetic

placement of fossil taxa with respect to extant taxa across the whole posterior sample of trees was visualized through the R package 'RoguePlots' (Klopfstein & Spasojevic, 2019).

A posteriori time scaling

In most cases, model-based biogeographic analyses make use of time-calibrated phylogenies (or timetrees). Bayesian methods to jointly estimate topology and divergence times of a phylogeny have become increasingly popular in the last 10 years and are now frequently used with datasets that include fossil taxa as tips, thanks to the development of complex models such as the Fossilized Birth Death (FBD) tree prior (Heath *et al*., 2014; Simões *et al.*, 2020; Mongiardino Koch *et al*., 2021). However, several features of the dataset assembled for this study suggest that Bayesian joint estimation of topology and divergence times might result in biased inference. First, there might be a strong temporal and phylogenetic bias in the fossil preservation potential in our dataset, as fossil bonytongues are temporally clustered in the Early Cretaceous and Paleogene. Most Paleogene bonytongues are osteoglossids, representing a phylogenetically biased sample. The abundance of fossil osteoglossids appears related to the origin of marine forms in this clade. Marine settings have higher preservation potential than most freshwater settings, thus representing a biased taphonomic window. Although the skyline FBD tree prior might be appropriate to model the temporal bias in fossil preservation (Simões *et al*., 2020), phylogenetic bias is not easily handled with current implementations of FBD, as the fossil preservation and recovery parameter is estimated as a single parameter for the whole tree within each temporal slice. Second, 'tip-dating' Bayesian phylogenetic analyses tend to retrieve trees with higher stratigraphic fit than parsimony or undated Bayesian analyses (King, 2020). While this is often a desirable feature of tip-dating methods, it can result in topologies where fossil taxa of similar age are spuriously grouped together when they are scored for a small percentage of morphological characters, or when the fossil sampling is highly uneven through time (Turner *et al*., 2017; King, 2020)—two conditions met by our dataset. Third, the morphological matrix employed for this study was originally designed to capture morphological variation across Osteoglossomorpha with the purpose of resolving relationships between major bonytongue clades. As a result, it does not capture well the morphological variation between closely related extant taxa, especially in the Notopteroidei (Notopteridae + *Gymnarchus* + Mormyridae) portion

of the tree. Applying a 'clock-like' model to this dataset would result in apparent morphological stasis after the divergence of major extant lineages, thus potentially biasing divergence time estimates. A massive genus and species-level osteological revision of Notopteroidei will be needed to overcome this issue and derive more accurate divergence time estimates.

Instead of an approach that jointly estimates topology and divergence, we opted for an *a posteriori* time scaling (APT) approach that does not use branch length information to estimate a timetree: the Lloyd *et al*. (2016) algorithm based on the node-dating approach of Hedman (2010). This method uses the age distribution of successive fossil outgroups to derive probability distribution for node ages. Because not every node can be appropriately dated this way (see Lloyd *et al*., 2016), the algorithm obtains missing dates via a randomization process. Non-osteoglossomorph tips (*Amia*, *Elops*, *Dorosoma*, and †Ellimmichthyiformes) were pruned from all trees before APT. Tip ages of extant taxa were fixed to zero, whereas tip ages of extinct taxa were randomly drawn from a uniform distribution bound by the ages of chronostratigraphic boundaries with stage-level resolution. Minimum ages of successive outgroups to the study group have to be provided to the Hedman APT algorithm in order to calibrate the root node. We used the phylogenetic hypothesis of Bean & Arratia (2020) to identify successive outgroups to Osteoglossomorpha. When the oldest known fossil of an earlier-diverging outgroup is younger than the oldest fossil in its sister clade, the minimum age of that outgroup is the age of the oldest fossil of its sister clade. These are the successive outgroups to Osteoglossomorpha, with oldest known fossils and oldest minimum age for that outgroup as used in the Hedman APT algorithm:

- Clupeocephala: †*Tischlingerichthys viohli*, late Tithonian (Arratia, 2000; min age 143.1 Ma)
- †Orthogonikleithridae: †*Leptolepides* spp., late Kimmeridgian (Konwert, 2016; min age 149.2 Ma)
- †Luisiellidae: †*Luisiella feruglioi*, Oxfordian–Tithonian (Sferco *et al*., 2015; min age 149.2 Ma)
- Elopomorpha: †*Anaethalion* spp., Kimmeridgian (Poyato-Ariza, 1999; min age 149.2 Ma)
- †Varasichthyidae: †*Varasichthys ariasi*, Oxfordian (Arratia, 2008; min age 154.8 Ma)

- †Ascalaboidae: †*Ebertichthys ettlingensis*, late Kimmeridgian (Arratia, 2016; min age 154.8 Ma)
- †*Leptolepis coryphaenoides*, early Toarcian, †*Harpoceras falciferum* Ammonite Zone (Konwert & Stumpf, 2017; min age 181.3 Ma)
- †*Dorsetichthys bechei*, Sinemurian, *semicostatum* to *obtusum* Ammonite Zones (Hart *et al*., 2020; min age 197.2 Ma)
- †Ankylophoridae: †*Steurbautichthys aequatorialis*, Middle Jurassic (Taverne, 2011; Arratia, 2013; min age 197.2 Ma)
- †Pholidophoriformes: †*Pholidophoretes salvus*, Carnian, Julian stage (Arratia, 2013; min age 233.6 Ma)
- †*Prohalecites porroi*, Ladinian-Carnian boundary (Arratia & Tintori, 1999; min age 237 Ma)

To better constrain node ages, we manually grafted phylogenetic tips representing the oldest known fragmentary fossils that can be attributed to an extant osteoglossomorph genus. These include: *Gymnarchus* (teeth, late Bartonian; Otero *et al*., 2015), *Hyperopisus* (parasphenoid, Messinian–early Zanclean; Van Neer, 1994), *Heterotis* (squamules, Rupelian; Otero & Gayet, 2001), and *Arapaima* (basioccipital complex, Serravallian; Lundberg & Chernoff, 1992). These manually grafted tips were pruned from trees after APT, and thus removed from downstream biogeographic analyses.

The Hedman APT algorithm was applied to the Bayesian consensus majority rule tree, with 1000 dates estimated for each node. The resulting time-calibrated consensus tree with mean node ages was used in downstream biogeographic analyses. A time-calibrated consensus tree with minimum node ages and one with maximum node ages were also used to assess the sensitivity of our analyses to inferred divergence times. To evaluate sensitivity of downstream biogeographic analyses to phylogenetic uncertainty, we also applied the Hedman APT algorithm to a random sample of 50 trees with different topologies from the Bayesian posterior distribution.

All APT were performed in R. Scripts can be found in Appendix F.

Biogeographic analyses

Two different strategies were employed to define geographic areas for the biogeographic analysis. In both, continental land masses were divided into 7 areas encompassing the whole distribution of extant and extinct Osteoglossomorpha and corresponding to major biogeographical regions for extant freshwater fishes (Leroy *et al*., 2019): Nearctic, Neotropical, Ethiopian, Palearctic, Sinean, Indo-Malayan (or Oriental), and Australian. The maximum number of areas that could be occupied by a single lineage at any one point was fixed to 3, to reduce the number of allowed geographic states and reduce computational time. The difference between the two strategies pertains to how fossil taxa from marine deposits were treated. In the first strategy ('MarineAsArea'), we considered the marine realm as an additional geographic area (bringing the total to 8 areas), and fossil taxa found in marine deposits were scored as occurring exclusively in the marine area. In the second strategy ('MarineAsTrait'), we scored marine fossil taxa as occurring in the continental geographic areas where their fossils have been found. Instead of being treated as a geographic area, marine environment was included as a possible state of a binary ecological trait, where state 1 indicates freshwater or brackish water ecology and state 2 indicates marine ecology. For this binary trait, extinct taxa were scored according to the depositional setting where their fossils have been found. The 'MarineAsTrait' strategy for area coding was used to evaluate the impact of trait-dependent dispersal models on biogeographic inference when fossil taxa reveal traits and geographic distributions that fall outside the present-day range displayed by the examined group.

We did not apply any time-stratified biogeographic model (e.g., models where area connectivity changes between time slices as a consequence of continental rearrangement or sea-level changes) because the confidence intervals on node dates derived from the Hedman APT are too broad and encompass several time slices with different continental arrangements to allow for such a model to be meaningful.

All biogeographic analyses were performed in the R package 'BioGeoBEARS' (Matzke, 2014).

'MarineAsArea' strategy

For the 'MarineAsArea' strategy, we applied the standard biogeographic models implemented in BioGeoBEARS. These include DEC (Dispersal-Extinction-Cladogenesis; Ree & Smith, 2008),

DIVALIKE (a likelihood interpretation of the parsimony DIVA, DIspersal Vicariance Analysis, model; Ronquist, 1997), and BAYAREALIKE (a simplified likelihood interpretation of the Bayesian model implemented in the software 'BayArea'; Landis *et al*., 2013). Additionally, we ran variants of these three models that include a 'jump dispersal' parameter, *j*, which allows for founder-event jump dispersal during cladogenesis (see Ree & Sanmartín, 2018 for a critique of the DEC+*j* model, and Klaus & Matzke, 2020 for a partial response to that critique). Standard tools for statistical model comparison (Burnham & Anderson, 2002) – including likelihood ratio test for pairs of nested models and Akaike Information Criterion for small sample sizes (AICc) – were used to evaluate model support.

To test how different divergence time estimates impact the results of the biogeographic analysis, we applied the best-fitting biogeographic model to the APT consensus tree with minimum node ages and to the APT consensus tree with maximum node ages. To test how phylogenetic uncertainty affects biogeographic inference, we applied the best-fitting model to 50 phylogenies with different topologies from the Bayesian posterior distribution, scaled with Hedman APT. The results from these 50 analyses were summarized by recording marginal ancestral area reconstructions for 8 clades: total-group Osteoglossomorpha (root node), crown Osteoglossomorpha (*Hiodon alosoides* + *Osteoglossum bicirrhosum*), crown Osteoglossiformes (*Pantodon buccholzi* + *Osteoglossum bicirrhosum* node), crown Osteoglossidae (*Osteoglossum bicirrhosum* + *Arapaima gigas* node), crown Osteoglossinae (*Osteoglossum bicirrhosum* + *Scleropages leichardti* node), crown Arapaiminae (*Arapaima gigas* + *Heterotis niloticus* node), crown Notopteroidei (*Notopterus notopterus* + *Mormyrus ovis* node), and crown Notopteridae (*Notopterus notopterus* + *Papyrocranus afer* node). For these clades, we tabulated the most likely marginal ancestral state across the 50 phylogenies, and we calculated average marginal probabilities for each possible state (these correspond to empirical Bayesian posterior probabilities).

To explore how the inclusion of fossil data impacts biogeographic inference, we ran the standard BioGeoBEARS models listed above on the Bayesian consensus tree pruned of all fossil taxa. We compared marginal ancestral states found in this extant-only analysis with marginal ancestral states recovered from the previous integration of 50 posterior phylogenies with fossil taxa.

'MarineAsTrait' strategy

For the 'MarineAsTrait' strategy, we used a recently developed model variant that allows a binary trait to affect dispersal ability, and in turn allows geographic data to affect inference of the evolution of that trait (Klaus & Matzke, 2020). These trait-dependent dispersal models can be used to test the correlation between a particular trait and dispersal probability by comparing the fit of a model with trait-dependent dispersal to the fit of a similar model but with dispersal probability independent from the trait. Because we would expect marine fishes to have a higher probability of dispersing between continents than freshwater fishes, this type of model is particularly appealing for the osteoglossomorph case, as dispersal of extinct marine forms followed by transition to freshwater settings might be one of the causes for the current disjunct distribution of the group. Trait-dependent dispersal models add three parameters to the standard BioGeoBEARS models mentioned above. Two parameters (*t12* and *t21*) represent the rate of transition of the binary trait from state 1 to state 2 and from state 2 to state 1 (that is, the rate of transition from freshwater to marine ecology, and vice versa). The third parameter (*m1*) is a multiplier on the rate of dispersal when a lineage is in state 1 (freshwater). The dispersal multiplier *m2* when a lineage is in state 2 (marine) was fixed to 1, such that the estimated *m1* represents the probability ratio of dispersal when in state 1 relative to when in state 2. For calculation purposes, it is recommended to fix the multiplier corresponding to the state that *a priori* is thought to be correlated with higher dispersal probability (in this case, the marine dispersal multiplier). We constrained the range of possible values that *m1* could assume between 0.001 and 5, such that dispersal probability for freshwater forms could never be zero and could potentially be higher than for marine forms. The dispersal multiplier is applied only to the dispersal parameter *d* in DEC, DIVALIKE and BAYAREALIKE models, while it is applied to both *d* and the jump dispersal parameter *j* in DEC+*j*, DIVALIKE+*j* and BAYAREALIKE+*j* models.

To compare trait-dependent dispersal models to trait-independent models, we applied the six standard BioGeoBEARS models (DEC(+*j*), DIVALIKE(+*j*) and BAYAREALIKE(+*j*)) to the geographic data without binary trait, and independently estimated best-fit parameters for a 2-rate (*t12* and *t21*) model of discrete character evolution using the binary trait data. The log-likelihood of the geographic data and the log-likelihood of the trait data were added together to obtain the

log-likelihood of combined data under trait-independent dispersal models. These were compared to log-likelihoods of combined data under trait-based dispersal models (DEC(+*j*)+*t12*+*t21*+*m1*, DIVALIKE(+*j*) +*t12*+*t21*+*m1* and BAYAREALIKE(+*j*) +*t12*+*t21*+*m1*) by calculating AICc weights and by likelihood ratio tests for pairs of nested models.

The script for the BioGeoBEARS analysis under the 'MarineAsTrait' strategy can be found in Appendix F. Scripts for other BioGeoBEARS analyses performed in this study can be requested from the author of this Dissertation.

Institutional abbreviations

FMNH, The Field Museum, Chicago, IL, USA; NHMUK, The Natural History Museum, London, UK; UMMP, University of Michigan Museum of Paleontology, Ann Arbor, MI, USA; UMMZ, University of Michigan Museum of Zoology, Ann Arbor, MI, USA.

Dagger symbols

Following the convention of Patterson & Rosen (1977), the dagger symbol (†) precedes extinct taxa.

RESULTS

Phylogeny

Most major phylogenetic relationships recovered in the Bayesian total-evidence (morphology + DNA) phylogenetic analysis (Fig. 5.1) are compatible with previous morphological and molecular studies (e.g., Lavoué & Sullivan, 2004; Wilson & Murray, 2008; Inoue *et al*., 2009; Lavoué *et al*., 2011, 2012; Chapter 4 of this Dissertation; see Hilton & Lavoué, 2018 for a review of osteoglossomorph systematics). These include mooneyes (Hiodontidae) as living sister group to all other extant bonytongues (Osteoglossiformes); elephantfishes (Mormyridae) and the aba (Gymnarchidae) as closely related to Old World knifefishes (Notopteridae); and arapaimas and relatives (Arapaiminae) being closely related to arowanas (Osteoglossinae) and forming the clade Osteoglossidae. The African butterflyfish *Pantodon* was recovered as living sister group to all other extant Osteoglossiformes, a position that is not supported by morphological characters alone (Hilton, 2003; Wilson & Murray, 2008; Chapter 4 of this Dissertation) but which is often

found in molecular phylogenetic analyses (Lavoué *et al*., 2012; Betancur-R *et al*., 2017). Intergeneric relationships within the species-rich Mormyridae are the same as in Sullivan *et al*. (2016), with *Petrocephalus* sister to all other elephantfishes (Mormyrinae), and the recognition of two major large clades within Mormyrinae: one including *Brevimyrus*, *Hyperopisus*, *Hippopotamyrus*, *Marcusenius*, *Cyphomyrus*, *Genyomyrus*, *Gnathonemus*, and *Campylomormyrus*; the other including *Pollimyrus*, *Stomatorhinus*, *Cryptomyrus*, *Boulengeromyrus*, *Ivindomyrus*, and *Paramormyrops*. An unexpected result involves the interrelationships of extant Osteoglossinae, with the genus *Scleropages* being resolved as paraphyletic with respect to *Osteoglossum*. Specifically, the Southeast Asian *S. formosus* is recovered as more closely related to both South American species of *Osteoglossum* (*O. bicirrhosum* and *O. ferreirai*) than to other species of *Scleropages* (the Australian *S. leichardti* and the Australian and New Guinean *S. jardinii*).

The posterior probabilities of several nodes forming the 'backbone' of the osteoglossomorph tree are extremely low due to the uncertain position of several fossil taxa (Fig. 5.1). However, even when posterior probabilities of nodes including fossil taxa are very low, those taxa might consistently resolve in few distinct positions in respect to extant taxa (see Klopfstein & Spasojevic, 2019). Hence, exploring where fossil taxa fall with respect to extant ones when considering the posterior distribution of phylogenies (especially when the relationships between extant lineages are well supported) can be more illuminating than just examining node supports.

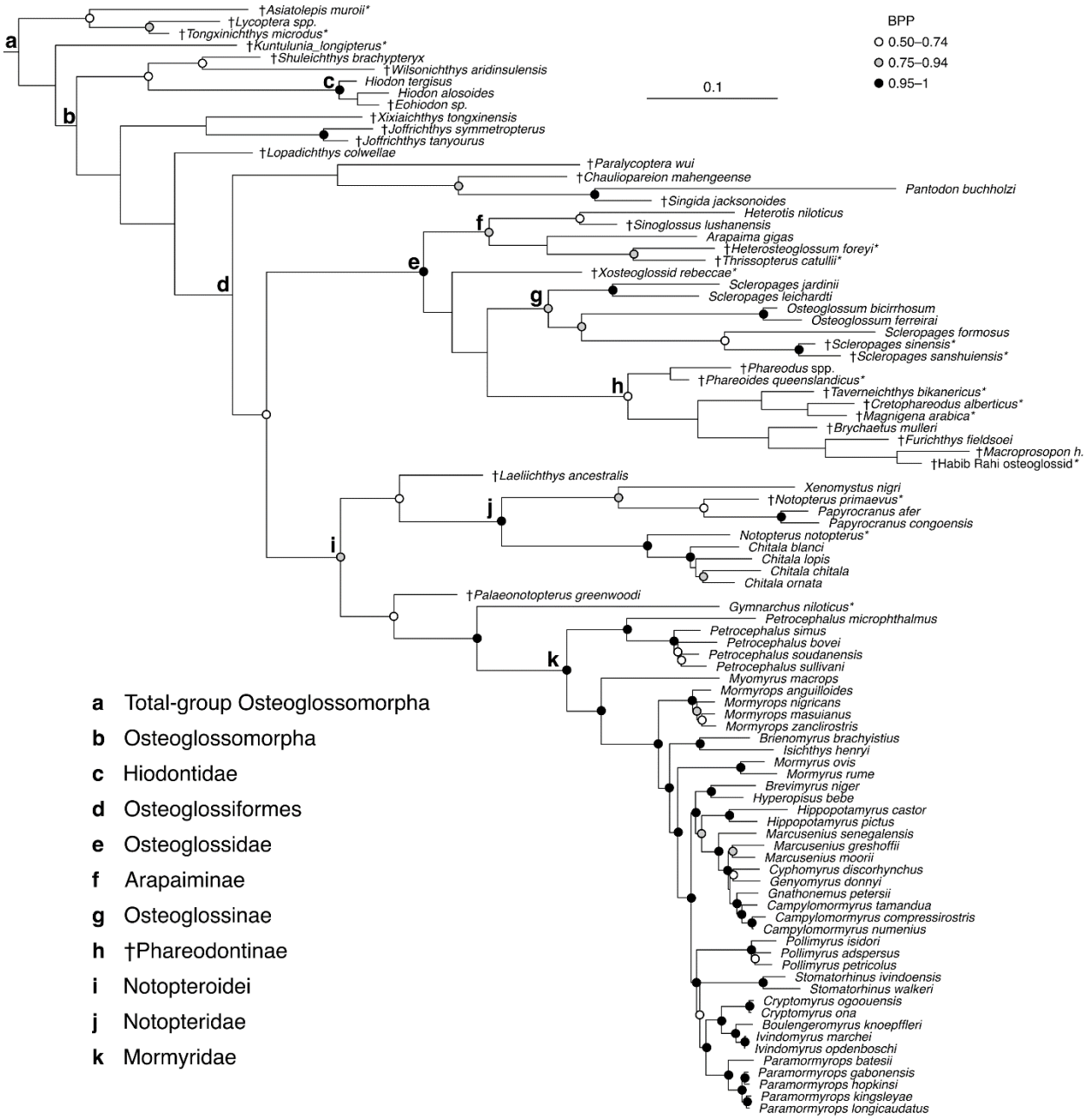

Fig. 5.1. Bayesian consensus tree showing phylogenetic relationships of Osteoglossomorpha. Bayesian posterior probabilities (BPP) for nodes are indicated by shaded circles. Nodes without a right semicircle have less than 0.50 BPP. Asterisks indicate taxa with new or revised scoring of morphological characters compared to the phylogenetic analysis in Chapter 4 of this Dissertation. Scale bar indicates 0.1 changes per site (averaged over all partitions).

An ensemble of Early Cretaceous Chinese taxa that can be grouped in the extinct family †Lycopteridae (†*Asiatolepis*, †*Lycoptera*, and †*Tongxinichthys*) are reconstructed as either stem

Osteoglossomorpha or as early-diverging Hiodontiformes (total-group Hiodontidae). Another Early Cretaceous Chinese taxon, †*Kuntulunia*, can alternatively be placed as a stem Osteoglossomorpha, Hiodontiformes or stem Osteoglossiformes. The Early Cretaceous †*Shuleichthys* and the Late Cretaceous †*Wilsonichthys* are always recovered within Hiodontiformes. The early Eocene †*Eohiodon* and extant *Hiodon* species form a strongly supported clade. In fact, †*Eohiodon* has been synonymized with *Hiodon* by some authors (Hilton & Grande, 2008). †*Xixiaichthys* and the two species of †*Joffrichthys* are most likely stem Osteoglossiformes or members of Hiodontiformes (with <0.1 probability of being either stem Osteoglossomorpha or stem Pantodontidae). The deep-bodied †*Lopadichthys* from the Paleocene of North America is one of the most poorly resolved taxon in our phylogeny, with possible placements including stem Osteoglossiformes and several basal branches within crown Osteoglossiformes (with <0.1 probability of being either a stem Osteoglossomorpha or a stem Notopteroidei). The Early Cretaceous †*Paralycoptera* from China is recovered as a crown member of Osteoglossomorpha, either on the osteoglossiform stem or as a close relative of *Pantodon* or as a member of Hiodontiformes. The Early Cretaceous South American †*Laeliichthys* is either a stem notopterid, or sister to Notopteridae + Mormyridae. The Eocene †*Notopterus primaevus* from Sumatra is always reconstructed as a notopterid, and most often as being more closely aligned to extant African knifefishes (*Papyrocranus* and *Xenomystus*) than to extant Asian knifefishes (*Notopterus* and *Chitala*). The mid-Cretaceous †*Palaeonotopterus* from Morocco represents most likely a stem member of the Mormyridae + *Gymnarchus* clade; however, in more than 20% of posterior trees it is reconstructed in more rootward positions (including as a stem osteoglossomorph). This result is likely due to the low percentage of morphological characters scored for †*Palaeonotopterus* (76% of characters could not be scored), which reflects that little is known of this taxon outside of its braincase anatomy.

All marine taxa included in this analysis (†*Brychaetus*, †*Furichthys*, †*Heterosteoglossum*, †*Macroprosopon*, †*Magnigena*, †*Thrissopterus*, †*Xosteoglossid* and the undescribed Habib Rahi taxon) are recovered as either crown or stem members of Osteoglossidae. Some of them are often grouped together with extinct freshwater taxa from various continents (†*Phareodus*, †*Phareoides*, †*Taverneichthys* and †*Cretophareodus*), comprising †Phareodontinae (*sensu* Chapter 4 of this Dissertation). Among these, the Late Cretaceous †*Cretophareodus* can also occupy a position outside of †Phareodontinae as a stem osteoglossid. The position of

†Phareodontinae within Osteoglossidae is not well resolved, although a close relationship with Osteoglossinae is more favored than other placements. The marine taxa †*Heterosteoglossum* and †*Thrissopterus* often form a clade to the exclusion of other osteoglossids (0.8 posterior probability) and are consistently recovered as either stem or crown members of Arapaiminae. †*Sinoglossus* from the late Eocene–Oligocene of China is either reconstructed as sister to the African *Heterotis*, or as sister to the South American *Arapaima*. The Eocene Chinese species of *Scleropages* (†*S. sinensis* and †*S. sanshuiensis*) resolve as crown Osteoglossinae, most often closely related to the Southeast Asian extant representative of the genus, *S. formosus*.

Biogeographic analysis

'MarineAsArea' strategy

The best-fitting biogeographic model for the geographic data in the 'MarineAsArea' strategy is the DEC+*j* model (Figs. 5.2-5.3, Table 5.1), with AICc model weight of 67%, followed by the DIVALIKE+*j* (AICc weight: 19%) and the BAYAREALIKE+*j* models (AICc weight: 13%). Models that do not include the jump dispersal parameter *j* are strongly rejected by the AICc and by the likelihood ratio test when compared to their +*j* counterparts (Table 5.1), suggesting that long-distance dispersal played a significant role in the biogeographic history of bonytongues. Marginal ancestral range probabilities are very similar across the three best models (see Appendix G), with the only major difference being at the root node, reconstructed as Nearctic + Sinean in the DEC+*j* model and as Sinean in the DIVALIKE+*j* and BAYAREALIKE+*j* models. Applying the DEC+*j* model to the trees with minimum and maximum node ages resulted in no significant differences in the model parameter estimates and in the marginal ancestral ranges (Appendix G), suggesting that differences in node age have little impact – if any – to biogeographic models that are not time-stratified.

Phylogenetic uncertainty has a considerable impact on the estimate of ancestral ranges, as shown by the results of the DEC+*j* analysis applied to 50 different phylogenies (Fig. 5.2, Tables 5.2-5.3). However, some strong patterns emerge even when taking this uncertainty into account. The most basal nodes of osteoglossomorph phylogeny are reconstructed as Nearctic + Sinean + Ethiopian, or as subsets of this broad geographic range. The ancestral range for crown

Osteoglossiformes is estimated as either Ethiopian (marginal probability = 0.65) or Nearctic (marginal probability = 0.12). Crown osteoglossids are reconstructed as ancestrally marine with marginal probability = 0.62, and in 84% of phylogenies from the posterior sample the marine state is the most likely for this node. Alternatively, the freshwater Ethiopian realm has marginal probability = 0.19 to be the ancestral area for crown osteoglossids. Arapaiminae is estimated as being ancestrally either marine, Ethiopian, or Neotropical. Osteoglossinae is reconstructed as ancestrally Australian with a relatively low marginal probability of 0.47, but this was the most likely ancestral range in 84% of topologies. Finally, the Ethiopian realm is the most likely ancestral area for both Notopteroidei (Notopteridae + Mormyridae + *Gymnarchus*) and Notopteridae. However, Notopteridae are reconstructed as ancestrally Indo-Malayan in 20% of phylogenies considered here (marginal probability = 0.18).

'MarineAsArea' strategy	**lnL**	**np**	***d***	***e***	***j***	**AICc weight**	**adding *j* P-value**
DEC	-267	2	0.01	0.01	0	1.70E-75	
DEC+*j*	-94.15	3	1.00E-12	1.00E-12	0.02	0.67	3.80E-77
DIVALIKE	-130.7	2	0.0007	0.0002	0	2.70E-16	
DIVALIKE+*j*	-95.39	3	1.00E-12	1.00E-12	0.02	0.19	4.40E-17
BAYAREALIKE	-261.2	2	0.01	0.01	0	5.70E-73	
BAYAREALIKE+*j*	-95.78	3	1.00E-12	1.00E-12	0.02	0.13	6.50E-74

Table 5.1. Biogeographic model comparison with the 'MarineAsArea' strategy. lnL = log-likelihood; np = number of parameters.

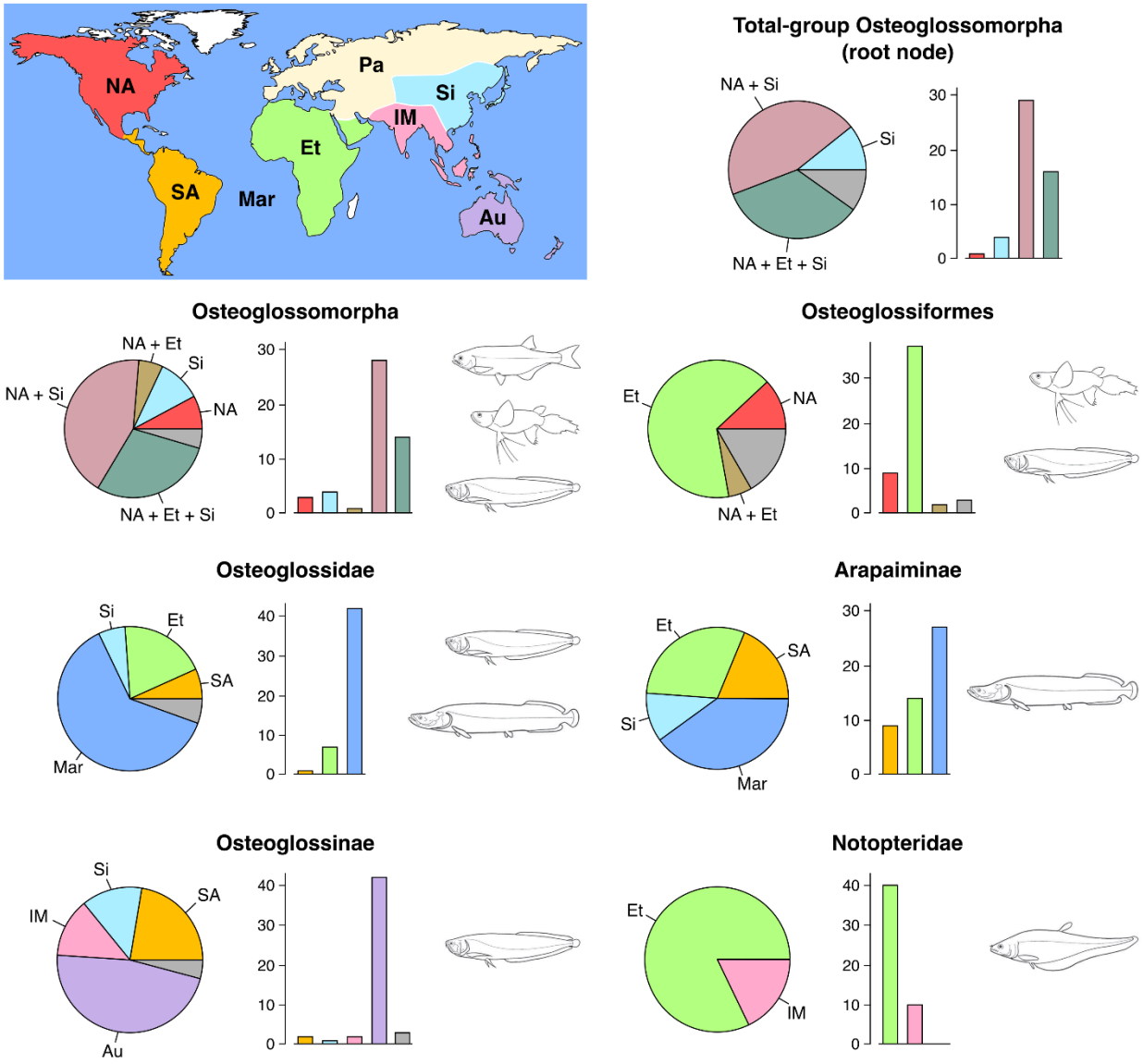

Fig. 5.2. Average marginal probabilities (*pie charts*) and most-likely ancestral ranges (*bar plots*) for key nodes on the bonytongue phylogeny under the best-fitting biogeographic model (DEC+*j*) with the 'MarineAsArea' strategy, calculated across a sample of 50 phylogenies with different topologies from the Bayesian posterior. Ancestral ranges with <0.05 marginal probability were collapsed into an unlabeled category (in grey). Abbreviations for biogeographic regions: NA = Nearctic; SA = Neotropical; Et = Ethiopian; Pa = Palearctic; Si = Sinean; IM = Indo-Malayan; Au = Australian; Mar = marine. Line drawings of various extant osteoglossomorphs used here and in Fig. 5.3 are taken from Nelson *et al.* (2016).

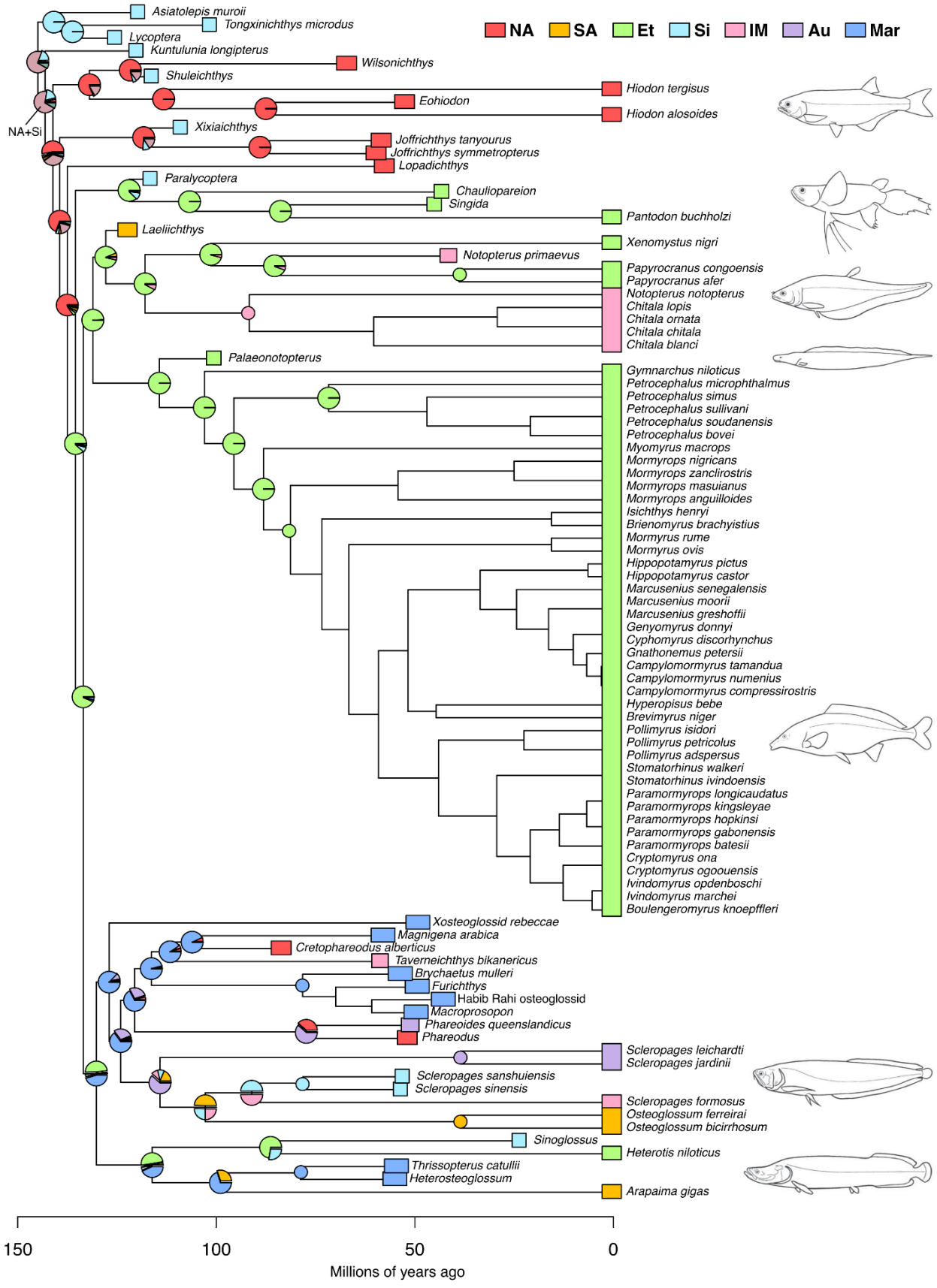
NA
SA
Et
Si
IM
Au
Mar
NA+Si
Asiatolepis muroii
Tongxinichthys microdus
Lycoptera
Kuntulunia longipterus
Wilsonichthys
Shuleichthys
Hiodon tergisus
Eohiodon
Hiodon alosoides
Xixiaichthys
Joffrichthys tanyourus
Joffrichthys symmetropterus
Lopadichthys
Paralycoptera
Chaulioparelon
Singida
Pantodon buchholzi
Laeliichthys
Xenomystus nigri
Notopterus primaevus
Papyrocranus congoensis
Papyrocranus afer
Notopterus notopterus
Chitala lopis
Chitala ornata
Chitala chitala
Chitala blanci
Palaeonotopterus
Gymnarchus niloticus
Petrocephalus microphthalmus
Petrocephalus simus
Petrocephalus sullivani
Petrocephalus soudanensis
Petrocephalus bovei
Myomyrus macrops
Mormyrops nigricans
Mormyrops zanclirostris
Mormyrops masuianus
Mormyrops anguilloides
Isichthys henryi
Brienomyrus brachyistius
Mormyrus rume
Mormyrus ovis
Hippopotamyrus pictus
Hippopotamyrus castor
Marcusenius senegalensis
Marcusenius moorii
Marcusenius greshoffii
Genyomyrus donnyi
Cyphomyrus discorhynchus
Gnathonemus petersii
Campylomormyrus tamandua
Campylomormyrus numenius
Campylomormyrus compressirostris
Hyperopisus bebe
Brevimyrus niger
Pollimyrus isidori
Pollimyrus petricolus
Pollimyrus adspersus
Stomatorhinus walkeri
Stomatorhinus ivindoensis
Paramormyrops longicaudatus
Paramormyrops kingsleyae
Paramormyrops hopkinsi
Paramormyrops gabonensis
Paramormyrops batesii
Cryptomyrus ona
Cryptomyrus ogoouensis
Ivindomyrus opdenboschi
Ivindomyrus marchei
Boulengeromyrus knoepffleri
Xosteoglossid rebeccae
Magnigena arabica
Cretophareodus alberticus
Taverneichthys bikanericus
Brychaetus mulleri
Furichthys
Habib Rahi osteoglossid
Macroprosopon
Phareoides queenslandicus
Phareodus
Scleropages leichardti
Scleropages jardinii
Scleropages sanshuiensis
Scleropages sinensis
Scleropages formosus
Osteoglossum ferreirai
Osteoglossum bicirrhosum
Sinoglossus
Heterotis niloticus
Thrissopterus catullii
Heterosteoglossum
Arapaima gigas
150
100
50
0
Millions of years ago

Fig. 5.3 (*previous page*). Ancestral geographic range estimates under the best-fitting model (DEC+*j)* with the 'MarineAsArea' strategy, applied to the Bayesian consensus tree. See Fig. 5.2 for abbreviations of geographic areas. Pie charts for nodes deriving from unambiguously reconstructed nodes are not displayed.

'MarineAsArea' strategy, DEC+*j* model	**NA**	**SA**	**Et**	**Si**	**IM**	**Au**	**Mar**	**NA + Et**	**NA + Si**	**NA + Et + Si**	**Other**
Total-group Osteoglossomorpha	1	0	0	4	0	0	0	0	29	16	0
Osteoglossomorpha	3	0	0	4	0	0	0	1	28	14	0
Osteoglossiformes	9	0	37	0	0	0	0	1	0	2	1
Osteoglossidae	0	0	7	1	0	0	42	0	0	0	0
Arapaiminae	0	9	14	0	0	0	27	0	0	0	0
Osteoglossinae	0	2	0	1	2	42	3	0	0	0	0
Notopteroidei	0	0	50	0	0	0	0	0	0	0	0
Notopteridae	0	0	40	0	10	0	0	0	0	0	0

Table 5.2. Most likely ancestral ranges for key nodes on the bonytongue phylogeny. Numbers indicate in how many (out of 50) distinct phylogenetic hypotheses a certain geographic range was the most likely. Highest scores for each node are highlighted. See Fig. 5.2 for abbreviations of geographic areas. Clade names refer to crown clades unless specified otherwise.

'MarineAsArea' strategy, DEC+*j* model	**NA**	**SA**	**Et**	**Si**	**IM**	**Au**	**Mar**	**NA + Et**	**NA + Si**	**NA + Et + Si**	**Other**
Total-group Osteoglossomorpha	0.022	0	0.001	0.107	0	0	0	0.028	0.451	0.343	0.048
Osteoglossomorpha	0.078	0	0.004	0.102	0	0	0	0.057	0.427	0.29	0.042
Osteoglossiformes	0.12	0.012	0.658	0.044	0.002	0.001	0.011	0.055	0.005	0.029	0.063
Osteoglossidae	0.005	0.067	0.194	0.06	0.012	0.036	0.624	0	0	0	0.002
Arapaiminae	0	0.188	0.301	0.111	0	0	0.399	0	0	0.001	0
Osteoglossinae	0	0.224	0	0.135	0.131	0.469	0.041	0	0	0	0
Notopteroidei	0	0.016	0.961	0	0.021	0	0	0	0	0	0.002
Notopteridae	0	0	0.821	0	0.178	0	0	0	0	0	0.001

Table 5.3. Average marginal probabilities (or empirical Bayesian posterior probabilities) for ancestral geographic ranges at key nodes on the bonytongue phylogeny. Highest probabilities for each node are highlighted. Probabilities <0.001 are tabulated as 0. See Fig. 5.2 for abbreviations of geographic areas. Clade names refer to crown clades unless specified otherwise.

After pruning fossil taxa from the analysis, the best-fitting biogeographic model is still the DEC+*j* model, with AICc of 63%. The only other viable model is the DIVALIKE+*j*, with AICc of 37%. Contrary to the analysis including fossil taxa, the BAYAREALIKE+*j* model had an

AICc weight several orders of magnitude lower than the two best models (Appendix G). The jump dispersal parameter *j* was estimated to be around 4 times lower in the DEC+j analysis without fossils than in the same model with fossil (0.0048 vs 0.02). More importantly, reconstructed ancestral ranges differ significantly when ignoring fossil taxa (Table 5.4, Fig. 5.4). The absence of any marine species and of continental Chinese fossils means that geographic ranges including the Sinean region or the marine realm are extremely unlikely to be found as probable ancestral ranges. Another effect of the exclusion of fossil taxa is that the Neotropical region is often included in reconstructed ancestral ranges at key nodes, which contrasts strikingly with the results of the analysis when including fossils.

'MarineAsArea' strategy, DEC+*j* model excluding fossils	**Most likely ancestral range (marginal probability)**
Osteoglossomorpha	NA + SA + Et (0.753)
Osteoglossiformes	SA + Et (0.761)
Osteoglossidae	SA + Et (0.658)
Arapaiminae	SA + Et (0.642)
Osteoglossinae	SA (0.820)
Notopteroidei	Et (0.977)
Notopteridae	Et (0.972)

Table 5.4. Most likely ancestral ranges at key nodes in the bonytongue phylogeny under the best-fitting model (DEC+*j*) when excluding fossil taxa from the analysis. See Fig. 5.2 for abbreviations of geographic areas.

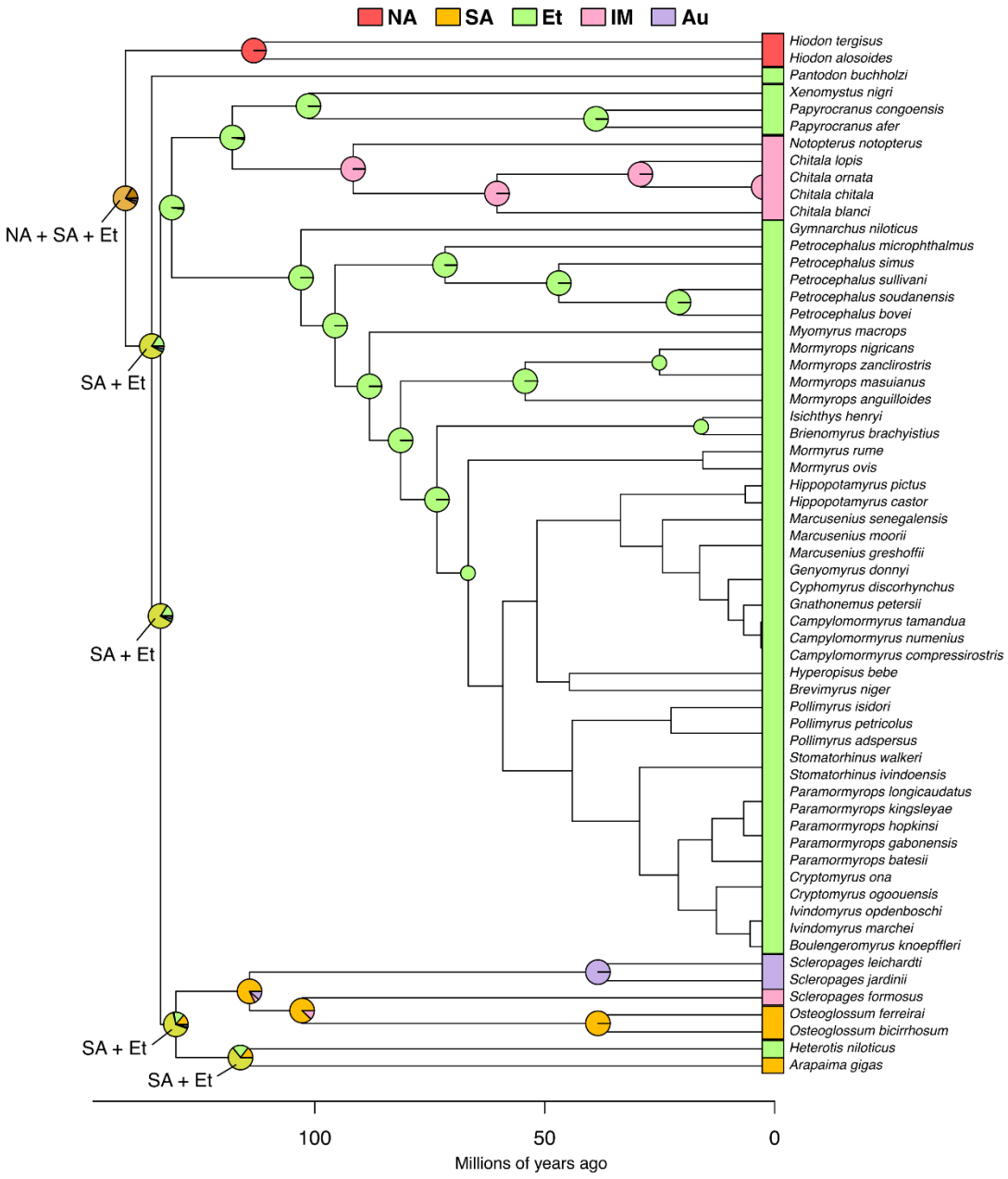

Fig. 5.4. Ancestral geographic range estimates under the best-fitting model (DEC+*j)* when fossil taxa are excluded from the analysis. See Fig. 5.2 for abbreviations of geographic areas. Pie charts for nodes deriving from unambiguously reconstructed nodes are not displayed.

'MarineAsTrait' strategy

The best-fitting biogeographic model for the geographic data in the 'MarineAsTrait' strategy is the DEC+*j+t12+t21+m1* model (Fig. 5.5), with AICc weight of 70%, followed by the

DIVALIKE+*j*+*t12*+*t21*+*m1* and BAYAREALIKE+*j*+*t12*+*t21*+*m1* models, with AICc weights of 16% and 14%, respectively. Trait-dependent dispersal models were always favored when compared to their trait-independent analogues, suggesting that taking into account freshwater-to-marine transitions (and vice versa) is fundamental to accurately reconstruct the biogeographic history of Osteoglossomorpha (Table 5.5). Moreover, as in the 'MarineAsArea' strategy, models that did not include the jump dispersal parameter *j* were strongly rejected. The DIVALIKE+*j*+*t12*+*t21*+*m1* model yields ancestral range estimates comparable with the best-fitting DEC+*j*+*t12*+*t21*+*m1*, except for the tree root and for crown Osteoglossiformes, which are reconstructed as occurring respectively in the Sinean and Nearctic realm rather than in a combined range of both. Instead, the BAYAREALIKE+*j*+*t12*+*t21*+*m1* results in more pronounced differences compared to the best-fitting model, such as crown Notopteridae being estimated as Indo-Malayan instead of Ethiopic, and crown Osteoglossidae as Palearctic instead of Ethiopic. More interestingly, ancestral state reconstruction for the ecological binary trait (freshwater/marine) differs majorly between trait-independent dispersal and trait-dependent dispersal models (Fig. 5.6). In the first class of models, where trait evolution is estimated independently from geography, the rate of transition from freshwater to marine environments (*t12*) is 100 times higher than the rate of transition from marine to freshwater (*t21*). As a consequence, these models reconstruct four independent invasions of marine settings by bonytongues, and no return to freshwater by marine taxa. Conversely, the best-fitting DEC+*j*+*t12*+*t21*+*m1* model, in which trait evolution is co-estimated with biogeographic evolution, strongly favors a scenario with a single invasion of marine settings and multiple (up to 11) independent returns to freshwater environments by marine taxa. In fact, all trait-dependent dispersal models estimate a rate of transition from marine to freshwater (*t21*) higher than the opposite transition (37 times higher under DEC+*j*+*t12*+*t21*+*m1*). The maximum likelihood estimate of the dispersal multiplier *m1* under the best-fitting model is 0.0041, indicating that dispersal rate for marine lineages is almost 250 times higher than the dispersal rate for freshwater lineages.

'MarineAsTrait' strategy		geog lnL	trait lnL	joint lnL	np	d	e	j	t12	t21	m1	m2	AICc weight	adding j P-value	adding m1 P-value
Trait-independent dispersal models	**DEC+t12+t21**	-266.96	-16.67	-283.62	4	0.01	0.01	0	0.001021	1.00E-05	1	1	9.40E-76		
	DEC+j+t12+t21	-110.26	-16.67	-126.93	5	8.84E-05	1.00E-12	0.026532	0.001021	1.00E-05	1	1	3.90E-08	3.98E-70	
	DIVALIKE+t12+t21	-268.44	-16.67	-285.11	4	0.01	0.01	0	0.001021	1.00E-05	1	1	2.10E-76		
	DIVALIKE+j+t12+t21	-112.18	-16.67	-128.84	5	1.05E-04	1.00E-12	0.027356	0.001021	1.00E-05	1	1	5.70E-09	6.13E-70	
	BAYAREALIKE+t12+t21	-284.96	-16.67	-301.6	4	0.01	0.01	0	0.001021	1.00E-05	1	1	1.40E-83		
	BAYAREALIKE+j+t12+t21	-112.93	-16.67	-129.60	5	7.97E-05	1.00E-12	0.027981	0.001021	1.00E-05	1	1	2.70E-09	8.33E-77	
Trait-dependent dispersal models	**DEC+t12+t21+m1**			-168.30	5	0.200448	0.001939	0	0.002405	0.086423	1.00E-04	1	4.20E-26		4.32E-52
	DEC+j+t12+t21+m1			-109.22	6	0.000522	1.00E-12	1.289991	0.000433	0.016057	0.0004094	1	0.7	1.60E-27	2.67E-09
	DIVALIKE+t12+t21+m1			-131.72	5	0.005603	1.00E-12	0	1.00E-05	0.030375	1.00E-04	1	3.20E-10		1.10E-68
	DIVALIKE+j+t12+t21+m1			-110.68	6	0.000616	1.00E-12	0.630209	0.000428	0.016132	0.011736	1	0.16	8.73E-11	1.67E-09
	BAYAREALIKE+t12+t21+m1			-172.26	5	0.004941	0.010131	0	1.00E-05	0.026105	1.00E-04	1	7.90E-28		3.25E-58
	BAYAREALIKE+j+t12+t21+m1			-110.85	6	0.000324	1.00E-12	0.271437	1.00E-05	0.026358	0.00363	1	0.14	1.52E-28	9.16E-10

Table 5.5. (*previous page*). Biogeographic model comparison with the 'MarineAsTrait' strategy. lnL = log-likelihood; np = number of parameters.

The 'MarineAsArea' and 'MarineAsTrait' strategies for including marine taxa are remarkably similar in results (Figs. 5.3 and 5.5). Excluding the nodes resulting as 'Marine' in the 'MarineAsArea' strategy, most likely ancestral ranges are almost identical under the two strategies (Tables 5.3 and 5.6). Moreover, there is a strong correspondence between nodes with a probable 'Marine' ancestral geographic range in the 'MarineAsArea' strategy, and nodes for which the ecological binary trait was reconstructed as being in the marine state under the 'MarineAsTrait' strategy (although support for a marine ancestor of all extant osteoglossids is much stronger when the 'MarineAsTrait' strategy is employed).

'MarineAsTrait' strategy, DEC+*j*+*t12*+*t21*+*m1* model	**Most likely ancestral range (marginal probability)**
Total-group Osteoglossomorpha	NA + Si (0.773)
Osteoglossomorpha	NA + Si (0.453)
Osteoglossiformes	Et (0.827)
Osteoglossidae	Et (0.805)
Arapaiminae	Et (0.719)
Osteoglossinae	Au (0.446)
Notopteroidei	Et (0.949)
Notopteridae	Et (0.890)

Table 5.6. Most likely ancestral ranges at key nodes in the bonytongue phylogeny under the best-fitting model (DEC+*j*+*t12*+*t21*+*m1*) with the 'MarineAsTrait' strategy, applied to the Bayesian consensus tree. See Fig. 5.2 for abbreviations of geographic areas. Clade names refer to crown clades unless specified otherwise.

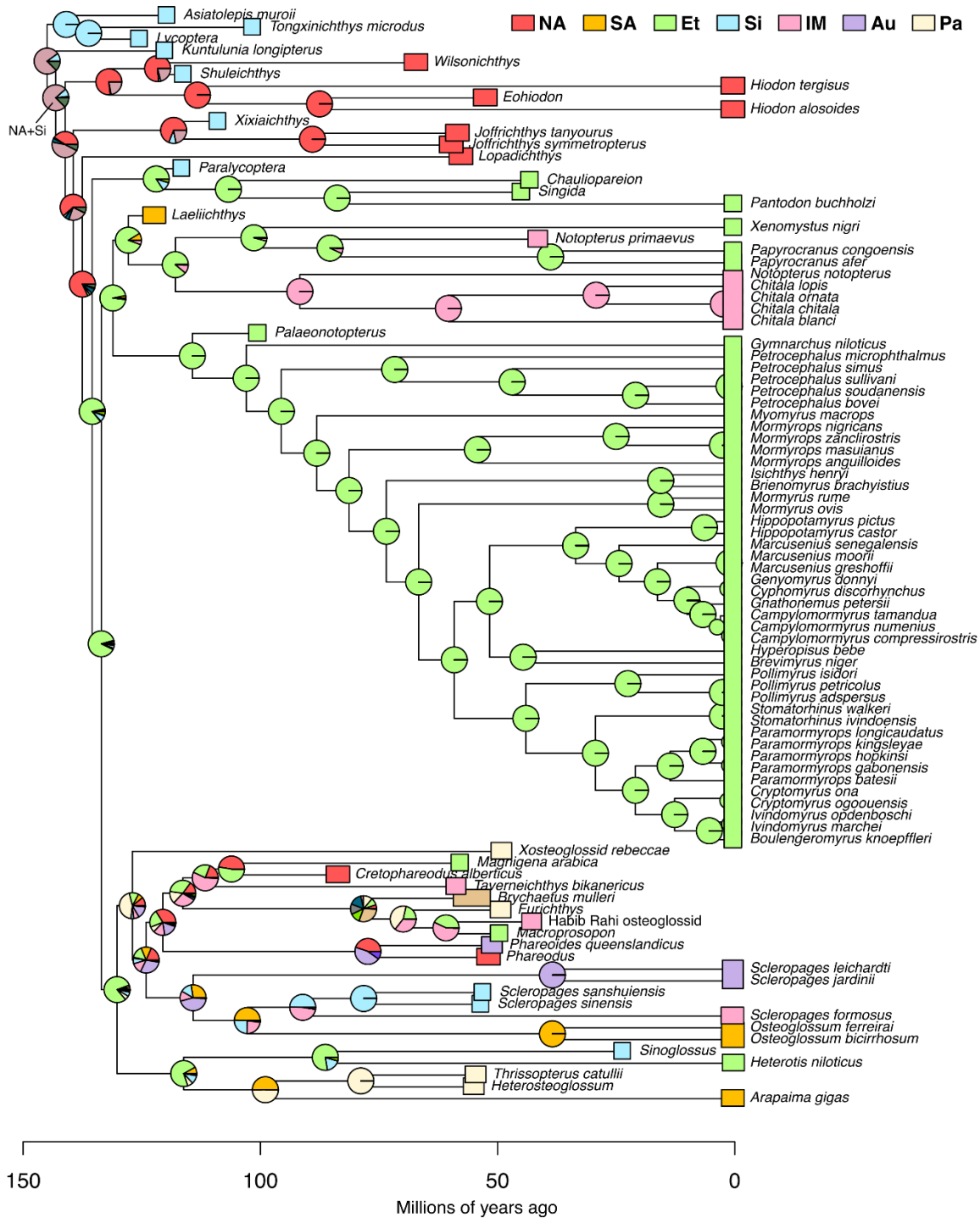

Fig. 5.5. Ancestral geographic range estimates under the best-fitting model (DEC+*j*+*t12*+*t21*+*m1)* with the 'MarineAsTrait' strategy, applied to the Bayesian consensus tree. See Fig. 5.2 for abbreviations of geographic areas.

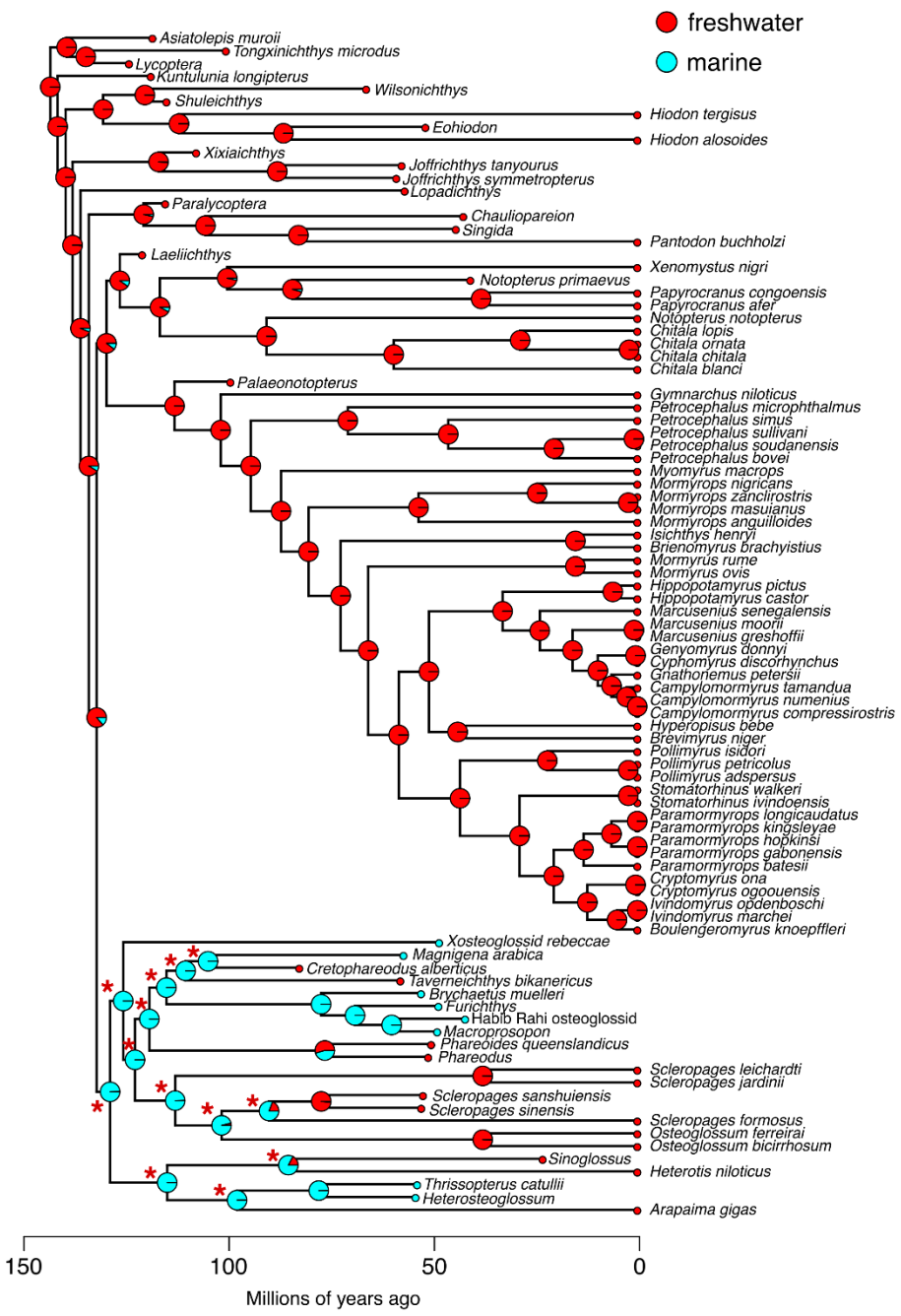

Fig. 5.6. Ancestral estimates for the ecological binary trait (freshwater/marine) under the best-fitting model (DEC+*j*+*t12*+*t21*+*m1)* with the 'MarineAsTrait' strategy, applied to the Bayesian consensus tree. Red asterisks mark ancestrally marine nodes that are reconstructed as freshwater under a simple 2-rate model (*t12*+*t21*) of discrete character evolution that is independent from geographic data.

DISCUSSION

Phylogenetic resolution and uncertainty within Osteoglossomorpha

The current study provides the most comprehensive phylogeny of Osteoglossomorpha to date in terms of taxonomic breadth, including all extant genera (except for one mormyrid genus) and 32 extinct taxa. Most high-level relationships between extant bonytongue subclades are strongly supported by both morphological and molecular data, and as such are confirmed by our analysis. The position of the African butterflyfish *Pantodon* as sister to all other osteoglossiforms has yet to be recovered by morphological data only. *Pantodon* has been resolved as nested within osteoglossids and as sister taxon of osteoglossines in several morphological analyses (Li & Wilson, 1996; Taverne, 1998; Hilton, 2003). However, the inclusion of more fossil taxa and the addition of characters has shifted its position rootward as sister to osteoglossids (Wilson & Murray, 2008; Chapter 4 of this Dissertation). It is likely that the similar morphological features shared by *Pantodon* and osteoglossids are either ancestral osteoglossiform characters that were lost in the very distinctive and morphologically derived mormyrids and notopterids, or the result of convergent evolution. New fossil discoveries of early osteoglossiforms could reveal a character distribution more compatible with the molecular topology and progressively align the position of *Pantodon* in morphological trees to its position in molecular and total-evidence analyses. One such fossil taxon that could impact phylogenetic inference is a yet undescribed osteoglossiform from the Early Cretaceous of Tunisia that shares some similarities with *Pantodon* (Ali *et al*., 2018).

Our analysis suggests that the osteoglossid *Scleropages* might not be monophyletic, with its Indo-Malayan species (*S. formosus*) more closely related to the South American *Osteoglossum* than to the Australian *Scleropages* species (*S. leichardti* and *S. jardinii*). *S. formosus* was recovered as being most closely related to the extinct species †*S. sinensis* and †*S. sanshuiensis* from the early Eocene of China, which were included for the first time in a phylogenetic analysis. Molecular phylogenetic analyses have yielded contrasting results, with most studies recovering a monophyletic *Scleropages* (Mu *et al*., 2012; Lavoué, 2015; de Bello Cioffi, 2019), but at least one other study supporting the same paraphyletic topology found here (Rabosky *et al*., 2018). Strikingly, the caudal skeleton of the Australian species of *Scleropages* is morphologically very different not only from that of *S. formosus*, but also from that of all other osteoglossids (Xu & Chang, 2009; Hilton & Britz, 2010). Hilton & Lavoué (2018) already noted that there are no known morphological synapomorphies supporting the monophyly of *Scleropages*. Our study strengthens the need for a careful systematic revision of the genus

Scleropages and of its close relative *Osteoglossum* on both morphological and molecular grounds, in order to test the monophyly of *Scleropages*.

Despite our best efforts to include as many marine bonytongue taxa as possible, some of them were left out of this analysis, due to either poor preservation quality or high incompleteness (†*Brychaetoides*, †*Monopteros*, †*Ridewoodichthys*), or pending more in-depth study (†*Opsithrissops*, indeterminate osteoglossid from London Clay; see Forey & Hilton, 2010). It is likely that these taxa fit somewhere in the 'cloud' of marine bonytongues surrounding the base of Osteoglossidae, as they often display a mix of osteoglossine and arapaimine characters. In Chapter 4 of this Dissertation, we highlighted some osteological features that we interpreted as synapomorphies of an Arapaiminae + †Phareodontinae clade. While the phylogenetic analysis in this study slightly favors an osteoglossine (rather than arapaimine) affinity for †phareodontines, several of the new taxa added in this Chapter cannot be scored for the key characters highlighted in Chapter 4, due to incomplete preservation. Thus, more complete specimens will be needed to stabilize the three-way relationships between arapaimines, osteoglossines and †phareodontines.

Biogeographic history of Osteoglossomorpha

The biogeographic analyses included in the present study allow us to paint a comprehensive picture of the biogeographic history of bonytongue fishes. The origin of bonytongues can be traced back to at least the Middle–Late Jurassic (Capobianco & Friedman, 2019). At that time, they likely inhabited Laurasian landmasses (North America and continental Asia), with the possible inclusion of Africa. Although, as part of the Gondwanan supercontinent, Africa was geographically separated from Laurasian landmasses by the Tethys ocean since the Middle–Late Jurassic (Blakey, 2008), the fossil record of African terrestrial vertebrates in the Jurassic suggests striking similarities with the biotas of Laurasia (Ezcurra & Agnolín, 2012; Dunhill *et al*., 2016; Haddoumi *et al*., 2016). Intriguingly, the earliest putative bonytongue fossils are represented by fragmentary squamules found in Middle Jurassic continental deposits of Morocco (Haddoumi *et al*., 2016). Similar specimens – although initially reported as indeterminate actinopterygians – have been found in the Late Jurassic Morrison Formation of Wyoming (Foster & Heckert, 2011; Haddoumi *et al*., 2016). Thus, a Laurasian + African ancestral range for bonytongues is plausible on the basis of both our phylogeny-based biogeographic analysis and of

the paleogeographic distribution of fossil specimens that are too fragmentary to be included in a phylogenetic framework. The divergence between hiodontiforms and osteoglossiforms is compatible with a vicariance event where hiodontiforms become restricted to Laurasia and osteoglossiforms to Africa. This hypothesis has been previously proposed in the literature (Cavin, 2008, 2017) and finds support in our analysis. We point out that, if North American and Asian taxa such as †*Xixiaichthys*, †*Joffrichthys*, †*Lopadichthys* and potentially †*Paralycoptera* are indeed stem osteoglossiforms, then the vicariance event separating Laurasian from African lineages likely happened after the hiodontiform-osteoglossiform divergence and before the origin of crown osteoglossiforms. There is some evidence for multiple dispersals between Asia and North America during the Cretaceous, as shown by potential sister-group relationships between †*Shuleichthys* and †*Wilsonichthys* and between †*Xixiaichthys* and †*Joffrichthys.* Hiodontiforms lived in Asia potentially up to the Paleogene (Hilton & Grande, 2008) before becoming extinct there and surviving exclusively in North America, where the extant mooneyes *Hiodon* are found today. Biotic connections between Asia, North America, and Europe during the Cretaceous and Paleogene – likely driven by transient land bridges between these continents – are not uncommon in freshwater fishes, as shown by the fossil record and phylogenetic history of sturgeons, paddlefishes, bowfins, cypriniforms and pikes (Cavin, 2008, 2017; Capobianco & Friedman, 2019).

An African origin of osteoglossiforms has often been proposed based on the presence of all extant osteoglossiform families in Africa (Nelson, 1969; Wilson & Murray, 2008). This hypothesis is strongly favored by our analyses. Crown osteoglossiforms must have originated at least in the Barremian (around 125 million years ago), as that is the age of its oldest known representatives (Ali *et al*., 2018; Brito *et al*., 2020). At that time, Africa and South America were still joined in a western Gondwanan landmass. Indeed, several freshwater or euryhaline fish groups (including †mawsoniids, polypterids, †vidalamiins, †cladocyclids and chanids) are commonly found in Early Cretaceous and mid-Cretaceous (Cenomanian) deposits on both continents (Capobianco & Friedman, 2019). Thus, the almost complete absence of osteoglossiforms from South American deposits of that age is quite surprising. The only exception is represented by †*Laeliichthys* from Brazil, recently re-described as a stem notopterid (Brito *et al*., 2020), a position that is confirmed by the current study. Given the age and phylogenetic placement of †*Laeliichthys*, if the ancestral range of osteoglossiforms encompassed

western Gondwana then we would expect stem pantodontids, stem mormyroids and stem osteoglossids to have inhabited South America at some point before going regionally extinct. In alternative, the presence of †*Laeliichthys* in South America could be the result of dispersal from a more restricted African range. Additional fossil discoveries made in Cretaceous deposits of South America and Africa might give further insight into this part of the biogeographic history of bonytongues.

In our most likely reconstruction, notopterid knifefishes originated in Africa and reached Indo-Malaya through dispersal. This hypothesis has been previously proposed on the basis of Asian notopterids being nested within an African clade (Notopteridae + Mormyroidea) and likely younger than the separation between Africa and eastern Gondwana (including the Indian subcontinent; Lavoué, 2016). The two possible scenarios within this hypothesis involve an ancient dispersal from Africa to the Indo-Malagasy landmass through the Mozambique Channel in the Cretaceous or a more recent dispersal from Africa to Eurasia in the Neogene after the closure of the Tethys ocean (Inoue *et al*., 2009; Barby *et al*., 2018). Otoliths attributed to notopterids are already present in latest Cretaceous deposits of India (Nolf *et al*., 2008), supporting the ancient dispersal scenario. However, a different hypothesis could be drawn from the inferred relationships of †*Notopterus primaevus*. This species, found in Eocene freshwater deposits of Sumatra, is resolved in our study as being closer to the extant African notopterids than to the Asian ones. While it is more often found as sister to *Papyrocranus* only, this position is driven by the scoring of a single character (character 94: posterior process of the hyomandibula longer than half the length of its dorsal articular surface). Remarkably, all posterior trees in which †*Notopterus primaevus* is sister to African notopterids (*Papyrocranus* + *Xenomystus*) result in an Indo-Malayan (rather than African) origin of crown Notopteridae. In this scenario, notopterids originated in the Indian subcontinent during the Cretaceous (either through vicariance when it separated from Africa, or through dispersal across the Mozambique Channel), dispersed into the rest of Southeast Asia after the Indian subcontinent migrated northward and docked into continental Asia, then dispersed more recently from Asia to Africa during the Cenozoic. Strikingly, this hypothesis would be congruent with general biogeographic patterns of the African freshwater fish fauna, in which up to 20% of the extant species diversity in the continent seems to derive from Asia-to-Africa dispersal events since the Oligocene, whereas dispersals from Africa to Asia appear to be much rarer (Lavoué, 2020). A detailed

redescription of †*Notopterus primaevus* and a reassessment of the osteological characters differentiating the four extant genera of notopterid knifefishes will be necessary to better evaluate the affinities of this striking fossil species and to test alternative biogeographic scenarios for the origin of notopterids.

Osteoglossid bonytongues are by far the most geographically widespread lineage within Osteoglossomorpha, with extant members in South America, Africa, Southeast Asia and Australia, and fossils known from all continents except Antarctica. We find strong support for a marine origin of osteoglossids, followed by multiple independent marine-to-freshwater transitions, as the primary cause for this widespread distribution. Under this scenario, the freshwater Paleogene taxa †*Phareodus* and †*Cretophareodus* from North America, †*Phareoides* from Australia, and †*Taverneichthys* from India all derive from marine ancestors that dispersed between different continents and reentered freshwater environments. Additionally, arapaimines are recovered as ancestrally marine both when treating 'marine' as an ecological trait and when treating it as a geographic state. Marine ancestry and dispersal could not only explain the disjunct distribution of *Arapaima* in South America and *Heterotis* in Africa, but also the presence of the fossil †*Sinoglossus* in China. Alternatively, if †*Sinoglossus* is more closely related to *Heterotis* than to *Arapaima*, an Asia-to-Africa (or vice versa) overland dispersal in the Oligocene would also represent a viable scenario. Osteoglossine biogeographic history remains more uncertain, as their last common ancestor is reconstructed as marine only when treating 'marine' as an ecological trait. Nonetheless, their geographic distribution is difficult to explain otherwise. *Scleropages* has always represented a biogeographic conundrum, as it is the only strictly freshwater fish genus to occur on both sides of the Wallace line separating the Indo-Malayan and Australian regions (Lavoué, 2015). A possible explanation for the distribution of osteoglossines in both Australia and South America involves overland dispersal through Antarctica around the latest Cretaceous, when these landmasses were still connected. However, this scenario cannot explain the presence of *Scleropages* in Southeast Asia and China, nor is it consistent with the topology recovered in the study (where the South American *Osteoglossum* is more closely related to the Southeast Asian than to the Australian *Scleropages*).

Marine origin and dispersal in bonytongue fishes

Marine dispersal in osteoglossid bonytongues has been proposed since the recognition of †*Brychaetus* as an osteoglossid (Patterson, 1975). Several more taxa have been subsequently described from marine deposits. (Taverne, 1998; Bonde, 2008; Forey & Hilton, 2010). However, the lack of a phylogenetic framework or the uncertain systematics of these marine forms hindered any formal test of the marine dispersal hypothesis. Until the current study, the strongest arguments in favor of marine dispersal in osteoglossids came from estimated divergence times younger than continental break-ups (Lavoué, 2015; 2016) and from the observation that closely related Paleogene taxa (sometimes even classified in the same genus) have been found in freshwater deposits as distant as Wyoming is from Australia (Capobianco *et al*., 2021). We find that, rather than forming a single clade or being randomly interspersed across bonytongue phylogeny, marine bonytongues form a 'cloud' at the base of Osteoglossidae from which all three major osteoglossid subclades (Arapaiminae, Osteoglossinae and †Phareodontinae) emerged. Trait-dependent dispersal models strongly support a single freshwater-to-marine transition on the osteoglossid stem, followed by around 10 independent marine-to-freshwater reversals. This result is remarkable for several reasons. Firstly, to our knowledge, this is the first time that a group whose extant members and closer outgroups are all exclusively freshwater is reconstructed as ancestrally marine. Secondly, major environmental transitions such as the marine-to-freshwater one are usually quite rare in teleost fish groups (Bloom & Lovejoy, 2012; Conway *et al*., 2017). Lastly, this reconstruction implies that several distinct lineages of osteoglossids were wiped out from marine environments during the Ypresian or later, and these fishes never reinvaded the sea afterwards. At this point is difficult to speculate whether competition with other predatory fishes such as several acanthomorph lineages, or severe climate change towards colder temperatures might have played some role in this extinction.

Extant osteoglossids do not possess anatomical or physiological feature that are obviously associated with marine ancestry. However, the possibility that some of the peculiarities displayed by osteoglossids might be eventually reinterpreted in light of their marine origin cannot be discounted. As an example, we point out that osteoglossids are unique among osteoglossomorphs in being mouthbrooders (except for *Arapaima*; Koenig & Gallant, 2021). Mouthbrooding is a type of parental care found in some shallow marine clades and few freshwater lineages, including cichlids, but it is unknown whether its independent appearance in several fish groups could be correlated with a particular environmental setting. Rather than suggesting that

mouthbrooding (or any other character of extant osteoglossid) evolved in marine forms first, we would just like to stimulate researchers working on the anatomy, physiology and behavior of osteoglossid bonytongues to consider their probable marine origins when investigating their evolution.

When osteoglossids invaded marine environments remains an open question. Because occurrences of bonytongue fossils in marine environments are mostly restricted to the Paleogene, it has been proposed that the diversification of marine bonytongues happened on the wake of the Cretaceous-Paleogene mass extinction that wiped out several large predatory taxa, and triggered the diversification of new specialized clades, including most of the predatory fish groups that roam in the seas today (Capobianco *et al*., 2020; 2021). However, if †*Cretophareodus* is part of †phareodontine radiation and descended from marine ancestors, as suggested by our Bayesian consensus tree, then the minimum age for the marine invasion is at least Campanian (around 80 million years ago). Given the poor state of preservation of †*Cretophareodus* and the high phylogenetic uncertainty that surrounds several fossil osteoglossids, it is possible that future attempts at co-estimation of topology and divergence times using FBD-derived models will not give a reasonably precise time of origin for marine bonytongues. A more promising avenue might be undertaken by asking whether the absence of osteoglossid specimens from Cretaceous marine deposits (apart from few possible occurrences in the Maastrichtian; see Capobianco *et al*., 2021) is a genuine absence or a taphonomic artifact. Given that even fragmentary osteoglossid remains such as teeth, vertebrae and scales can be correctly identified, how likely it is that we never found any of those in pre-Maastrichtian marine deposits knowing how many of those deposits yielded fish material? Some approaches have been recently developed to answer this kind of question (see for example Saulsbury & Baumiller, in review), and we believe that these are more likely to advance our knowledge of the timing of the marine bonytongue radiation.

Impact of fossils on biogeographic inference

Paleontological data dramatically changes estimates for ancestral geographic ranges across the bonytongue phylogeny. The most likely biogeographic history when ignoring fossil taxa is more compatible with a vicariant scenario in which *Heterotis* and *Arapaima* are descended from a western Gondwanan ancestor, and the Australian *Scleropages* descended from a South American

ancestor that likely dispersed through Antarctica into Australia when these three continents still maintained some land connection. Moreover, ignoring fossil taxa removes continental Asia from any consideration on the area of origin of osteoglossomorphs. More importantly, fossil taxa and their paleoenvironmental context are the only line of evidence for the spectacular—and totally unexpected—marine phase of bonytongue evolution.

Trait-based dispersal models represent an exciting recent development in historical biogeography, as they reintegrate the biology of the studied organisms into biogeographic models and allow for questions beyond ancestral range estimation (Sukumaran & Knowles, 2018). Practical application of trait-based dispersal models has been so far limited to datasets of extant taxa (Blom *et al*., 2019; Nicolaï & Matzke, 2019; Klaus & Matzke, 2020), with in one case the addition of few recently extinct taxa with ranges and traits within the variability displayed by living species (Garcia-R & Matzke, 2021). Our study showcases how insightful these models can be when applied to extant groups with an abundant fossil record that displays a range of geographic distributions and ecological traits going beyond those observed in modern species.

A potential refinement for future studies of osteoglossomorph biogeography would be to explicitly include a time factor into the biogeographic model (allowing for different parameter estimates and different plausible ranges at each pre-determined time slice) and include a distance multiplier (whereby dispersal probability can be dependent on the geographic distance between two areas, and that distance can change at different time slices). In order to do this, an accurate time-calibrated tree of osteoglossomorphs would be required. The development of FBD models that allow for clade-specific fossil preservation potential, and a detailed revision at a genus and species level of the osteology of extant osteoglossiforms would be necessary preliminary steps to complete such a task.

ACKNOWLEDGMENTS

The authors would like to thank Danielle Goodvin (UMMP) and Kyle Kramer (Undergraduate Research Opportunity Program at the University of Michigan) for great help with segmentation of CT data. We also thank for specimen access Douglas Nelson and Randy Singer at the

University of Michigan Museum of Zoology, William F. Simpson at the Field Museum of Natural History, Chicago, Bo Schultz and René L. Sylvestersen at the Fur Museum, Bent E. K. Lindow at the Natural History Museum of Denmark, Emma Bernard at the Natural History Museum, London, Mariagabriella Fornasiero at the Istituto Geologico dell'Università di Padova, Anna Vaccari and Roberta Salmaso at the Museo Civico di Storia Naturale, Verona, Gaël Clement at the Muséum National d'Histoire Naturelle, Paris and Florias Mees at the Musée Royal de l'Afrique Centrale. We would like to thank Zach Randall and the Florida Museum of Natural History, University of Florida for permission to use CT data stored on Morphosource. This study includes data produced in the CTEES facility at University of Michigan, supported by the Department of Earth & Environmental Sciences and College of Literature, Science, and the Arts. This work was supported by funding from the Department of Earth and Environmental Sciences of the University of Michigan (Scott Turner Student Research Grant Award 2017, to A.C.), by the Rackham Graduate School of the University of Michigan (Rackham Predoctoral Fellowship Award 2020-2021, to A.C.) and by the Society of Systematic Biologists (2017 SSB Graduate Student Research Award, to A.C.).

REFERENCES

Aberer, A. J., Krompass, D., & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic biology*, 62, 162-166.

Ali, W. B., Cavin, L., Boukhalfa, K., Ouaja, M., & Soussi, M. (2018). Fish assemblage and palaeoenvironment of Early Cretaceous (Barremian) neap-spring tidal rhythmites from Sidi Aïch Formation of the Chotts basin (Southern Tunisia). *Cretaceous Research*, 92, 31-42.

Arratia, G. (2000). Remarkable teleostean fishes from the Late Jurassic of southern Germany and their phylogenetic relationships. *Fossil Record*, 3, 137-179.

Arratia, G. (2008). The varasichthyid and other crossognathiform fishes, and the break-up of Pangaea. *Geological Society, London, Special Publications*, 295, 71-92.

Arratia, G. (2013). Morphology, taxonomy, and phylogeny of Triassic pholidophorid fishes (Actinopterygii, Teleostei). *Journal of Vertebrate Paleontology*, 33(sup1), 1-138.

Arratia, G. (2016). New remarkable Late Jurassic teleosts from southern Germany: Ascalaboidae n. fam., its content, morphology, and phylogenetic relationships. *Fossil Record*, 19, 31-59.

Arratia, G., & Tintori, A. (1999). The caudal skeleton of the Triassic actinopterygian †*Prohalecites* and its phylogenetic position. *Mesozoic fishes*, *2*, 121-142.

Barby, F. F., Ráb, P., Lavoué, S., Ezaz, T., Bertollo, L. A. C., Kilian, A., ... & Cioffi, M. D. B. (2018). From chromosomes to genome: insights into the evolutionary relationships and biogeography of Old World knifefishes (Notopteridae; Osteoglossiformes). *Genes*, 9, 306.

Bauer, J. E. (2021). Paleobiogeography, paleoecology, diversity, and speciation patterns in the Eublastoidea (Blastozoa: Echinodermata). *Paleobiology*, 47, 221-235.

Bean, L. B., & Arratia, G. (2020). Anatomical revision of the Australian teleosts *Cavenderichthys talbragarensis* and *Waldmanichthys koonwarri* impacting on previous phylogenetic interpretations of teleostean relationships. *Alcheringa: An Australasian Journal of Palaeontology*, 44, 121-159.

de Bello Cioffi, M., Ráb, P., Ezaz, T., Antonio Carlos Bertollo, L., Lavoué, S., Aguiar de Oliveira, E., Sember, A., Franco Molina, W., de Souza, H. S., Majtanova, Z. & Liehr, T. (2019). Deciphering the evolutionary history of arowana fishes (Teleostei, Osteoglossiformes, Osteoglossidae): Insight from comparative cytogenomics. *International journal of molecular sciences*, 20, 4296.

Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., & Orti, G. (2017). Phylogenetic classification of bony fishes. *BMC evolutionary biology*, 17, 1-40.

Blakey, R. C. (2008). Gondwana paleogeography from assembly to breakup—A 500 my odyssey. *Geological Society of America Special Papers*, 441, 1-28.

Blom, M. P., Matzke, N. J., Bragg, J. G., Arida, E., Austin, C. C., Backlin, A. R., ... & Moritz, C. (2019). Habitat preference modulates trans-oceanic dispersal in a terrestrial vertebrate. *Proceedings of the Royal Society B*, 286, 20182575.

Bloom, D. D., & Lovejoy, N. R. (2012). Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *Journal of Evolutionary Biology*, 25, 701-715.

Bonde, N. (2008). Osteoglossomorphs of the marine Lower Eocene of Denmark–with remarks on other Eocene taxa and their importance for palaeobiogeography. *Geological Society, London, Special Publications*, 295, 253-310.

Brito, P. M., Figueiredo, F. J., & Leal, M. E. C. (2020). A revision of *Laeliichthys ancestralis* Santos, 1985 (Teleostei: Osteoglossomorpha) from the Lower Cretaceous of Brazil: Phylogenetic relationships and biogeographical implications. *Plos one*, 15, e0241009.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference—a practical information-theoretic approach*. New York: Springer.

Capobianco, A., & Friedman, M. (2019). Vicariance and dispersal in southern hemisphere freshwater fish clades: a palaeontological perspective. *Biological Reviews*, 94, 662-699.

Capobianco, A., Beckett, H. T., Steurbaut, E., Gingerich, P. D., Carnevale, G., & Friedman, M. (2020). Large-bodied sabre-toothed anchovies reveal unanticipated ecological diversity in early Palaeogene teleosts. *Royal Society Open Science*, 7, 192260.

Capobianco, A., Foreman, E., & Friedman, M. (2021). A Paleocene (Danian) marine osteoglossid (Teleostei, Osteoglossomorpha) from the Nuussuaq Basin of Greenland, with a brief review of Palaeogene marine bonytongue fishes. *Papers in Palaeontology*, 7, 625-640.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, 17, 540-552.

Cavin, L. (2008). Palaeobiogeography of cretaceous bony fishes (Actinistia, Dipnoi and Actinopterygii). *Geological Society, London, Special Publications*, 295, 165-183.

Cavin, L. (2017). *Freshwater Fishes: 250 Million Years of Evolutionary History*. ISTE Press-Elsevier, London and Oxford.

Conway, K. W., Kim, D., Rüber, L., Pérez, H. S. E., & Hastings, P. A. (2017). Molecular systematics of the New World clingfish genus Gobiesox (Teleostei: Gobiesocidae) and the origin of a freshwater clade. *Molecular phylogenetics and evolution*, 112, 138-147.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Dunhill, A. M., Bestwick, J., Narey, H., & Sciberras, J. (2016). Dinosaur biogeographical structure and Mesozoic continental fragmentation: a network-based approach. *Journal of Biogeography*, 43, 1691-1704.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32, 1792-1797.

Eme, D., Anderson, M. J., Struthers, C. D., Roberts, C. D., & Liggins, L. (2019). An integrated pathway for building regional phylogenies for ecological studies. *Global Ecology and Biogeography*, 28, 1899-1911.

Ezcurra, M. D., & Agnolín, F. L. (2012). A new global palaeobiogeographical model for the late Mesozoic and early Tertiary. *Systematic Biology*, 61, 553-566.

Forey, P. L. & Hilton, E. J. (2010). Two new tertiary osteoglossid fishes (Teleostei: Osteoglossomorpha) with notes on the history of the family. In *Morphology, Phylogeny and Paleobiogeography of Fossil Fishes* (eds D. K. Elliott, J. G. Maisey, X. Yu and D. Miao), pp. 215– 246. Verlag Dr. F. Pfeil, München.

Foster, J. R., & Heckert, A. B. (2011). Ichthyoliths and other microvertebrate remains from the Morrison Formation (Upper Jurassic) of northeastern Wyoming: a screen-washed sample indicates a significant aquatic component to the fauna. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 305, 264-279.

Garcia-R, J. C., & Matzke, N. J. (2021). Trait-dependent dispersal in rails (Aves: Rallidae): Historical biogeography of a cosmopolitan bird clade. *Molecular phylogenetics and evolution*, 159, 107106.

Grande, L. (1985). The use of paleontology in systematics and biogeography, and a time control refinement for historical biogeography. *Paleobiology*, 11, 234-243.

Haddoumi, H., Allain, R., Meslouh, S., Metais, G., Monbaron, M., Pons, D., Rage, J. C., Vullo, R., Zouhri, S., & Gheerbrant, E. (2016). Guelb el Ahmar (Bathonian, Anoual Syncline, eastern Morocco): first continental flora and fauna including mammals from the Middle Jurassic of Africa. *Gondwana Research*, 29, 290-319.

Hart, M. B., Arratia, G., Moore, C., & Ciotti, B. J. (2020). Life and death in the Jurassic seas of Dorset, Southern England. *Proceedings of the Geologists' Association*, 131, 629-638.

Heath, T. A., Huelsenbeck, J. P., & Stadler, T. (2014). The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111, E2957-E2966.

Hedman, M. M. (2010). Constraints on clade ages from fossil outgroups. *Paleobiology*, 36, 16-31.

Hills, E. S. (1934). Tertiary fresh water fishes from southern Queensland. *Memoirs of the Queensland Museum* 10, 163– 172.

Hilton, E. J. (2003). Comparative osteology and phylogenetic systematics of fossil and living bony-tongue fishes (Actinopterygii, Teleostei, Osteoglossomorpha). *Zoological Journal of the Linnean Society* 137, 1– 100.

Hilton, E. J., & Britz, R. (2010). The caudal skeleton of osteoglossomorph fishes, revisited: comparisons, homologies, and characters. *Origin and phylogenetic interrelationships of teleosts*, 219-237.

Hilton, E. J., & Grande, L. (2008). Fossil mooneyes (Teleostei: Hiodontiformes, Hiodontidae) from the Eocene of western North America, with a reassessment of their taxonomy. *Geological Society, London, Special Publications*, 295, 221-251.

Hilton, E. J., & Lavoué, S. (2018). A review of the systematic biology of fossil and living bony-tongue fishes, Osteoglossomorpha (Actinopterygii: Teleostei). *Neotropical Ichthyology*, 16, e180031.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, 35, 518-522.

Humphries, C. J., & Parenti, L. R. (1986). *Cladistic Biogeography*. Oxford Monographs on Biogeography 2. Oxford: Clarendon Press.

Inoue, J. G., Kumazawa, Y., Miya, M., & Nishida, M. (2009). The historical biogeography of the freshwater knifefishes using mitogenomic approaches: a Mesozoic origin of the Asian notopterids (Actinopterygii: Osteoglossomorpha). *Molecular Phylogenetics and Evolution*, 51, 486-499.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14, 587-589.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30, 772-780.

King, B. (2021). Bayesian tip-dated phylogenetics in paleontology: topological effects and stratigraphic fit. *Systematic Biology*, 70, 283-294.

Klaus, K. V., & Matzke, N. J. (2020). Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. *Systematic Biology*, 69, 61-75.

Klopfstein, S., & Spasojevic, T. (2019). Illustrating phylogenetic placement of fossils using RoguePlots: An example from ichneumonid parasitoid wasps (Hymenoptera, Ichneumonidae) and an extensive morphological matrix. *PloS one*, *14*(4), e0212942.

Koenig, L. A., & Gallant, J. R. (2021). Sperm competition, sexual selection and the diverse reproductive biology of Osteoglossiformes. *Journal of Fish Biology*, Early View.

Konwert, M. (2016). First record of *Leptolepides haerteisi* (Teleostei, Orthogonikleithridae) from the Upper Jurassic Plattenkalks of Schamhaupten (Bavaria, Germany). *Archaeopteryx*, 33, 54-61.

Konwert, M., & Stumpf, S. (2017). Exceptionally preserved Leptolepidae (Actinopterygii, Teleostei) from the late Early Jurassic Fossil-Lagerstätten of Grimmen and Dobbertin (Mecklenburg-Western Pomerania, Germany). *Zootaxa*, 4243, 249-296.

Kumar, K., Rana, R. S. & Paliwal, B. S. (2005). Osteoglossid and lepisosteid fish remains from the Paleocene Palana Formation, Rajasthan, India. *Palaeontology* 48, 1187–1210.

Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. *Systematic biology*, 62, 789–804.

Landis, M. J., Freyman, W. A., & Baldwin, B. G. (2018). Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution*, 72, 2343–2359.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular biology and evolution*, 34, 772-773.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30, 3276-3278.

Lassmann, T., & Sonnhammer, E. L. (2005). Automatic assessment of alignment quality. *Nucleic acids research*, 33, 7120-7128.

Lavoué, S. (2015). Testing a time hypothesis in the biogeography of the arowana genus Scleropages (Osteoglossidae). *Journal of Biogeography*, 42, 2427-2439.

Lavoué, S. (2016). Was Gondwanan breakup the cause of the intercontinental distribution of Osteoglossiformes? A time-calibrated phylogenetic test combining molecular, morphological, and paleontological evidence. *Molecular phylogenetics and evolution*, 99, 34-43.

Lavoué, S. (2020). Origins of Afrotropical freshwater fishes. *Zoological Journal of the Linnean Society*, 188, 345-411.

Lavoué, S., & Sullivan, J. P. (2004). Simultaneous analysis of five molecular markers provides a well-supported phylogenetic hypothesis for the living bony-tongue fishes (Osteoglossomorpha: Teleostei). *Molecular phylogenetics and evolution*, 33, 171-185.

Lavoué, S., Miya, M., Arnegard, M. E., McIntyre, P. B., Mamonekene, V., & Nishida, M. (2011). Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. *Proceedings of the Royal Society B: Biological Sciences*, 278, 1003-1008.

Lavoué, S., Miya, M., Arnegard, M. E., Sullivan, J. P., Hopkins, C. D., & Nishida, M. (2012). Comparable ages for the independent origins of electrogenesis in African and South American weakly electric fishes. *PLoS one*, 7, e36287.

Leroy, B., Dias, M. S., Giraud, E., Hugueny, B., Jézéquel, C., Leprieur, F., Oberdorff, T., & Tedesco, P. A. (2019). Global biogeographical regions of freshwater fish species. *Journal of biogeography*, 46, 2407-2419.

Li, G.-Q. (1994). Systematic position of the Australian fossil osteoglossid fish †*Phareodus* (= *Phareoides*) *queenslandicus* Hills. *Memoirs of the Queensland Museum* 37, 287– 300.

Li, G.-Q. (1996). A new species of Late Cretaceous osteoglossid (Teleostei) from the Oldman Formation of Alberta, Canada, and its phylogenetic relationships. In *Mesozoic Fishes. Systematics and Paleoecology* (eds G. Arratia and G. Viohl), pp. 285– 298. Verlag Dr. F. Pfeil, München.

Li, G.-Q. & Wilson, M. V. H. (1996). The discovery of Heterotidinae (Teleostei: Osteoglossidae) from the Paleocene Paskapoo formation of Alberta, Canada. *Journal of Vertebrate Paleontology* 16, 198– 209.

Lieberman, B. S. (2003). Paleobiogeography: the relevance of fossils to biogeography. *Annual Review of Ecology, Evolution, and Systematics*, 34, 51-69.

Lloyd, G. T., Bapst, D. W., Friedman, M., & Davis, K. E. (2016). Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biology letters*, 12, 20160609.

Lundberg, J. G. & Chernoff, B. (1992). A Miocene fossil of the Amazonian fish *Arapaima* (Teleostei, Arapaimidae) from the Magdalena River region of Colombia--Biogeographic and evolutionary implications. *Biotropica* 24, 2– 14.

Maddison, W. P. & D.R. Maddison. 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61 http://www.mesquiteproject.org

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic biology*, 63, 951-970.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37, 1530-1534.

Mongiardino Koch, N., Garwood, R. J., & Parry, L. A. (2021). Fossils improve phylogenetic analyses of morphological characters. *Proceedings of the Royal Society B*, 288, 20210044.

Mu, X. D., Song, H. M., Wang, X. J., Yang, Y. X., Luo, D., Gu, D. E., Luo, J. R. & Hu, Y. C. (2012). Genetic variability of the Asian arowana, Scleropages formosus, based on mitochondrial DNA genes. *Biochemical Systematics and Ecology*, *44*, 141-148.

Nelson, G. J. (1969). Infraorbital bones and their bearing on the phylogeny and geography of osteoglossomorph fishes. *American Museum novitates*; 2394.

Nicolaï, M. P., & Matzke, N. J. (2019). Trait-based range expansion aided in the global radiation of Crocodylidae. *Global Ecology and Biogeography*, 28, 1244-1258.

Nolf, D., Rana, R. S. & Prasad, G. V. (2008). Late Cretaceous (Maastrichtian) fish otoliths from the Deccan intertrappean beds, India: a revision. *Bulletin: Sciences de la Terre* 78, 239– 259.

Otero, O. & Gayet, M. (2001). Palaeoichthyofaunas from the Lower Oligocene and Miocene of the Arabian Plate: palaeoecological and palaeobiogeographical implications. *Palaeogeography, Palaeoclimatology, Palaeoecology* 165, 141– 169.

Otero, O., Pinton, A., Cappetta, H., Adnet, S., Valentin, X., Salem, M., & Jaeger, J. J. (2015). A fish assemblage from the Middle Eocene from Libya (Dur At-Talah) and the earliest record of modern African fish genera. *PLoS One*, 10, e0144358.

Patterson, C. (1975). The distribution of Mesozoic freshwater fishes. *Mémoires du Muséum national d'histoire naturelle Série A Zoologie*, 88, 156-174.

Patterson, C., & Rosen, D. E. (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.

Poyato-Ariza, F. J. (1999). The elopiform fish† *Anaethalion angustus* restored, with comments on individual variation. *Mesozoic fishes–systematics and fossil record. Verlag Dr. Friedrich Pfeil, Munich*, 361-368.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559, 392-395.

Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography*, 45, 741-749.

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57, 4-14.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, 46, 195-203.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61, 539-542.

Sanders, M. (1934). Die Fossilen Fische der Alttertiären Süsswasserablagerungen aus Mittel-Sumatra. *Verhandelingen van het Geologisch-Mijnbouwkundig Genootschap voor Nederland en Koloni'n Geologische Series* 11, 1– 144.

Santaquiteria, A., Siqueira, A. C., Duarte-Ribeiro, E., Carnevale, G., White, W. T., Pogonoski, J. J., Baldwin, C. C., Ortí, G., Arcila, D., & Betancur-R., R. (2021). Phylogenomics and historical biogeography of seahorses, dragonets, goatfishes, and allies (Teleostei: Syngnatharia): Assessing factors driving uncertainty in biogeographic inferences. *Systematic Biology*, 1–18.

Saulsbury, J., & Baumiller, T. In review. Dispersal from the ancient West Tethys as a source of the modern Indo-West Pacific marine biodiversity hotspot.

Sferco, E., López-Arbarello, A., & Báez, A. M. (2015). Phylogenetic relationships of †*Luisiella feruglioi* (Bordas) and the recognition of a new clade of freshwater teleosts from the Jurassic of Gondwana. *BMC Evolutionary Biology*, 15, 1-15.

Simões, T. R., Caldwell, M. W., & Pierce, S. E. (2020). Sphenodontian phylogeny and the impact of model choice in Bayesian morphological clock estimates of divergence times and evolutionary rates. *BMC biology*, 18, 1-30.

Sukumaran, J., & Knowles, L. L. (2018). Trait-dependent biogeography:(re) integrating biology into probabilistic historical biogeographical models. *Trends in ecology & evolution*, 33, 390-398.

Sullivan, J. P., Lavoué, S., & Hopkins, C. D. (2016). Cryptomyrus: a new genus of Mormyridae (Teleostei, Osteoglossomorpha) with two new species from Gabon, West-Central Africa. *ZooKeys*, 561, 117.

Taverne, L. (1972). Ostéologie des genres *Mormyrus* Linné, *Mormyrops* Müller, *Hyperopisus* Gill, *Isichthys* Gill, *Myomyrus* Boulenger, *Stomatorhinus* Boulenger et *Gymnarchus* Cuvier: considérations générales sur la systématique des poissons de l'ordre des Mormyriformes. *Musée royal de l'Afrique centrale—Tervuren, Belgique Annales—Serie in-8°—Sciences Zoologiques,* 200, 1– 198.

Taverne, L. (1978). Osteologie, phylogénèse et systématique des Téléostéens fossiles et actuels de super–ordre des Ostéoglossomorphes. Deuxième partie. Ostéologie des genres *Phareodus*, *Phareoides*, *Brychaetus*, *Musperia*, *Pantodon*, *Singida*, *Notopterus*, *Xenomyst us* et *Papyrocranus*. *Mémoires de la Classe des Sciences, Académie Royale de Belgique, Collection in–8°, 2e Série* 42, 1– 213.

Taverne, L. (1998). Les ostéoglossomorphes marins de l'Éocène du Monte Bolca (Italie): *Monopteros* Volta 1796, *Thrissopterus* Heckel, 1856 et *Foreyichthys* Taverne, 1979. Considérations sur la phylogénie des téléostéens ostéoglossomorphes. *Studi e Ricerche sui Giacimenti Terziari di Bolca, Miscellanea Paleontologica* 7, 67– 158.

Taverne, L. (2009). New insights on the osteology and taxonomy of the osteoglossid fishes *Phareodus*, *Brychaetus* and *Musperia* (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* 79, 175– 190.

Taverne, L. (2011). Ostéologie et relations phylogénétiques de *Steurbautichthys* (‘*Pholidophophorus*’) *aequatorialis* gen. nov.(Teleostei,‘Pholidophoriformes’) du Jurassique

moyen de Kisangani, en République Démocratique du Congo. *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique: Sciences de la Terre*, 81, 129-173.

Taverne, L. (2016). *Chanopsis lombardi* (Teleostei, Osteoglossiformes) from the continental Lower Cretaceous of the Democratic Republic of Congo. Comments on the evolution of the caudal skeleton within osteoglossiform fishes. *Geologica Belgica* 19, 291– 301.

Taverne, L., Kumar, K. & Rana, R. S. (2009). Complement to the study of the Indian Paleocene osteoglossid fish genus *Taverneichthys* (Teleostei, Osteoglossomorpha). *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Sciences de la Terre* 79, 155– 160.

Turner, A. H., Pritchard, A. C., & Matzke, N. J. (2017). Empirical and Bayesian approaches to fossil-only divergence times: a study across three reptile clades. *PLoS One*, 12, e0169885.

Upchurch, P., Hunn, C. A., & Norman, D. B. (2002). An analysis of dinosaurian biogeography: evidence for the existence of vicariance and dispersal patterns caused by geological events. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269, 613-621.

Van Neer, W. (1994). Cenozoic fish fossils from the Albertine Rift Valley in Uganda. Geology and Palaeontology of the Albertine Rift Valley, Uganda–Zaire, II: Paleobiology. *CIFEG Occasional Publication* 29, 89– 127.

Wallace, A. R. (1876). *The geographical distribution of animals; with a study of the relations of living and extinct faunas as elucidating the past changes of the Earth's surface.* Harper & Brothers, New York.

Wilkinson, M. (1996). Majority-rule reduced consensus trees and their use in bootstrapping. *Molecular Biology and evolution*, 13, 437-444.

Wilson, M. V. H., & Murray, A. M. (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.

Xu, G. H. & Chang, M. (2009). Redescription of †*Paralycoptera wui* Chang. & Chou, 1977 (Teleostei: Osteoglossoidei) from the Early Cretaceous of eastern China. *Zoological Journal of the Linnean Society* 157, 83– 106.

Yabumoto, Y. (2008). A new Early Cretaceous osteoglossomorph fish from Japan, with comments on the origin of the Osteoglossiformes. In *Mesozoic Fishes 4 - Homology and Phylogeny* (eds G. Arratia, H.-P. Schultze and M. V. H. Wilson), pp. 217– 228. Verlag Dr. F. Pfeil, München.

Zhang, J.-Y. (1998). Morphology and phylogenetic relationships of †*Kuntulunia* (Teleostei: Osteoglossomorpha). *Journal of Vertebrate Paleontology* 18, 280– 300.

Zhang, J.-Y. (2010). Validity of the osteoglossomorph genus †*Asiatolepis* and a revision of †*Asiatolepis muroii* (†*Lycoptera muroii*). In *Origin and Phylogenetic Interrelationships of Teleosts* (eds J. S. Nelson, H.-P. Schultze and M. V. H. Wilson), pp. 239– 249. Verlag Dr. F. Pfeil, München.

Zhang, J. Y. (2020). A new species of Scleropages (Osteoglossidae, Osteoglossomorpha) from the Eocene of Guangdong, China. *Vertebrata PalAsiatica*, 58, 100-119.

Zhang, J.-Y. & Jin, F. (1999). A revision of †*Tongxinichthys* Ma 1980 (Teleostei: Osteoglossomorpha) from the Lower Cretaceous of northern China. In *Mesozoic Fishes 2 – Systematics and Fossil Record* (eds G. Arratia and H.-P. Schultze), pp. 385– 396. Verlag Dr. F. Pfeil, München.

Zhang, J.-Y. & Wilson, M. V. H. (2017). First complete fossil *Scleropages* (Osteoglossomorpha). *Vertebrata PalAsiatica* 55, 1– 23.

CHAPTER 6

Conclusions

Although by no means a complete treatment of the evolutionary and biogeographic history of bonytongue fishes, the work in this Dissertation constitutes a significant advancement in our knowledge of the deep-time evolution of this charismatic group of freshwater fishes. More importantly, it features bonytongue fishes as a case study to demonstrate how paleontological data derived from fossils provide a wide array of information that can be integrated in several ways into quantitative and qualitative analyses aimed at biogeographic reconstruction. The two descriptive chapters (Chapters 3 and 4) exemplify how different kinds of fossil can provide different types of information—all of them important to consider. In Chapter 3, a fragmentary fossil that can only be assigned to a family-level taxon provides new fundamental data about the temporal and geographic distribution of that taxon. While fragmentary specimens are often difficult to include in phylogeny-based comparative analyses (due to a low number of preserved morphological characters), they are of great value for analyses such as the one conducted in Chapter 2, where temporal and geographic data from fossil occurrences can be used to test biogeographic hypotheses. In Chapter 4, a more complete, three-dimensionally preserved, and articulated fossil adds key morphological information that are used to derive new phylogenetic hypotheses. This kind of data is fundamental for phylogeny-based comparative approaches such as the one employed in Chapter 5. Thus, overlooking fossil data—no matter their state of preservation or degree of completeness—can result in loss of information that diminishes the power and interpretative value of downstream evolutionary analyses.

While bonytongues have an exceptionally diverse fossil record compared to other extant freshwater fish groups, including fossil data in a more comprehensive assessment of biogeographic history might be possible—if not desirable—for other taxa as well. Cyprinodontiforms (killifishes and allies), cypriniforms (carps, minnows, loaches, and allies) and anabantiforms (snakeheads, gouramies, and allies) come to mind as prominent examples, as both have been the subject of historical biogeography studies and are known from several articulated

fossil taxa with complex geographic distributions, yet fossils of these organisms have been rarely integrated with neontological data in a 'total-evidence' approach to biogeography.

I believe that the main findings of Chapter 5 demonstrate that there is an additional way in which fossils can influence the reconstruction of historical biogeography of living organisms beyond the ones listed by Grande (1985)—which can be summarized in age and geographic distribution. The example of marine bonytongues shows that fossils can also demonstrate past ecologies (either through their morphological features or through their depositional environment and taphonomy) that radically change our way of seeing the biology of their extant relatives, which in turn influences the kind of hypotheses that we formulate and test to explain observed patterns, including biogeographic ones. The inclusion of fossil data in biogeographic analyses is synergistic with recently developed models that include trait-dependent dispersal, distance-dependent dispersal, and time-stratified parameters. Several future developments in historical biogeography might help to increase such synergy. Biogeographic models that explicitly account for area-specific fossil preservation potentials will be needed to distinguish whether biogeographic patterns revealed by the fossil record are spurious representation of differential preservation and sampling effort across geographic areas, or whether they genuinely reflect biogeographic history. Integration between phylogeny-based models used in historical biogeography and ecological models that employ paleoclimatic and paleoenvironmental data (e.g., ecological niche modeling) is another promising yet mostly unexplored research avenue that will be paramount to put biogeographic processes such as dispersal and vicariance in the context of a dynamic, ever-changing Earth system (see Landis *et al*., 2021 for an example of this approach).

The case of marine fossil bonytongues completely overthrowing what we knew about the biogeography, evolutionary biology and ecology of these organisms based on their living representatives is a cautionary tale that extends beyond the scope of biogeography and embraces the study of macroevolution as a whole. Whenever the ancestral state of any kind of trait (e.g., morphological, ecological, geographic) is estimated, the result will always be some kind of phylogenetically weighted average that is strictly bound by the range of observed states. This issue is particularly marked in the case of continuous traits, such as linear measurements, shapes, and dietary indexes. Although recognizing that these approaches are most often the best that can

be applied to available data, we should always be careful when interpreting and weighing this kind of results, especially when derived from taxonomic samples only including extant species.

I would like to spend some final words about the near future of evolutionary biology as a research field. Recent technological and theoretical advancements in sequencing and computational techniques have made molecular sequences of any kind of organism relatively easy to obtain and process for researchers. Indeed, it is not too unrealistic to think that in the next few decades we will have access to a substantial portion of the genome of most—if not all—known living organisms. This massive amount of data has had and will have an enormous impact on systematics—including the resolution of a substantial portion of the 'Tree of Life'—and on the study of evolutionary processes at a molecular level. Yet, these data alone are incapable of elucidating evolution at an organismal level. As such, it is concerning that the exponential increase in genotypic information has not been paralleled by a similar increase in available phenotypic data. Although current inter-institutional projects such as oVert ('open Vertebrate' project) are going to collect and make freely available an unprecedented amount of anatomical data in the form of computed tomography-generated digital specimens, these are not ready-to-use for evolutionary analyses in the same way as gene sequences are. Converting the complex morphology of a specimen into strings of discrete and/or continuous characters that are required for analytical approaches requires careful, time-consuming descriptive and comparative work carried on by specialists of a given taxon. While such work provides the raw data for most macroevolutionary analyses—including those that form the core of high-impact papers published in top-tier journals, the primary product of this type of research—morphologically descriptive papers—is often relegated to relatively obscure journals with limited circulation and low citation potential. Because of this, early-career researchers in evolutionary biology are not incentivized to focus on descriptive work, and need instead to rely on published literature and existing databases to gather part of the input data necessary for analytical studies. However, for many extant groups of organisms the descriptive literature on their morphology goes back to the pre-phylogenetic era of biology, and the internal (including skeletal) anatomy of most species has never been described at all. As the focus of modern evolutionary systematics shifts from figuring out the relationships between different organisms to understanding how and why specific phenotypes evolve in relation to surrounding environmental conditions, we have more need of high-quality morphological descriptions and comparative anatomical studies than ever before. No analytical

approach can yield meaningful insight into phenotypic macroevolution if it is based on highly incomplete or—even worse—incorrect data. Thus, I believe that a renewed interest and push towards descriptive anatomical studies (not only for newly discovered fossil and extant species, but especially for species that have been known since decades or centuries ago) will be a necessary step to keep up with the enormous amount of genomic data and ask relevant questions about phenotypic evolution at different timescales, and to enter a truly integrative era of macroevolutionary biology.

REFERENCES

Grande, L. (1985). The use of paleontology in systematics and biogeography, and a time control refinement for historical biogeography. *Paleobiology*, 11, 234-243.

Landis, M., Edwards, E. J., & Donoghue, M. J. (2021). Modeling phylogenetic biome shifts on a planet with a past. *Systematic biology*, 70, 86-107.

APPENDIX A

Character Definitions and Morphological Matrix from Chapter 4

Character definitions

This character list is a modification of the Murray *et al*. (2018) character list, with the addition of characters 89–96 and changes to preexisting characters detailed in Chapter 4.

(1) Temporal fossa

0 = absent; 1 = present, with the exoccipital making a contribution to the border; 2 = present, bordered by epioccipital, pterotic and parietal; 3 = present, bordered by epioccipital and pterotic

(2) Shape of extrascapular

0 = expanded; 1 = reduced and irregularly shaped; 2 = reduced and tubular

(3) Shape of frontal bones

0 = anterior margin narrower than posterior margin; 1 = anterior margin about equal in width to posterior margin; 2 = anterior margin wider than posterior margin

(4) Supraorbital shelf of frontal bone

0 = absent; 1 = present

(5) Length of frontal bone

0 = over twice as long as parietal; 1 = less than twice as long as parietal

(6) Relationship of nasal bones

0 = some part separated by anterior portion of frontals; 1 = separated only by ethmoid bones; 2 = meet each other in midline

(7) Nasal bones

0 = tubular but not curved; 1 = tubular and strongly curved; 2 = gutter-like; 3 = flat and broad

(8) Parasphenoid teeth

0 = absent; 1 = small; 2 = large and found along the length of the parasphenoid; 3 = large and restricted to the basal portion of the parasphenoid

(9) Basipterygoid process

0 = absent; 1 = present

(10) Supratemporal commissure passing through the parietals

0 = absent; 1 = present

(11) Supraorbital sensory canal

0 = ending in parietal; 1 = ending in frontal

(12) Orbitosphenoid

0 = present; 1 = absent

(13) Basisphenoid

0 = present; 1 = absent

(14) Basioccipital process of the parasphenoid

0 = divided; 1 = median

(15) Ventral occipital groove

0 = present; 1 = absent

(16) Intercalar

0 = present; 1 = absent

(17) Foramen/foramina for anteroventral lateral line nerve plus cranial nerve V

0 = in the prootic; 1 = straddling the suture between the prootic and pterosphenoid; 2 = straddling the suture between the sphenotic and pterosphenoid; 3 = foramina separate from each other, one

straddling the suture between the prootic, sphenotic and the pterosphenoid (dorsally) and one straddling the suture between the prootic, pterosphenoid and parasphenoid (ventrally)

(18) Suture between the parasphenoid and sphenotic

0 = absent; 1 = present

(19) Foramen for cranial nerve VI

0 = opens within the prootic bridge; 1 = opens anterior to the prootic bridge

(20) Supraorbital bone

0 = present; 1 = absent

(21) Otic and supraorbital sensory canal

0 = in bony canals; 1 = partially or completely in grooves

(22) Number of bones in the infraorbital series, not including the dermosphenotic or the antorbital if present

0 = five; 1 = four

(23) First infraorbital

0 = does not contribute or only partially contributes to the anterior margin of the orbit; 1 = is the only bone that contributes to the anterior margin of the orbit

(24) Condition of the infraorbital sensory canal in at least some infraorbitals

0 = enclosed in a bony canal; 1 = open in a gutter

(25) Palatoquadrate area behind and below the orbit

0 = not completely covered by the infraorbitals; 1 = completely covered by infraorbitals

(26) Dermosphenotic

0 = flattened, plate-like; 1 = triradiate; 2 = tubular

(27) Posterior extent of the fossa on the neurocranium for the hyomandibula

0 = formed of pterotic; 1 = formed of pterotic and intercalar; 2 = formed of pterotic and exoccipital; 3 = formed of exoccipital and intercalary

(28) Neurocranial heads of the hyomandibula

0 = one head or two heads but continuous; 1 = two heads, separate; 2 = two heads, bridged

(29) Anterior process (wing) of the hyomandibula that contacts the entopterygoid

0 = absent; 1 = present

(30) Bones of palatoquadrate

0 = two lateral elements; 1 = one lateral element; 2 = one element, laterally and medially

(31) Autopalatine bone

0 = present; 1 = absent

(32) Preopercular sensory canal

0 = opens by pores the entire length of the canal; 1 = opens by pores ventrally and by a groove dorsally; 2 = opens by pores dorsally and a groove ventrally; 3 = opens by a groove the entire length of the canal

(33) Opercle depth to width ratio

0 = less than two; 1 = about two or greater than two

(34) Posterodorsal spine on the opercle

0 = absent; 1 = present

(35) Subopercle bone

0 = large and ventral to the opercle; 1 = small and anterior to the opercle; 2 = absent

(36) Gular bone

0 = present; 1 = absent

(37) Ascending process of the premaxilla

0 = well developed; 1 = only slightly developed if at all

(38) Premaxillae

0 = paired; 1 = median

(39) Posterior portion of maxilla

0 = lies on angular; 1 = lies on dentary

(40) Supramaxillae

0 = present; 1 = absent

(41) Mandibular canal

0 = enclosed in a bony tube; 1 = open in a groove

(42) Posterior bones of the lower jaw

0 = angular and retroarticular bones fused; 1 = angular and articular bones fused; 2 = all separate; 3 = all fused

(43) Retroarticular bone

0 = included in the articulation with the quadrate; 1 = excluded from the articulation with the quadrate

(44) Medial wall of the Meckelian fossa of the lower jaw

0 = present; 1 = absent

(45) Bony elements associated with the second ventral gill arch

0 = absent; 1 = present as autogenous elements; 2 = present as a bony process on the second hypobranchial

(46) Toothplates associated with basibranchial 4

0 = present; 1 = absent

(47) Basihyal toothplate

0 = present; 1 = absent

(48) Basihyal toothplate

0 = flat; 1 = with ventrally directed processes

(49) Basibranchial toothplate and basihyal toothplate

0 = separate; 1 = continuous

(50) Basihyal

0 = present and ossified; 1 = present and cartilaginous; 2 = absent

(51) Hypohyals

0 = two ossified pairs present; 1 = one ossified pair present; 2 = one ossified pair present but greatly reduced in size

(52) Infrapharyngobranchial 3

0 = undivided; 1 = divided into two elements

(53) Infrapharyngobranchial 1

0 = present; 1 = absent

(54) Orientation of infrapharyngobranchial 1

0 = proximal tip anteriorly directed; 1 = proximal tip posteriorly directed

(55) Abdominal scutes

0 = absent; 1 = present as paired structures

(56) Epipleural bones

0 = absent; 1 = only a few bones in anterior caudal region; 2 = present throughout abdominal and caudal region

(57) Dorsal arm of the post-temporal bone

0 = less than 1.5 times as long as the ventral arm; 1 = more than twice as long as the ventral arm

(58) Lateral line that pierces the supracleithrum

0 = present; 1 = absent

(59) Cleithrum

0 = with no or only a slight medial lamina; 1= with a broad medial lamina

(60) Coracoid fenestra

0 = absent; 1 = present

(61) First pectoral fin ray

0 = normal; 1 = greatly expanded

(62) Post-pelvic bone

0 = absent; 1 = present

(63) Pelvic bone

0 = slender; 1 = possesses a thin deep lamella in dorsoventral plane

(64) Posterior end of anal fin

0 = separate from caudal fin; 1 = continuous with caudal fin

(65) Number of principal caudal fin rays

0 = 19 or more; 1 = 18; 2 = 17 or fewer

(66) Uroneurals

0 = three or more; 1 = two or one; 2 = absent

(67) Neural spine on ural centrum 1

0 = absent or rudimentary; 1 = one or more

(68) Epurals

0 = two or three; 1 = one; 2 = absent

(69) Neural spine on the first preural centrum

0 = complete; 1 = rudimentary; 2 = absent

(70) Number of neural spines on the second preural centrum

0 = one; 1 = two

(71) Number of hypurals

0 = seven; 1 = six or fewer

(72) Scales

0 = no reticulate furrows; 1 = both radial and reticulate furrows present; 2 = reticulate furrows only present over entire scale

(73) Pelvic fin ray number

0 = more than seven; 1 = seven; 2 = six or fewer

(74) Swimbladder-ear direct connection

0 = absent; 1 = present

(75) Intestine

0 = coils to right of esophagus and stomach; 1 = coils to left of esophagus and stomach

(76) Opercle shape dorsal to facet for articulation with hyomandibula

0 = rounded; 1 = flattened or truncated; 2 = flattened with posterior recurved process

(77) Upper hypurals and second ural

0 = not fused; 1 = fused

(78) Second infraorbital shape and size

0 = more or less slender or tubular and small in size; 1 = triangular or rectangular and smaller than third infraorbital; 2 = expanded and equivalent in size to or larger than third infraorbital

(79) Dorsal fin shape

0 = base moderately long, fin triangular or falcate; 1 = base very short, much shorter than fin height, or fin absent; 2 = base moderately long to very long, fin with rounded outline anteriorly and posteriorly

(80) Posterior rays of dorsal and anal fin

0 = shorter than anterior ones; 1 = longer than or as long as anterior ones

(81) 'Cheek wall' formed by enlargement of first to third infraorbitals

0 = absent; 1 = present

(82) Ventral part of preopercle

0 = extending anteriorly to beneath orbit or to level of posterior edge of orbit; 1 = anteriorly does not reach level of orbit

(83) Posterior edge of nasal when it is gutter-like or irregularly subrectangular

0 = straight or slightly curved; 1 = strongly curved and extending backward

(84) Angle of jaws

0 = anterior to middle vertical line of orbit; 1 = between middle vertical line and posterior edge of orbit; 2 = behind orbit

(85) Utriculus

0 = connected with sacculus and lagena; 1 = completely separated from sacculus and lagena

(86) Anal fin sexual dimorphism

0 = absent; 1 = present

(87) Ventral margin of opercle

0 = rounded or pointed and narrower than mid-point of opercle; 1 = curved but not greatly narrowed compared to midpoint of opercle; 2 = flattened or only very slightly rounded

(88) Parapophysis on the first centrum

0 = not expanded or hypertrophied; 1 = expanded and rounded, barely reaching below the occiput and not touching the parasphenoid; 2 = greatly hypertrophied and extending anteriorly to touch the parasphenoid, wedge-shaped in lateral view

(89) Dorso-occipital fossa

0 = absent; 1 = present

(90) Contact between dermosphenotic and anteriormost bone of the infraorbital series

0 = absent; 1 = present

(91) Depth of dorsal posterior infraorbital compared to ventral posterior infraorbital

0 = shallower; 1 = equal; 2 = deeper

(92) Scleral ossicles

0 = absent; 1 = present

(93) Postero-dorsal flange of the angular

0 = absent; 1 = present

(94) Posterior process of the hyomandibula

0 = short (less than half the length of the dorsal articulating surface of the hyomandibula); 1 = long (more than half the length of the dorsal articulating surface of the hyomandibula); 2 = absent or extremely reduced

(95) Endopterygoid dentition

0 = patch of shagreen-like fine teeth or small conical teeth; 1 = few rows of large conical teeth; 2 = one or more medio-dorsal rows of large conical teeth, bordered laterally by a patch of shagreen-like fine teeth; 3 = teeth absent or extremely reduced

(96) Number of branchiostegal rays

0 = 8 or less; 1 = between 9 and 13; 2 = 14 or more

Morphological matrix

Amia
0?10123110100000001?10010000000000000001002110??? 2100000000000000210000000001020001?2102000(0 1)01001

Ellimmichthyiformes
3?000100111011100000?????0000?0100110000 1????????1??0110000000001001010 1??00?00?0?0?00?00?100(0 2 3)(0 1)

Dorosoma
2000010001101010000000000000000000011010011111 1??10001110100000001001010 11?0000000?1?00000110030

Elops
1000000100100000000000000001000000001000000000000000000200000000000010000000?00011?2?00000010002

Lycoptera
??00110200000????????000000?0?0?0000000000??1??000?????0000??0?0010(0 1)100002??0010000?0?02?00?00001

Paralycoptera 0000(0 1)2321000???????1010010?00??2100100000111??????1???000??01?00201200101???00000001??1?00200101

Sinoglossus
??00?2????10???????1?11??0?????000??0?110??????????????0???????2????0??2??0?1211000??1??010????

Eohiodon
200011120010??00???1010001?200?00101100100 01??010?0???0010000?001001(0 1)0001??2010000?0?12?002?01?(0 1)

Hiodon 2000111200(0 1)0000000010100011200100101100100010001000000001000010010 01(0 1)(0 1)00111201000?001200001010(0 1)

'Joffrichthys_symmetropterus'??100132?1100??0???10110010????100?100102?????? ????000?000?001?1200?02??0?10010?1??000?100??0

'Joffrichthys_tanyourus' 20100132?0100??0???10?00001001?2100?100102?????????000?000?00101200002??00?00 10?1??0?0?0000?0

Lopadichthys 10100002?010???????10?000001????00??100?011???????????000?0?0??01102001 02??001001 0?0??0?00?000?1

Chauliopareion 2010003?1100???????101000000???2101?10110?1???????????0010?110102????0102??0?1?00 000?11??010000(0 1)

Shuleichthys 3010010200100??0?0?11???0?100??(0 1)0001?0?10210??00?01???001??0010010(0 1)100001??00?0000?1??2?0??00001

Wilsonichthys 2?100?00001????????11?0??0??0??0100?10110?????00??????0010??0?00110111001??00????0 ?0?????0?0???0

Xixiaichthys 0?100132??000??????1?1000??00??0102100010 2????0???0???001?0?0?00001200001??001000 002??1?0??000?1

Heterotis 22201230111010100001011010011100001110110200211??11100010101000021120012201011 21100000 2?00201010

Arapaima 32001232111010101011011010011110001100110200201??21100000110000021120012201111 211100001201201101

Phareodus ?22101331010?100???10100100 11?02101100010210??001?????000??11?002?1?00122??00121 0112?01210201121

Pantodon
2210113210101110100110001000011010210101010020001110 1?00100110102112001120101110011201000000011

Singida
2210103010000???????10100100???2102???010?1???????????0000?11010211200102??0010001 02?00100100?31

Scleropages 32100233111110000101010010001 1(0 1)2101110010110210011100000010110102112001220101021011200010110102 2

Osteoglossum
3210023311111000010101001000111210111001011021001110000001011010211200122010102101112000101000021

Petrocephalus
1000002210100011211111010220021000011111 13?021000210010 0?011000011020011?????20????????000200230

Gnathonemus
1000102210101011311111010220021000011111 13?1211??2200100?011000011020011??1???0????????000100230

Campylomormyrus
1000002210101011211111010220021000011111 13?1211??2200100?011000011020011?????20????????000100230

Chitala
1120022101100100101111010030011200210001 1110110101101?10010100012212001 0??11?21??1????0000100000

Xenomystus
1110022201100100101111010230011200210001 1110110101101?10010100012212001 0??11?21??0????0000200030

Papyrocranus

1110022201101100101111010230011200210001111011010110001001010 0?122120(0 1)10??11?21??0????0000100100

Palaeonotopterus

1000???200100?00110?1?????000????????????????????????????????1????????????????????????????? ??00??0????

Macroprosopon

?211?????11????????1?10010?1???????1?0010200????????????1???1??????????0???1?0??01?2 ???210211?2

Furichthys

??????3????????0??????????0?11??2000?10??021???????????0??1?11??????????0???0?????1??? ?2??0???1??

Brychaetus

??2101301??0100000010100101111???10?1100?0???21????????0?????????????????????0????01 11??0210(1 2)11102

Laeliichthys

?1100232?1100??????1110110?00??200210001111???0???????000101100011120010 2??1?200 00????0?00000?01

APPENDIX B

GenBank Accession Numbers, Morphological Matrix, and Taxon Sampling from Chapter 5

Table B.1. GenBank accession numbers of DNA sequences included in the total-evidence dataset

Genus	Species	co1	cytb	12srrna	16srrna	rag2	rag1	rhod	glyt	plagl2	sreb2	zic1	sh3px3
Amia	*calva*	KX145442	NC_004742	AY442347.RR1	AY442347.RR2	AF369079	JX190802	MN519151	EF032987	EF033013	JX191055	EF032909	EF033000
Arapaima	*gigas*	MH830243	MH830293	AY504824	EF523611.RR2	AY504843	JN230877	JN230972	JX190255				JX190970
Boulengeromyrus	*knoepffleri*	AP011568.COI	AF201573	AF201483	AF201528	AF201616							
Brevimyrus	*niger*	3281550*	KT820038	AF201487	AF201532	AF201620							
Brienomyrus	*brachyistius*	MK073996	AF201574	AF201484	AF201529	AF201617							
Campylomormyrus	*compressirostris*	HG779437.COI	KJ713961	HG779437.12S-RRNA	HG779437.16S-RRNA								
Campylomormyrus	*numenius*	KF036202	DQ231116	AF201490	AF201535	AF201622							
Campylomormyrus	*tamandua*	KT193439	KJ713963	AF201492	AF201537	AF201625							
Chitala	*blanci*	NC_012710	NC_012710	AP008921.RR1	AP008921.RR2								
Chitala	*chitala*	MK572123	GQ476731	KX894524.RR1	JF300361		FJ896406	KY026029	JX190259	JX190546	JX191060	FJ906625	

Chitala	*lopis*	NC_012711	NC_012711	AP008922.RR1	AP008922.RR2								
Chitala	*ornata*	MK049507	AB035243	AF201493	AF201538	AF201626		MG584732	JX190260	JX190547	JX191061	JX191261	JX190932
Cryptomyrus	*ogoouensis*	KT875221	KT875226	KT875213	KT875217	KT875230							
Cryptomyrus	*ona*	KT875223	KT875228	KT875215	KT875219	KT875232							
Cyphomyrus	*discorhynchus*	MN207907	AF201587	MN255576	AF201542	AF201630							
Dorosoma	*cepedianum*	131275 87*	MG570459.CYTB	MG570459.RR1	DQ912062	MG958284	MG958450	KX145707	EU002043		EU002121	EU001867	
Elops	*saurus*	403882 2*	GQ183882	AF454711	KC146867		FJ896408	MH796447	JX190247	EU002097	EU002123	FJ906627	EU002068
Genyomyrus	*donnyi*	NC_015086	NC_015086	AF201494	AF201539	AF201627							
Gnathonemus	*petersii*	HM882766	DQ231098	AF201495	AF201540	AF201628	FJ896407	KY982945	JX190258	JX190545		FJ906626	
Gymnarchus	*niloticus*	HM882950	AF201586	AF201496	AF201541	AF201629	JX190806	KY982943	JX190256	JX190542	JX191057	JX191256	JX190929
Heterotis	*niloticus*	HM882702	AY504820	FJ890318.RR1	MH767416	AY504842			JX190262		JX191063		JX190934
Hiodon	*alosoides*	KX145583	AY504821	AY430248	AY504835	AY504841	AY430200	MH796542		EU002095	EU002120	EU001866	EU002066
Hiodon	*tergisus*	EU524661	NC_015082	AP009499.RR1	AP009499.RR2			KX146110	JX190257	JX190544	JX191059	JX191258	JX190931
Hippopotamyrus	*castor*	KT875220	KT875224	KT875212	KT875216	KT875229							
Hippopotamyrus	*pictus*		KT820017	AF201498	AF201543	AF201631							
Hyperopisus	*bebe*	MG824600	KT820040	AF201500	AF201545	AF201633							
Isichthys	*henryi*	AP011573.COI	AF201590	AF201501	AF201546	AF201634							

Ivindomyrus	*marchei*	AP011574.COI	DQ166679	AP011574.RR1	AP011574.RR2							
Ivindomyrus	*opdenboschi*		DQ166690	AF201502	AF201547	AF201635						
Marcusenius	*greshoffii*	KT193430	AF201594	AF201504	AF201549	AF201637						
Marcusenius	*moorii*	444820*	MK119181	AF201505	AF201550	AF201638						
Marcusenius	*senegalensis*	HM882735	NC_015090	AF201506	AF201551	AF201640						
Mormyrops	*anguilloides*	KT193437	AF095293	AP011576.RR1	AP011576.RR2		JN230878	JN230973				
Mormyrops	*masuianus*	444820*	AF201597	AF201508	AF201553	AF201641						
Mormyrops	*nigricans*	444806*	AF201598	AF201509	AF201554	AF201642						
Mormyrops	*zanclirostris*	KT193168	AF201599	AF201510	AF201555	AF201643						
Mormyrus	*ovis*		AF201600	AF201511	AF201556	AF201644						
Mormyrus	*rume*	AP011577.COI	AF201601	AF201512	AF201557	AF201645						
Myomyrus	*macrops*	KT193467	AF201602	AF201513	AF201558	AF201646						
Notopterus	*notopterus*	579658*	AY504822	AF508062	KT878235	AY504845	AF369063					
Osteoglossum	*bicirrhosum*	HM156437	AB035238	FJ549357	KX816045	AY504838	AY430201	JX255591	JX255648	EU002111	EU002142	EU001887
Osteoglossum	*ferreirai*	465019*	AB035239	X99172	X99171							
Pantodon	*buchholzi*	444818*	GU997128	AF201527	AF201572	AF201647	AF369061	KY982942				
Papyrocranus	*afer*	JF510515	AY504823	AY504826	AY504836	AY504844						

Papyrocranus	*congoensis*	KT193311	NC_012714	AP008926.RR1	AP008926.RR2					
Paramormyrops	*batesii*		AF201578	AF201488	AF201533	AF201621				
Paramormyrops	*gabonensis*	NC_015107	AF201603	AF201514	AF201559	AF201648				
Paramormyrops	*hopkinsi*		AF201575	AF201485	AF201530	AF201618				
Paramormyrops	*kingsleyae*	KT192764				XM_023842223	XM_023842067			XM_023810506
Paramormyrops	*longicaudatus*		AF201576	AF201486	AF201531	AF201619				
Petrocephalus	*bovei*		GU982922	AF201516	AF201561	AF201650				
Petrocephalus	*microphthalmus*	KX186413	EU770190	AP009609.RR1	AP009609.RR2					
Petrocephalus	*simus*	MK074582	EU770196	AF201515	AF201560	AF201649				
Petrocephalus	*soudanensis*	NC_015092	NC_015092	AF201518	AF201563	AF201652		KY982944		
Petrocephalus	*sullivani*		EU770180	AF201517	AF201562	AF201651				
Pollimyrus	*adspersus*	KT193313	AY124316	AY124317	AY124314	AY124315				
Pollimyrus	*isidori*	KX186560	KT820034	AF201519	AF201565	AF201654				
Pollimyrus	*petricolus*		AF201608	AF201520	AF201566	AF201653				
Scleropages	*formosus*	HM156403	NC_007012			XM_018765257	XM_018765326	XM_018762271	XM_029249684	XM_018753632
Scleropages	*jardinii*	KY123529	AB035236	KF481952.RR1	KX816044			KY026031		
Scleropages	*leichardti*	KJ669635	FJ890319.CYTB	AY504829	AY504832					

Stomator	*ivindoens*		AF20161	AF20152	AF2015	AF20165							
hinus	*is*		2	3	68	8							
Stomator		MK0746	AF20161	AF20152	AF2015	AF20165							
hinus	*walkeri*	46	0	2	67	6							
Xenomys		444806	AF20161	AP00892	AP0089	AF20166	KY98294	JX19	JX190	JX191	JX191	JX190	
tus	*nigri*	7*	4	7.RR1	27.RR2	0	1	0261	548	062	262	933	

Morphological matrix for the total-evidence analysis of Chapter 5. Character definition same as in Chapter 4 (Appendix A)

Amia_calva
0?10123110100000001?10010000000000000001002110??? 2100000000000000210000000001020001?2102000{0 1}01001

Ellimmichthyiformes
3?000100111011100000?????0000?0100110000 1????????1??0110000000001001010 1??00?00?0?0?00?00?100{0 2 3}{0 1}

Dorosoma_cepedianum
2000010001101010000000000000000000011010011111 1??10001110100000001001010 11?0000000?1?00000110030

Elops_saurus
1000000100100000000000000001000000001000000000000000000200000000000010000000?00011?2?00000010002

Lycoptera
??00110200000????????000000?0?0?0000000000??1??000?????0000??0?0010{0 1}100002??0010000?0?02?00?00001

Paralycoptera 0000{0 1}2321000???????1010010?00??2100100000111??????1???000??01?00201200101???00000001??1?00200101

Sinoglossus
??00?2????10???????1?11??0?????000??0?110??????????????0???????2????0??2??0?1211000??1??010????

Eohiodon
200011120010??00???1010001?200?00101100100 01??010?0???0010000?001001{0 1}0001??2010000?0?12?002?01?{0 1}

Hiodon_alosoides 2000111200{0 1}0000000010100011200100101100100010001000000001000010010010{0 1}{0 1}0011120100000?001200001010{0 1}

Hiodon_tergisus 2000111200{0 1}0000000010100011200100101100100010001000000001000010010010{0 1}{0 1}0011120100000?001200001010{0 1}

Joffrichthys_symmetropterus ??100132?1100??0???10110001000??100?1001021???????????000?000?001?1200?02??0?10010?1??000?100??0

Joffrichthys_tanyourus 20100132?0100??0???10?00001001?2100?1001021???????????000?000?00101200002??00?0010?1??0?0?0000?0

Lopadichthys 10100002?010???????10?000001????00??100?011???????????000?0?0??01102001020??0010010?0??0?00?000?1

Chauliopareion 2010003?1100???????101000000???2101?10110?1???????????0010?110102????0102??0?1?00000?11??010000{0 1}

Shuleichthys 3010010200100??0?0?11???0?100??{0 1}0001?0?10210??00?01???001??0010010{0 1}100001??00?0000?1??2?0??00001

Wilsonichthys 2?100?00001????????11?0??0??0??0100?10110?????00??????0010??0?00110111001??00????0?0?????0?0???0

Xixiaichthys 0?100132??000??????1?1000??00??0102100010200??0???0???001?0?0?00001200001??0010000002??1?0??000?1

Heterotis_niloticus
22201230111010100001011010011100001110110200211??111000101010000211200122010112100000 2?00201010

Arapaima_gigas
32001232111010101011011010011110001100110200201??211000001100000211200122011112111000 01201201101

Phareodus
?2210133101 0?100???10100100 11?02101100010210??001?????000??11?002?1?00122??00121 0112?01210201121

Pantodon_buchholzi
22101132101011101001100010000110102101010100200011101?00100110102112001120101110011 2010000001011

Singida
22101030100???????10100100???2102???010?1???????????0000?11010211200102??001000 102?00100100?31

Scleropages_formosus
32100233111110000101010010001112101110010110210011100000010110102112001220101021011 2000101101022

Scleropages_leichardti
32100233111110000101010010001102101110010110210011100000010110102112000220101021011 2000101101022

Osteoglossum_bicirrhosum
32100233111110000101010010001112101110010110210011100000010110102112001220101021011 2000101000021

Petrocephalus_simus
10000022101000112111110102200210000111111 3?02100021001 00?01100001102001 1?????20????????000200230

Gnathonemus_petersii

1000102210101011311111010220021000011111113?1211??2200100?011000011020011??1???0
????????000100230

Campylomormyrus_tamandua

10000022101010112111110102200210000111111 3?1211??2200100?011000011020011?????20
????????000100230

Chitala_chitala

1120022101100100101111010030011200210001111011010110 1?10010100012212001 0??11?2
1??1????0000100000

Xenomystus_nigri

1110022201100100101111010230011200210001111011010110 1?10010100012212001 0??11?2
1??0????0000200030

Papyrocranus_afer

1110022201101100101111010230011200210001111011010110 00010010100?122120{0
1}10??11?21??0????0000100100

Palaeonotopterus

1000???200100?00110?1?????000????????????????????????????1????????????????????????
??00??0????

Macroprosopon

?211?????11????????1?10010?1???????1?0010200???????????1???1??????????0???1?0??01?2
???210211 1?2

Brychaetus

??2101301??0100000010100101 11???10?1100?0???21????????0?????????????????????0????01
11??0210{1 2}11102

Furichthys

??????3????????0?????????0?11??2000?10??021???????????0??1?11??????????0???0?????1???
?2??0???1??

Laeliichthys
?1100232?1100??????1110110?00??2002100011111???0???????000101100011120010 2??1?20000????0?00000?01

Heterosteoglossum
?2?00?3?11?????0??10????0?11??000?11???0?????????????00???11?102?1?00121??11?20?0???0??1?0????

Habib_Rahi_osteoglossid
???1???01???????????0?001??11??????1?00?0????????????????????????????????????0??0???2??????1???2

Gymnarchus_niloticus
1000112000100011211112000220021200011011 0?0211??0101?00?0?100???2?2???1???1?020000001?0000?00230

Notopterus_notopterus
1010022201100100101111010230011200210001111011010110 1?10010100012212001 02??1021?00?0?00000100000

Notopterus_primaevus
1??01?22???0??0??0??1?010??00??200?1??01111???01??????100????00????????????1??1?00?0??100??0010?

Phareoides_queenslandicus
?22101?0101??0??????0??01??111?21011??01021???????????00?1??1??02??????2???0?????1?2??0?1?20?12?

Taverneichthys_bikanericus
??2??23????0????????0??01?????0?????00010??11??????2?????

Thrissopterus_catullii
3?000??31?1????????1??001??0???000?100010?????????????00????1010210?00122??11?2010?0???????0????

Xosteoglossid_rebeccae

??100?301?10????????0??????01???10??0000?????????????????1?10??????????2???0?????10?
?????????102

Scleropages_sinensis

?210003??011???????1110010??11?200?1100101 1???????????00?1??10?02112001 22??010210
111??0?0110??22

Scleropages_sanshuiensis

?2100033 1?11???????1110010?0111200111011011 0??????????0?????1??02112001 22??010210
111??0?01201?2?

Kuntulunia_longipterus

0100110010100 0???0?1011000?0?0?300111001011???00?01???0000100000100200001??0010
000?0??1?002000?0

Asiatolepis_muroii

?1000102?000???????1010000?000?00001101000 0???????0???0010??000020{0
1}201002??000000??0??1?000000?1

Tongxinichthys_microdus

??0011?2?000??????01000000?0???00001100001 1???0???0???0010?00000100100002??00?000
0?0??1??0100?01

Rogue taxa excluded from the analysis

Tetoriichthys

?0100??31?1??00?????1??0?2?????200??????????????????????????0????????????0????????????
?0???2?????

Chanopsis_lombardi

??211?33101?????????1??0???01????0??????????????????????????1?0?????0120011???0????????
?????????0??

Musperia_radiata
?????23?????????????????????1??210???0??0?????????????00????1??02112011 2???00?00????
??1???2??1??

Comments on the three rogue taxa excluded from the analysis

The three fossil taxa that we preliminarily identified as rogues – and thus we did not include in any of our analyses– deserve a more detailed discussion about their potential phylogenetic and biogeographic relevance. †*Tetoriichthys* from the Berrasian–Hauterivian of Japan could be the oldest known crown osteoglossomorph or even osteoglossiform (Yabumoto, 2008). Its disarticulated remains display a tantalizing mix of features, some resembling pantodontids, osteoglossids and even notopteroids. The age and geographic provenance of †*Tetoriichthys* are compatible with our reconstruction of the most basal osteoglossomorph nodes as including East Asia, but a phylogenetic position deeply nested within Osteoglossiformes could significantly change some of the ancestral ranges estimated in this study.

†*Chanopsis* from the Aptian–Albian of the Democratic Republic of Congo is a large fish (skull roof length over 15 cm) known from mostly disarticulated remains (Taverne, 2016). A striking feature of †*Chanopsis* is the presence of a broad supraorbital shelf on the frontal, a feature that is uniquely found in †phareodontines among osteoglossomorphs (Chapter 4 of this Dissertation). However, other morphological characters suggest a much more basal position, such as the presence of three uroneurals in the caudal skeleton (Taverne, 2016) when all other known osteoglossiforms have two or less uroneurals. It should be noted that the fossil specimen including the caudal skeleton referred to †*Chanopsis* could potentially belong to another taxon, as it comes from a different locality than the holotype and the only comparable material between the two specimens are scale fragments. Nonetheless, †*Chanopsis* could represent the oldest known †osteoglossid, predating †*Cretophareodus* by at least 20 million years (Capobianco *et al*., 2019). Thus, any reassessment of the systematics of *Chanopsis* can strongly impact the reconstruction of phylogenetic and biogeographic history of bonytongues.

†*Musperia* from the Eocene Sangkarewang Formation of Sumatra, the third and last rogue taxon excluded from our analyses, is an osteoglossid with teeth and lower jaws similar to those of

†*Phareodus*. Unfortunately, the best-preserved specimen of this taxon was potentially destroyed during World War II and only photos of it remain (Sanders, 1934). Not enough is known from †*Musperia* to confidently place it within phareodontines, but it could expand the geographic range of this clade to the Malay Archipelago. Interestingly, the co-occurrence of †*Musperia* and †*Notopterus primaevus* shows that osteoglossids and notopterids have inhabited Sumatra since at least 35 million years ago; however, those species were probably not close relatives of the osteoglossid and notopterids that occur in the island nowadays (*Scleropages formosus*, *Notopterus notopterus*, *Chitala* spp.).

APPENDIX C

R Scripts from Chapter 5

regPhylo script for semi-automated DNA sequence extraction and alignment

```
install.packages(c("bold", "seqinr", "ape",
                   "RJSONIO", "stringr", "fields", "parallel",
                   "caper", "phytools"))
library(bold)
library(seqinr)
library(ape)
library(RJSONIO)
library(stringr)
library(fields)
library(parallel)
library(caper)
library(phytools)
library(rentrez)

# The "geomedb" requires the latest version available on Github, to
download it,
# do the following:
install.packages("devtools")
library(devtools)
install_github("biocodellc/fimsR-access")
library(geomedb)

# Install the package from GitHub
install_github("dvdeme/regPhylo")
library(regPhylo)

###################################
# Prepare the species list

# Load the species list and classification into R.
SpList.Classif    =    read.csv("Osteo_SpeciesList_Classification.csv",
h=TRUE)

# Extract the species list only.
Sp.List = SpList.Classif$SpeciesName
# Replace the "_" with a space between the genus and species name.
Sp.List = gsub("_", " ", Sp.List)
```

```
Sp.List[1:10] # Display the first 10 species.

write.table(Sp.List,     file="Sp.List_forNCBITaxo.txt",     sep="\t",
row.names=F,
            col.names=F, quote=F)

###################################
# Taxonomic checks using NCBI taxonomic database

# Go to www.ncbi.nlm.nih.gov/Taxonomy/TaxIdentifier/tax_identifier.cgi
to check the taxonomic status of species and extract their id

taxReport=read.delim("tax_report.txt",
                     sep="\t", h=T, stringsAsFactors=FALSE)

# We remove the unnecessary columns 2, 4 and 6 containing "|" as
separator.
taxReport = taxReport[,-c(2,4,6)]
head(taxReport, n= 3)

taxReport$name[which(taxReport$code==3)]

# Run the function with the path to the file "tax_report.txt" exported
by the
# NCBI taxonomic web facility as input.
SpList.DF = Taxreport2Sp.List(input = "tax_report.txt")

# Extraction of the species list for NCBI search (first object of the
list)
SpList.NCBI = SpList.DF$SpList.NCBI
head(SpList.NCBI, n = 3)

dim(SpList.NCBI) # Number of species with a taxid.

# Extraction of the species list for BOLD search (second object of the
list)
SpList.BOLD = SpList.DF$SpList.BOLD
head(SpList.BOLD, n = 3)

dim(SpList.BOLD)


###############################################
# Extract the DNA sequences and associated metadata from different
sources and assemble the data

dir.create("Data_Extraction")

##########
# Extraction from Genbank, via the NCBI platform
```

```
Seq.NCBI.info = GetSeqInfo_NCBI_taxid(splist = SpList.NCBI, gene =
"ALL",
                                      filename                    =
"Data_Extraction/Seq.NCBI_Outgroups.txt", chunk_size= 100) #Accessed
07/29/2020

# Try changing parameter chunk_size if downloading error happens.

Seq.NCBI.all = read.delim("Data_Extraction/Seq.NCBI.txt", sep = "\t",
h = TRUE)
dim(Seq.NCBI.all) # You can check the number of DNA sequences
retrieved.

#########
# Extraction from BOLD

Seq.BOLD.info = GetSeq_BOLD(splist = SpList.BOLD,
                            filename = "Data_Extraction/Seq.BOLD.txt")

Seq.BOLD=read.delim("Data_Extraction/Seq.BOLD.txt", sep="\t", h=T)
dim(Seq.BOLD) # You can check the number of DNA sequences retrieved.


#########
# Assemble the data into a single table

AllSeqDF = Congr.NCBI.BOLD.perReposit(input.NCBI = Seq.NCBI.all,
input.BOLD=Seq.BOLD,
                                      output                      =
"Data_Extraction/AllSeqDF.txt")
dim(AllSeqDF) # Number of sequences in total after removing the
duplicates

length(which(AllSeqDF$OriginDatabase == "NCBI"))
length(which(AllSeqDF$OriginDatabase == "BOLD"))
length(which(AllSeqDF$OriginDatabase == "NCBI-BOLD"))


################################################
# Improve the spatial metadata associated with the DNA sequences in
three steps

dir.create("Geolocation")

##########
# Homogenize the geographic coordinates

AllSeqDF1 = GeoCoord.WGS84(input = AllSeqDF, output =
"Geolocation/AllSeqDF_Geo1.txt")
dim(AllSeqDF1)
names(AllSeqDF1) # Confirm that Latitude (i.e. "Lat") and Longitude
(i.e. "Long") are distinct.
```

```
# Evaluate the percentage of DNA sequences with geographic coordinates
(table(AllSeqDF3$Geo_accuracy)/length(AllSeqDF3$Geo_accuracy))*100


##################################
# Build the species-by-gene matrix while removing undesirable
sequences

dir.create("SeqPool")

Sp.DNAMat_naive=SpeciesGeneMat.Bl(input=AllSeqDF3,
output="SeqPool/SpAll.DNA.Mat_Naive_")

dim(Sp.DNAMat_naive[[1]])

# DNA regions table
head(Sp.DNAMat_naive[[2]])

# Species table
head(Sp.DNAMat_naive[[3]])

# Species removed when removing microsatellites and unassigned DNA
Sp.DNAMat_naive[[4]]

# Load the table of sequences and metadata without the microsatellites
and unassigned DNA
CleanDF_Naive=read.delim("SeqPool/SpAll.DNA.Mat_Naive__CleanDataset.tx
t", sep="\t", h=T)
dim(CleanDF_Naive)
dim(AllSeqDF3)[1]-dim(CleanDF_Naive)[1]


#########
# Minimum number of gene regions maximizing the species coverage

SelGene.MaxSpCov(input=Sp.DNAMat_naive[[1]])

# Example with a pre-defined selection of genes
SelectionGenes=c("co1", "cytb", "12srrna", "16srrna", "rag2", "rag1",
"rhod", "glyt", "plagl2", "sreb2", "zic1", "sh3px3")

SelGene.MaxSpCov(input=Sp.DNAMat_naive[[1]], NBGene=SelectionGenes)

##########
# Degree of species overlap between gene regions

Mat.overlap.genes_naive=Matrix.Overlap(input=Sp.DNAMat_naive[[1]],
                                       gene.Sel=SelectionGenes)

Mat.overlap.genes_naive[[1]]
min(Mat.overlap.genes_naive[[1]])
```

```
diag(Mat.overlap.genes_naive[[1]])=NA ### Removed the diagonal.
mean(Mat.overlap.genes_naive[[1]], na.rm=T)

##########
# Amount of missing data in the species-by-gene matrix for the
selected gene regions

AmMissData(input=Sp.DNAMat_naive[[1]], gene.list=SelectionGenes)


##############################
# Export all DNA sequence of selected gene regions and associated
metadata into a single file

CleanDF_Naive=read.delim("SeqPool/SpAll.DNA.Mat_Naive__CleanDataset.tx
t", sep="\t", h=T)

Sp.Genes.DF_naive=Select.DNA(input=CleanDF_Naive,
                                  gene.list=SelectionGenes,

output="SeqPool/Sp.Genes.Dataset_Naive_", timeout = 5)

dim(Sp.Genes.DF_naive)

##############################
# Extraction of all the sequences per species for the selected gene
regions

dir.create("Alignments")
dir.create("Alignments/AllSeqNaive_Ge")

Sp.Genes.DF_naive.export=SelBestSeq(input=Sp.Genes.DF_naive,

output="Alignments/AllSeqNaive_Ge/Alig_AllSeqNaive",
                                        Alignment=T,     MaxSeq="ALL",
gene.list=SelectionGenes,
                                        SeqChoice="Median")
list.files("Alignments/AllSeqNaive_Ge")



########################################
# First alignment of all sequences and detection of sequences that
should be reverse complemented

dir.create("Alignments/AllSeqNaive_Ge/FirstAlign")

input="Alignments/AllSeqNaive_Ge"
output="Alignments/AllSeqNaive_Ge/FirstAlign"
nthread=5
methods = c("mafftfftnsi")
```

```
Mafft.path = "D:/mafft-win/mafft"
Muscle.path = "D:/muscle/muscle3.8.31_i86win32.exe"

First.Align.All(input="Alignments/AllSeqNaive_Ge",
                output="Alignments/AllSeqNaive_Ge/FirstAlign",
                nthread=5, methods = c("mafftfftnsi"),
                Mafft.path = "D:/mafft-win/mafft")
list.files("Alignments/AllSeqNaive_Ge/FirstAlign")


########################################
# Detection of potential outlier sequences in a gene region alignment

S12_MisAlign0.6_1=Detect.Outlier.Seq(

inputal="Alignments/AllSeqNaive_Ge/FirstAlign/Mafftfftns1_Alig_AllSeqN
aive_12srrna.fas",
  Strat.DistMat="Comb", Dist.Th = 0.6,
  output = "12S_outliers_1.txt",
  Second.Outlier="No")
dim(S12_MisAlign0.6_1)

S16_MisAlign0.6_1=Detect.Outlier.Seq(

inputal="Alignments/AllSeqNaive_Ge/FirstAlign/Mafftfftns1_Alig_AllSeqN
aive_16srrna.fas",
  Strat.DistMat="Comb", Dist.Th = 0.6,
  output = "16S_outliers_1.txt",
  Second.Outlier="No")
dim(S16_MisAlign0.6_1)

co1_MisAlign0.8_1=Detect.Outlier.Seq(

inputal="Alignments/AllSeqNaive_Ge/FirstAlign/Mafftfftns1_Alig_AllSeqN
aive_co1.fas",
  Strat.DistMat="Comb", Dist.Th = 0.8,
  output = "co1_outliers_1.txt",
  Second.Outlier="No")
dim(co1_MisAlign0.8_1)


########################################
# Removal of the outlier sequences and rebuilding of the species-by-
gene matrix for the selected gene regions

### For NCBI
NCBI.SeqTrash=c("AB213318",    "AF201577",    "M64907",    "AF137168",
"AH008176",
                "AF137189",    "AF137190",    "AF137169",    "AF137166",
"AF369080")

### For BOLD
```

```
BOLD.SeqTrash = c("4448254")

Sp.DNAMat1=SpeciesGeneMat.Bl(input=Sp.Genes.DF_naive,
                             output="SeqPool/Sp.Genes.cl_",
                             NCBI.Trash=NCBI.SeqTrash,
                             BOLD.Trash = BOLD.SeqTrash)


CleanDF_cl1=read.delim("SeqPool/Sp.Genes.cl__CleanDataset.txt",
sep="\t", h=T)
dim(CleanDF_cl1)

dim(Sp.Genes.DF_naive)[1]-dim(CleanDF_cl1)[1]


########################################
# Selection of the best sequences per species and gene region

dir.create("Alignments/OneBest")

Sp.Genes.DF.export_cl_2=SelBestSeq(input=CleanDF_cl1,

output="Alignments/OneBest/AligIn_OutGr_1Best",
                                   Alignment=T,          MaxSeq=1,
gene.list=SelectionGenes,
                                   SeqChoice="Median")


########################################
# Multiple alignments for the selected gene regions

dir.create("Alignments/OneBest/MultiAlign")

# Load the objects into the R environment.
output="Alignments/OneBest/MultiAlign"
input="Alignments/OneBest"
nthread=8
methods = c("mafftfftns2", "mafftfftnsi", "muscle")
Mafft.path = "D:/mafft-win/mafft"
Muscle.path = "D:/muscle/muscle3.8.31_i86win32.exe"

# Run the function.
Multi.Align(input="Alignments/OneBest",
output="Alignments/OneBest/MultiAlign",
            nthread=8,   methods   =   c("mafftfftns2",   "mafftfftnsi",
"muscle"),
            Mafft.path = "D:/mafft-win/mafft",
            Muscle.path = "D:/muscle/muscle3.8.31_i86win32.exe")


########################################
```

```
# Estimate the level of similarity among alignments using the multiple
overlap score (MOS)

# This step was performed on http://msa.cgb.ki.se/cgi-bin/msa.cgi
because on Windows machine.
# This is the list of alignments with highest MOS score for each gene
region:
# 12srrna: Mafftfftns2
# 16srrna: Mafftfftns2
# co1: Muscle
# cytb: Mafftfftnsi
# glyt: Mafftfftns2
# plagl2: Mafftfftns2
# rag1: Mafftfftnsi
# rag2: Mafftfftns2
# rhod: Mafftfftns2
# sh3px3: Mafftfftns2
# sreb2: Mafftfftns2
# zic1: Mafftfftns2

########################################
# Trim poorly aligned positions and/or gappy positions


# We prepare a vector with the Type of DNA for each alignment.
# The first alignment is the 16srrna, so it must
# be coded "d", all the others must be coded "c" (all coding DNA).
Type.ali = c("d", "d", "c", "c", "c", "c", "c", "c", "c", "c", "c",
"c")

i=1
for(i in 1:length(list.ali)){
  outGblocks                                                    =
Filtering.align.Gblocks(input="Alignments/OneBest/ToTrim",
                                        target.file = list.ali[i],
                                        LessStringent="TRUE",
Type=Type.ali[i],
                                        output
="Alignments/OneBest/Trimmed_Gblocks",
                                        remove.empty.align = TRUE,
                                        Gblocks.path                =
"D:/Gblocks_0.91b")
}
outGblocks


# After trimming with Gblocks, alignments were visually examined and
edited when needed


########################################
# Concatenate the trimmed alignments into a single supermatrix
```

```
Align.Concat(input="Alignments/OneBest/ForConcat",
             Sp.List=NULL,
             outputConcat = "Alignments/OneBest/ForConcat/Concat")


########################################
# Select the best partitioning scheme and substitution model using
PARTITIONFINDER2

PartiFinder2(input = "Alignments/OneBest/ForConcat/Concat.fas",
             Partition                                                =
"Alignments/OneBest/ForConcat/Partitions_Concat.txt",
             codon        =        c(3:12),        nexus.file        =
"Alignments/OneBest/ForConcat/Concat.nex",
             Path.PartiF2            =            "D:/partitionfinder-
2.1.1/PartitionFinder.py",
             branchlengths = "linked", models = "all", model_selection
= "BIC", search = "greedy",
             Raxml = "TRUE", nthread = 5)
```

R script for the Lloyd *et al*. (2016) algorithm of the Hedman-based *a posteriori* time scaling

```
library(paleotree)
library(FossilSim)
library(RRphylo)
library(ape)

# Load functions and libraries:
source("http://www.graemetlloyd.com/pubdata/functions_7.r")

# Load Bayesian consensus tree:
tree <-
read.nexus("OsteoglossomorphTotalEvidenceUndatedBayesian.con.tre")

# Drop outgroups
tree <- drop.tip (tree, tip = c("Amia_calva", "Elops_saurus",
"Dorosoma_cepedianum", "Ellimmichthyiformes"))

############################
# Graft branches corresponding to fragmentary fossils with known
taxonomic attribution (don't worry about warnings)

tree <- bind.tip(tree, "GymnarchusFossil", where =
which(tree$tip.label=="Gymnarchus_niloticus"), position = 0.01)
tree <- bind.tip(tree, "HeterotisFossil", where =
which(tree$tip.label=="Heterotis_niloticus"), position = 0.01)
tree <- bind.tip(tree, "HyperopisusFossil", where =
which(tree$tip.label=="Hyperopisus_bebe"), position = 0.01)
```

```
tree <- bind.tip(tree, "ArapaimaFossil", where =
which(tree$tip.label=="Arapaima_gigas"), position = 0.01)


# Load ages
ages <- read.csv("Hedman_TipDates.txt")
min.ages <- ages$min
max.ages <- ages$max
names(max.ages) <- names(min.ages) <- rownames(ages)
max.ages <- max.ages[tree$tip.label]      # Collapse ages to just taxa
in tree
min.ages<- min.ages[tree$tip.label]       # Collapse ages to just taxa
in tree

point.estimate <- runif(length(min.ages), min = min.ages, max =
max.ages)
point.estimate <- as.numeric(round(point.estimate, 2))
names(point.estimate) <- names(max.ages)

# Outgroup ages
outgroup.ages <- c(237, 233.6, 197.2, 197.2, 181.3, 154.8, 154.8,
149.2, 149.2, 149.2, 143.1)

datedTree <- Hedman.tree.dates(tree, point.estimate, outgroup.ages, t0
= 400, resolution = 1000, conservative = TRUE)

TimeTree <- drop.tip(datedTree$tree, c("GymnarchusFossil",
"HeterotisFossil", "HyperopisusFossil", "ArapaimaFossil"))
write.tree(TimeTree, "HedmanTree.Consensus.tre")

#OnlyExtantTree <- prune.fossil.tips(TimeTree)
#write.tree(OnlyExtantTree, "HedmanOnlyExtant.tre")

########################################
# Write down multiple Hedman-calibrated timetrees for same topology

averageTree <- datedTree$tree
HedmanTrees.100 <- vector(mode = "list", length = 1000)
for (i in 1:length(HedmanTrees.100)){
  HedmanTrees.100[[i]] <- scaleTree(averageTree, node.ages =
datedTree$age.distributions[,i], min.branch=0.00001)
  HedmanTrees.100[[i]] <- drop.tip(HedmanTrees.100[[i]],
c("GymnarchusFossil", "HeterotisFossil", "HyperopisusFossil",
"ArapaimaFossil"))
}

class(HedmanTrees.100) <- "multiPhylo"

# Select the tree with youngest node ages
write.tree(HedmanTrees.100[1], "MinAge.HedmanTree.Consensus.tre")

# Select the tree with oldest node ages
```

```
write.tree(HedmanTrees.100[1000], "MaxAge.HedmanTree.Consensus.tre")


class(HedmanTrees.100) <- "multiPhylo"

write.tree(HedmanTrees.100[c(1:1000)], "HedmanTrees100.tree")
```

BioGeoBEARS script for the biogeographic analysis under the 'MarineAsTrait' strategy

```
##########Code                                                                    for
Biogeography#############################################################################
#                     Original   Code   Nick   Matzke's,   modified   for   Bauer's   dataset
#
#      Most of the comments are Nick's some are Adriane Lam's some are Jen Bauer's, some are mine
(AC)#
#                  See   http://phylo.wikidot.com/biogeobears   for   updated   files   by   Nick
#
##########Code                                                                    for
Biogeography#############################################################################

install.packages("GenSA")
library(GenSA)    # GenSA is better than optimx (although somewhat slower)
install.packages("FD")
library(FD)       # for FD::maxent() (make sure this is up-to-date)
install.packages("snow")
library(snow)      # (if you want to use multicore functionality; some systems/R versions prefer
library(parallel), try either)
library(parallel)

install.packages("rexpokit")
install.packages("cladoRcpp")
install.packages("devtools")

library(rexpokit)
library(cladoRcpp)
devtools::install_github(repo="nmatzke/BioGeoBEARS",  dependencies=T)  #  for  when  BioGeoBEARS
fails, install dependencies separately

library(BioGeoBEARS)

calc_loglike_sp = compiler::cmpfun(calc_loglike_sp_prebyte)     # crucial to fix bug in uppass
calculations
calc_independent_likelihoods_on_each_branch                                                =
compiler::cmpfun(calc_independent_likelihoods_on_each_branch_prebyte)


######################################################
# SETUP: YOUR WORKING DIRECTORY
######################################################
# You will need to set your working directory to match your local system

# Note these very handy functions!
# Command "setwd(x)" sets your working directory
# Command "getwd()" gets your working directory and tells you what it is.
# Command "list.files()" lists the files in your working directory
# To get help on any command, use "?".  E.g., "?list.files"

# Set your working directory for output files
# default here is your home directory ("~")
# Change this as you like
wd = "D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/MarineAsTrait/"
setwd(wd)

# Double-check your working directory with getwd()
getwd()
```

```
######################################################
# SETUP: Extension data directory
######################################################
# When R packages contain extra files, they are stored in the "extdata" directory
# inside the installed package.
#
# BioGeoBEARS contains various example files and scripts in its extdata directory.
#
# Each computer operating system might install BioGeoBEARS in a different place,
# depending on your OS and settings.
#
# However, you can find the extdata directory like this:
extdata_dir = np(system.file("extdata", package="BioGeoBEARS"))
extdata_dir
list.files(extdata_dir)

# "system.file" looks in the directory of a specified package (in this case BioGeoBEARS)
# The function "np" is just a shortcut for normalizePath(), which converts the
# path to the format appropriate for your system (e.g., Mac/Linux use "/", but
# Windows uses "\\", if memory serves).

# Even when using your own data files, you should KEEP these commands in your
# script, since the plot_BioGeoBEARS_results function needs a script from the
# extdata directory to calculate the positions of "corners" on the plot. This cannot
# be made into a straight up BioGeoBEARS function because it uses C routines
# from the package APE which do not pass R CMD check for some reason.

######################################################
# SETUP: YOUR TREE FILE AND GEOGRAPHY FILE
######################################################
# Example files are given below. To run your own data,
# make the below lines point to your own files, e.g.
# trfn = "/mydata/frogs/frogBGB/tree.newick"
# geogfn = "/mydata/frogs/frogBGB/geog.data"

######################################################
# Phylogeny file
# Notes:
# 1. Must be binary/bifurcating: no polytomies
#  2.  No  negative  branchlengths  (e.g.  BEAST  MCC  consensus  trees  sometimes  have  negative
branchlengths)
# 3. Be careful of very short branches, as BioGeoBEARS will interpret ultrashort branches as
direct ancestors
# 4. You can use non-ultrametric trees, but BioGeoBEARS will interpret any tips significantly
below the
#     top of the tree as fossils!  This is only a good idea if you actually do have fossils in
your tree,
#    as in e.g. Wood, Matzke et al. (2013), Systematic Biology.
# 5. The default settings of BioGeoBEARS make sense for trees where the branchlengths are in
units of
#    millions of years, and the tree is 1-1000 units tall. If you have a tree with a total height
of
#    e.g. 0.00001, you will need to adjust e.g. the max values of d and e, or (simpler) multiply
all
#    your branchlengths to get them into reasonable units.
# 6. DON'T USE SPACES IN SPECIES NAMES, USE E.G. "_"
######################################################

#Read in and look at your raw tree file
trfn = "HedmanTree.Consensus.tre"
moref(trfn)

# Look at your phylogeny:
tr = read.tree(trfn)
tr
plot(tr)


######################################################
# Geography file
# Notes:
```

```
# 1. This is a PHLYIP-formatted file. This means that in the
#    first line,
#    - the 1st number equals the number of rows (species)
#    - the 2nd number equals the number of columns (number of areas)
# 2. This is the same format used for C++ LAGRANGE geography files.
# 3. All names in the geography file must match names in the phylogeny file.
# 4. DON'T USE SPACES IN SPECIES NAMES, USE E.G. "_"
# 5. Operational taxonomic units (OTUs) should ideally be phylogenetic lineages,
#    i.e. genetically isolated populations.  These may or may not be identical
#    with species.  You would NOT want to just use specimens, as each specimen
#    automatically can only live in 1 area, which will typically favor DEC+J
#    models.  This is fine if the species/lineages really do live in single areas,
#    but you wouldn't want to assume this without thinking about it at least.
#    In summary, you should collapse multiple specimens into species/lineages if
#    data indicates they are the same genetic population.
######################################################

#Read in the geography file
geogfn = "OsteoGeographyWithoutRogues.txt"
moref(geogfn)

#Look at geographic range data
tipranges = getranges_from_LagrangePHYLIP(lgdata_fn=geogfn)
tipranges

#Maximum range size observed
max(rowSums(dfnums_to_numeric(tipranges@df)))

#IMPORTANT: set the number of areas any species can occupy; CANNOT be larger than the total
#number of areas, but can be smaller

max_range_size = 3

areas = getareas_from_tipranges_object(tipranges)
#areas = c("A", "B", "C", "D")

# This is the list of states/ranges, where each state/range
# is a list of areas, counting from 0
states_list_0based    =    rcpp_areas_list_to_states_list(areas=areas,    maxareas=max_range_size,
include_null_range=TRUE)

# How many states/ranges, by default: 93
length(states_list_0based)

# Make the list of ranges
ranges_list = NULL
for (i in 1:length(states_list_0based))
{
  if ( (length(states_list_0based[[i]]) == 1) && (is.na(states_list_0based[[i]])) )
  {
    tmprange = "_"
  } else {
    tmprange = paste(areas[states_list_0based[[i]]+1], collapse="")
  }
  ranges_list = c(ranges_list, tmprange)
}

# Look at the ranges list
ranges_list

######################################################
# KEY HINT: The number of states (= number of different possible geographic ranges)
# depends on (a) the number of areas and (b) max_range_size.
# If you have more than about 500-600 states, the calculations will get REALLY slow,
# since the program has to exponentiate a matrix of e.g. 600x600.  Often the computer
# will just sit there and crunch, and never get through the calculation of the first
# likelihood.
#
# (this is also what is usually happening when LAGRANGE hangs: you have too many states!)
#
# To check the number of states for a given number of ranges, try:
```

```
numstates_from_numareas(numareas=7, maxareas=3, include_null_range=TRUE)
numstates_from_numareas(numareas=7, maxareas=3, include_null_range=FALSE)
numstates_from_numareas(numareas=7, maxareas=4, include_null_range=TRUE)
numstates_from_numareas(numareas=7, maxareas=4, include_null_range=FALSE)

#If you limit the range size above using max_range_size, you should be OK

######################################################################
######################################################
# Traits-only model -- 2 rates  - WITHOUT GEOGRAPHY
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$print_optim = TRUE
BioGeoBEARS_run_object$calc_ancprobs=TRUE        # get ancestral states from optim run
BioGeoBEARS_run_object$max_range_size = 1
BioGeoBEARS_run_object$num_cores_to_use=1
BioGeoBEARS_run_object$use_optimx="GenSA"
BioGeoBEARS_run_object$speedup=TRUE
BioGeoBEARS_run_object$geogfn = "OsteoGeography_1area.txt"   ## You NEED to create a different
geography file with 1 single area and all taxa scored as present in that area (AC)
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE
BioGeoBEARS_run_object$on_NaN_error = -1000000
BioGeoBEARS_run_object$force_sparse = FALSE  # works with kexpmv, but compare to dense,

# Set up DEC model, but set all rates to 0 (data are 1 invariant area)
# (nothing to do; defaults)

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["a","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["a","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["a","est"] = 0.0

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = 0.0

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = 0.0

# Set up BAYAREALIKE model
# No subset sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

# No vicariance
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","est"] = 0.0

# No jump dispersal/founder-event speciation
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = 0.01
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = 0.01

# Adjust linkage between parameters
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "1-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/1"
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "1-j"

# Only sympatric/range-copying (y) events allowed, and with
# exact copying (both descendants always the same size as the ancestor)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","init"] = 0.9999
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","est"] = 0.9999


# Load trait data
trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)


# Look at the params table
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

######################################################
# Manual modifications of trait-based model
######################################################
# Edit t12 and t21 rates

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = 0.001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = 0.001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = 0.001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = 0.001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = 1

# No multipliers on geog (set m1 and m2 to 1)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = 1.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = 1.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1",    "desc"]    =    "trait-based
dispersal rate multipliers m1"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2",    "desc"]    =    "trait-based
dispersal rate multipliers m2"

# Run this to check inputs. Read the error messages if you get them!
BioGeoBEARS_run_object = fix_BioGeoBEARS_params_minmax(BioGeoBEARS_run_object)
check_BioGeoBEARS_run(BioGeoBEARS_run_object)


# For a slow analysis, run once, then set runslow=FALSE to just
# load the saved result.
runslow = T
resfn = "OsteoMarineAsTrait_TraitOnly.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2  =  as.data.frame(matrix(data=c("TraitOnly_2rates",  res$total_loglikelihood,  txtout1),
nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
```

```
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)
  resTrait_2rates = res
} else {
  # Loads to "res"
  load(resfn)
  resTrait_2rates = res
}


######################################################

######################################################
######################################################
# DEC AND DEC+J ANALYSIS
######################################################
######################################################
# NOTE: The BioGeoBEARS "DEC" model is identical with
# the Lagrange DEC model, and should return identical
# ML estimates of parameters, and the same
# log-likelihoods, for the same datasets.
#
# Ancestral state probabilities at nodes will be slightly
# different, since BioGeoBEARS is reporting the
# ancestral state probabilities under the global ML
# model, and Lagrange is reporting ancestral state
# probabilities after re-optimizing the likelihood
# after fixing the state at each node. These will
# be similar, but not identical. See Matzke (2014),
# Systematic Biology, for discussion.
#
# Also see Matzke (2014) for presentation of the
# DEC+J model.
######################################################
######################################################

######################################################
# Run DEC
######################################################

# Intitialize a default model (DEC model)
BioGeoBEARS_run_object = define_BioGeoBEARS_run()

# Give BioGeoBEARS the location of the phylogeny Newick file
BioGeoBEARS_run_object$trfn = trfn

# Give BioGeoBEARS the location of the geography text file
BioGeoBEARS_run_object$geogfn = geogfn

# Input the maximum range size
BioGeoBEARS_run_object$max_range_size = max_range_size

BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"    # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4

BioGeoBEARS_run_object$force_sparse = FALSE     # force_sparse=TRUE causes pathology & isn't much
faster at this scale
```

```
# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DEC model
# (nothing to do; defaults)

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

# Run this to check inputs. Read the error messages if you get them!
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#Make sure you change the file name below to match your own data!!!
runslow = T
resfn = "OsteoMarineAsTrait_DEC.Rdata"

if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2  =  as.data.frame(matrix(data=c("DEC",  res$total_loglikelihood,  txtout1),  nrow=1),
stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)
  resDEC = res
} else {
  # Loads to "res"
  load(resfn)
  resDEC = res
}



######################################################
# Run DEC+J
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"    # if FALSE, use optim() instead of optimx()
```

```
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$force_sparse = FALSE     # force_sparse=TRUE causes pathology & isn't much
faster at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE     # get ancestral states from optim run

# Set up DEC+J model
# Get the ML parameter values from the 2-parameter nested model
# (this will ensure that the 3-parameter model always does at least as good)
dstart = resDEC$outputs@params_table["d","est"]
estart = resDEC$outputs@params_table["e","est"]
jstart = 0.0001

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Add j as a free parameter
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#Make sure to change the file name to match your data!
resfn = "OsteoMarineAsTrait_DECJ.Rdata"
runslow = T

if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2  =  as.data.frame(matrix(data=c("DEC+J",  res$total_loglikelihood,  txtout1),  nrow=1),
stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resDECj = res
} else {
  # Loads to "res"
  load(resfn)
  resDECj = res
}



######################################################
# PDF plots
#MAKE SURE TO CHANGE FILE NAMES TO MATCH YOUR OWN!!!
######################################################
pdffn = "OsteoMarineAsTrait_DECJ.pdf"
pdf(pdffn, width=8.5, height=11)

######################################################
```

```
# Plot ancestral states - DEC
#I modified this code to NOT plot pie charts at the corners of the
#phylogeny, ONLY at the nodes. If you want to change this, I still
#have old emails from Nick on how to turn these on and off -ARL
######################################################
analysis_titletxt ="BioGeoBEARS DEC on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resDEC
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

######################################################
# Plot ancestral states - DECJ
#I modified this code to NOT plot pie charts at the corners of the
#phylogeny, ONLY at the nodes. If you want to change this, I still
#have old emails from Nick on how to turn these on and off -ARL
######################################################
analysis_titletxt ="BioGeoBEARS DEC+J on Osteoglossomorpha - MarineAsTraitstrategy"

# Setup
results_object = resDECj
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

dev.off()  # Turn off PDF
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr) # Plot it


######################################################
# Run DEC + t12 + t21 + m1 (trait-based dispersal model)
######################################################

# Intitialize a default model (DEC model)
BioGeoBEARS_run_object = define_BioGeoBEARS_run()

# Give BioGeoBEARS the location of the phylogeny Newick file
BioGeoBEARS_run_object$trfn = trfn

# Give BioGeoBEARS the location of the geography text file
BioGeoBEARS_run_object$geogfn = geogfn

# Input the maximum range size
BioGeoBEARS_run_object$max_range_size = max_range_size

BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
```

```
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"    # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4

BioGeoBEARS_run_object$force_sparse = FALSE     # force_sparse=TRUE causes pathology & isn't much
faster at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DEC + t12 + t21 + m1 model
# Get the ML parameter values from the DEC model
# (speeds up ML calculations) (AC)

dstart = resDEC$outputs@params_table["d","est"]
estart = resDEC$outputs@params_table["e","est"]

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

########################################################################################

trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)

#######################################################
# Manual modifications of trait-based model
#######################################################
# Edit t12 and t21 rates
t12_start = 0.01
t21_start = 0.01
m1_start = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = round(max(t12_start,
t21_start)*10, 3)

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = t21_start
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = round(max(t12_start,
t21_start)*10, 3)


# Set 0/1 multipliers on dispersal rate
# Multiplier fixed at 1 for marine (m2)
# For freshwater (m1), max multiplier is 5, min multiplier is 0.0001, and estimated
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "min"] = 0.01
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "min"] = 0.0001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "max"] = 5




# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

################################################################
# INPUT the NEW states list into the BioGeoBEARS_run_object
#BioGeoBEARS_run_object$states_list = states_list_0based_NEW
################################################################

# Run this to check inputs. Read the error messages if you get them!
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#Make sure you change the file name below to match your own data!!!
runslow = T
resfn = "OsteoMarineAsTrait_DECm1.Rdata"

if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2  =  as.data.frame(matrix(data=c("DEC+t12+t21+m1",  res$total_loglikelihood,  txtout1),
nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)
  resDECm1 = res
} else {
  # Loads to "res"
  load(resfn)
  resDECm1 = res
}


######################################################
# Run DEC + j + t12 + t21 + m1 (trait-based dispersal model)
######################################################

# Intitialize a default model (DEC model)
BioGeoBEARS_run_object = define_BioGeoBEARS_run()

# Give BioGeoBEARS the location of the phylogeny Newick file
BioGeoBEARS_run_object$trfn = trfn
```

```
# Give BioGeoBEARS the location of the geography text file
BioGeoBEARS_run_object$geogfn = geogfn

# Input the maximum range size
BioGeoBEARS_run_object$max_range_size = max_range_size

BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"    # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4

BioGeoBEARS_run_object$force_sparse = FALSE     # force_sparse=TRUE causes pathology & isn't much
faster at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DEC+J+t12+t21+m1 model
# Get the ML parameter values from the DEC+t12+t21+m1 nested model
# (this will ensure that the +J model always does at least as good)
dstart = resDECm1$outputs@params_table["d","est"]
estart = resDECm1$outputs@params_table["e","est"]
jstart = 0.0001

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Add j as a free parameter
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

########################################################################################

trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)

#####################################################
# Manual modifications of trait-based model
#####################################################
# Edit t12 and t21 rates
t12_start = resDECm1$outputs@params_table["t12","est"]
```

```
t21_start = resDECm1$outputs@params_table["t21","est"]
m1_start = resDECm1$outputs@params_table["m1","est"]

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = round(max(t12_start,
t21_start)*10, 3)

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = round(max(t12_start,
t21_start)*10, 3)


# Set 0/1 multipliers on dispersal rate
# Multiplier fixed at 1 for marine (m2)
# For freshwater (m1), max multiplier is 5, min multiplier is 0.0001, and estimated
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "min"] = 0.01
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "min"] = 0.0001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "max"] = 5




# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

# Run this to check inputs. Read the error messages if you get them!
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#Make sure you change the file name below to match your own data!!!
runslow = T
resfn = "OsteoMarineAsTrait_DECJm1.Rdata"

if (runslow)
{
   # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2  =  as.data.frame(matrix(data=c("DEC+J+t12+t21+m1",  res$total_loglikelihood,  txtout1),
nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)
  resDECJm1 = res
} else {
  # Loads to "res"
  load(resfn)
  resDECJm1 = res
}

########################################
#############################
# Plotting ancestral areas after trait-based dispersal model (AC)
```

```
# Collapse the geog+traits probabilities to just traits
resDEC_wTraits = resDECm1
resDECj_wTraits = resDECJm1

numcols_yTrait = ncol(resDEC_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node)
numcols_nTrait = ncol(resDEC_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node) / 2

resDEC_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node                                =
resDEC_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,1:numcols_nTrait]             +
resDEC_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,(numcols_nTrait+1):numcols_yTra
it]

resDEC_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node                          =
resDEC_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,1:numcols_nTrait]       +
resDEC_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,(numcols_nTrait+1):numcol
s_yTrait]


resDECj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node                               =
resDECj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,1:numcols_nTrait]            +
resDECj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,(numcols_nTrait+1):numcols_yTr
ait]

resDECj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node                         =
resDECj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,1:numcols_nTrait]      +
resDECj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,(numcols_nTrait+1):numco
ls_yTrait]


######################################################
# PDF plots
######################################################
pdffn = "OsteoMarineAsTrait_DECJ_m1.pdf"
pdf(pdffn, width=8.5, height=11)

######################################################
# Plot ancestral states - DEC+t12+t21+m1
######################################################
analysis_titletxt ="BioGeoBEARS DEC+t12+t21+m1 on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resDEC_wTraits
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)


######################################################
# Plot ancestral states - DECJ+t12+t21+m1
######################################################
analysis_titletxt ="BioGeoBEARS DEC+J+t12+t21+m1 on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resDECj_wTraits
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
```

```
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

dev.off()  # Turn off PDF
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr) # Plot it


# Plot binary trait

results_object = resTrait_2rates

allowed_geog_states                                                                                <-
length(results_object$inputs$all_geog_states_list_usually_inferred_from_areas_maxareas)

results_marginal <- results_object$ML_marginal_prob_each_state_at_branch_top_AT_node
trait_marginal <- data.frame(matrix(ncol = 2, nrow = nrow(results_marginal)))
trait_marginal[,1] <- rowSums(results_marginal[, 1:allowed_geog_states])
trait_marginal[,2] <- rowSums(results_marginal[, (allowed_geog_states+1):ncol(results_marginal)])

tipStates <- trait_marginal[1:(tree$Nnode+1), ]
nodeStates <- trait_marginal[(tree$Nnode+2):nrow(trait_marginal), ]

pdffn = "OsteoMarineAsTrait_2Rates_Trait.pdf"
pdf(pdffn, width=8.5, height=11)

plotTree(tree)
tiplabels(pie=unlist(matrix(tipStates)), cex = 0.3)
nodelabels(pie = unlist(matrix(nodeStates)), cex=0.5)
dev.off()  # Turn off PDF
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr) # Plot it


######################################################
######################################################
# DIVALIKE AND DIVALIKE+J ANALYSIS
######################################################
######################################################
# NOTE: The BioGeoBEARS "DIVALIKE" model is not identical with
# Ronquist (1997)'s parsimony DIVA. It is a likelihood
# interpretation of DIVA, constructed by modelling DIVA's
# processes the way DEC does, but only allowing the
# processes DIVA allows (widespread vicariance: yes; subset
# sympatry: no; see Ronquist & Sanmartin 2011, Figure 4).
#
# DIVALIKE is a likelihood interpretation of parsimony
# DIVA, and it is "like DIVA" -- similar to, but not
# identical to, parsimony DIVA.
#
# I thus now call the model "DIVALIKE", and you should also. ;-)
######################################################
######################################################

######################################################
# Run DIVALIKE
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.
```

```
# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE             # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DIVALIKE model
# Remove subset-sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "2-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "ysv*1/2"

# Allow classic, widespread vicariance; all events equiprobable
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","init"] = 0.5
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","est"] = 0.5

# No jump dispersal/founder-event speciation
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = 0.01
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = 0.01

check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR OWN DATA
runslow = T
resfn = "OsteoMarineAsTrait_DIVALIKE.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2 = as.data.frame(matrix(data=c("DIVALIKE", res$total_loglikelihood, txtout1), nrow=1),
stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resDIVALIKE = res
} else {
  # Loads to "res"
  load(resfn)
  resDIVALIKE = res
}

#If you get a warning message like: In (function(npt=min(n + 2L, 2L * n), rhobeg=NA, rhoend=NA;
```

```
#      unused control arguments ignored
#Just ignore it. IDK what it means and Nick seemed fine with it -ARL

######################################################
# Run DIVALIKE+J
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DIVALIKE+J model
# Get the ML parameter values from the 2-parameter nested model
# (this will ensure that the 3-parameter model always does at least as good)
dstart = resDIVALIKE$outputs@params_table["d","est"]
estart = resDIVALIKE$outputs@params_table["e","est"]
jstart = 0.0001

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Remove subset-sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "2-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "ysv*1/2"

# Allow classic, widespread vicariance; all events equiprobable
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","init"] = 0.5
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","est"] = 0.5

# Add jump dispersal/founder-event speciation
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

# Under DIVALIKE+J, the max of "j" should be 2, not 3 (as is default in DEC+J)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","max"] = 1.99999
```

```
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR DATA
resfn = "OsteoMarineAsTrait_DIVALIKEJ.Rdata"
runslow = T
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2 = as.data.frame(matrix(data=c("DIVALIKE+J", res$total_loglikelihood, txtout1), nrow=1),
stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resDIVALIKEj = res
} else {
  # Loads to "res"
  load(resfn)
  resDIVALIKEj = res
}

######################################################
# Plot ancestral states - DIVALIKE
#Same here- I turned off pie charts and area reconstructions at corners -ARL
######################################################
pdffn = "OsteoMarineAsTrait_DIVALIKEJ.pdf"
pdf(pdffn, width=8.5, height=11)

analysis_titletxt ="BioGeoBEARS DIVALIKE on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resDIVALIKE
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

######################################################
# Plot ancestral states - DIVALIKE+J
######################################################
analysis_titletxt ="BioGeoBEARS DIVALIKE+J on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resDIVALIKEj
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
```

```
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

dev.off()
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr)


######################################################
# Run DIVALIKE +t12 + t21 + m1
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE     # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DIVALIKE model
# Remove subset-sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "2-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "ysv*1/2"

# Allow classic, widespread vicariance; all events equiprobable
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","init"] = 0.5
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","est"] = 0.5

# No jump dispersal/founder-event speciation
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = 0.01
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = 0.01

# Set up DIVALIKE + t12 + t21 + m1 model
# Get the ML parameter values from the DIVALIKE model
# (speeds up ML calculations) (AC)

dstart = resDIVALIKE$outputs@params_table["d","est"]
estart = resDIVALIKE$outputs@params_table["e","est"]

# Input starting values for d, e
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

########################################################################################

trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)

#####################################################
# Manual modifications of trait-based model
#####################################################
# Edit t12 and t21 rates
t12_start = 0.01
t21_start = 0.01
m1_start = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = round(max(t12_start,
t21_start)*10, 3)

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = round(max(t12_start,
t21_start)*10, 3)


# Set 0/1 multipliers on dispersal rate
# Multiplier fixed at 1 for marine (m2)
# For freshwater (m1), max multiplier is 5, min multiplier is 0.0001, and estimated
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "min"] = 0.01
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "min"] = 0.0001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "max"] = 5


# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR OWN DATA
runslow = T
resfn = "OsteoMarineAsTrait_DIVALIKEm1.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
```

```
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2 = as.data.frame(matrix(data=c("DIVALIKE + t12 + t21 + m1", res$total_loglikelihood,
txtout1), nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                    =                  c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resDIVALIKEm1 = res
} else {
  # Loads to "res"
  load(resfn)
  resDIVALIKEm1 = res
}


######################################################
# Run DIVALIKE+J+t12+t21+m1
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up DIVALIKE+J model
# Get the ML parameter values from the 2-parameter nested model
# (this will ensure that the 3-parameter model always does at least as good)
dstart = resDIVALIKEm1$outputs@params_table["d","est"]
estart = resDIVALIKEm1$outputs@params_table["e","est"]
jstart = 0.0001

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Remove subset-sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "2-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "ysv*1/2"

# Allow classic, widespread vicariance; all events equiprobable
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","init"] = 0.5
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","est"] = 0.5

# Add jump dispersal/founder-event speciation
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

# Under DIVALIKE+J, the max of "j" should be 2, not 3 (as is default in DEC+J)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","max"] = 1.99999

######################################################################################

trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)

#####################################################
# Manual modifications of trait-based model
#####################################################
# Edit t12 and t21 rates
t12_start = resDIVALIKEm1$outputs@params_table["t12","est"]
t21_start = resDIVALIKEm1$outputs@params_table["t21","est"]
m1_start = resDIVALIKEm1$outputs@params_table["m1","est"]

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = round(max(t12_start,
t21_start)*10, 3)

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = round(max(t12_start,
t21_start)*10, 3)


# Set 0/1 multipliers on dispersal rate
# Multiplier fixed at 1 for marine (m2)
# For freshwater (m1), max multiplier is 5, min multiplier is 0.0001, and estimated
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "min"] = 0.01
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "min"] = 0.0001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "max"] = 5


# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table
```

```
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR DATA
resfn = "OsteoMarineAsTrait_DIVALIKEJm1.Rdata"
runslow = T
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2    =    as.data.frame(matrix(data=c("DIVALIKE+J+t12+t21+m1",    res$total_loglikelihood,
txtout1), nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                    =                    c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resDIVALIKEjm1 = res
} else {
  # Loads to "res"
  load(resfn)
  resDIVALIKEjm1 = res
}


########################################
############################
# Plotting ancestral areas after trait-based dispersal model (AC)

# Collapse the geog+traits probabilities to just traits
resDIVALIKE_wTraits = resDIVALIKEm1
resDIVALIKEj_wTraits = resDIVALIKEjm1

numcols_yTrait = ncol(resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node)
numcols_nTrait = ncol(resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node) / 2

resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node                              =
resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,1:numcols_nTrait]        +
resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,(numcols_nTrait+1):numcols
_yTrait]

resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node                        =
resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,1:numcols_nTrait]   +
resDIVALIKE_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,(numcols_nTrait+1):n
umcols_yTrait]


resDIVALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node                             =
resDIVALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,1:numcols_nTrait]       +
resDIVALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,(numcols_nTrait+1):numcol
s_yTrait]

resDIVALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node                       =
resDIVALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,1:numcols_nTrait]  +
resDIVALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,(numcols_nTrait+1):
numcols_yTrait]



########################################################
# PDF plots
########################################################
pdffn = "OsteoMarineAsTrait_DIVALIKEJ_m1.pdf"
pdf(pdffn, width=8.5, height=11)

########################################################
```

```
# Plot ancestral states - DIVALIKE+t12+t21+m1
######################################################
analysis_titletxt ="BioGeoBEARS DIVALIKE+t12+t21+m1 on Osteoglossomorpha - MarineAsTrait
strategy"

# Setup
results_object = resDIVALIKE_wTraits
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2 = plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="text", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=FALSE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="pie", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=FALSE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)


######################################################
# Plot ancestral states - DIVALIKEJ+t12+t21+m1
######################################################
analysis_titletxt ="BioGeoBEARS DIVALIKE+J+t12+t21+m1 on Osteoglossomorpha - MarineAsTrait
strategy"

# Setup
results_object = resDIVALIKEj_wTraits
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1 = plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="text", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=FALSE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="pie", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=FALSE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

dev.off()  # Turn off PDF
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr) # Plot it

# Plot binary trait

results_object = resDIVALIKEjm1

allowed_geog_states                                                                    <-
length(results_object$inputs$all_geog_states_list_usually_inferred_from_areas_maxareas)

results_marginal <- results_object$ML_marginal_prob_each_state_at_branch_top_AT_node
trait_marginal <- data.frame(matrix(ncol = 2, nrow = nrow(results_marginal)))
trait_marginal[,1] <- rowSums(results_marginal[, 1:allowed_geog_states])
trait_marginal[,2] <- rowSums(results_marginal[, (allowed_geog_states+1):ncol(results_marginal)])

tipStates <- trait_marginal[1:(tree$Nnode+1), ]
nodeStates <- trait_marginal[(tree$Nnode+2):nrow(trait_marginal), ]

plotTree(tree)
tiplabels(pie=unlist(matrix(tipStates)), cex = 0.3)
nodelabels(pie = unlist(matrix(nodeStates)), cex=0.5)
```

```
######################################################
######################################################
# BAYAREALIKE AND BAYAREALIKE+J ANALYSIS
######################################################
######################################################
# NOTE: As with DIVA, the BioGeoBEARS BayArea-like model is
# not identical with the full Bayesian model implemented
# in the "BayArea" program of Landis et al. (2013).
#
# Instead, this is a simplified likelihood interpretation
# of the model.  Basically, in BayArea and BioGeoBEARS-BAYAREALIKE,
# "d" and "e" work like they do in the DEC model of Lagrange
# (and BioGeoBEARS), and then BayArea's cladogenesis assumption
# (which is that nothing in particular happens at cladogenesis) is
# replicated by BioGeoBEARS.
#
# This leaves out 3 important things that are in BayArea:
# 1. Distance dependence (you can add this with a distances
#    matrix + the "x" parameter in BioGeoBEARS, however)
# 2. A correction for disallowing "e" events that drive
#    a species extinct (a null geographic range)
# 3. The neat Bayesian sampling of histories, which allows
#    analyses on large numbers of areas.
#
# The main purpose of having a "BAYAREALIKE" model is
# to test the importance of the cladogenesis model on
# particular datasets. Does it help or hurt the data
# likelihood if there is no special cladogenesis process?
#
# BAYAREALIKE is a likelihood interpretation of BayArea,
# and it is "like BayArea" -- similar to, but not
# identical to, Bayesian BayArea.
# I thus now call the model "BAYAREALIKE", and you should also. ;-)
######################################################
######################################################

######################################################
# Run BAYAREALIKE
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up BAYAREALIKE model
# No subset sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

# No vicariance
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","est"] = 0.0

# No jump dispersal/founder-event speciation
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = 0.01
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = 0.01

# Adjust linkage between parameters
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "1-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/1"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "1-j"

# Only sympatric/range-copying (y) events allowed, and with
# exact copying (both descendants always the same size as the ancestor)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","init"] = 0.9999
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","est"] = 0.9999


# Check the inputs
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR OWN
runslow = T
resfn = "OsteoMarineAsTrait_BAYAREALIKE.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2   =   as.data.frame(matrix(data=c("BAYAREALIKE",   res$total_loglikelihood,   txtout1),
nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                      =                    c("model_num",                     "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resBAYAREALIKE = res
} else {
  # Loads to "res"
  load(resfn)
  resBAYAREALIKE = res
}

######################################################
# Run BAYAREALIKE+J
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE             # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
```

```
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE     # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up BAYAREALIKE+J model
# Get the ML parameter values from the 2-parameter nested model
# (this will ensure that the 3-parameter model always does at least as good)
dstart = resBAYAREALIKE$outputs@params_table["d","est"]
estart = resBAYAREALIKE$outputs@params_table["e","est"]
jstart = 0.0001

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# No subset sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

# No vicariance
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","est"] = 0.0

# *DO* allow jump dispersal/founder-event speciation (set the starting value close to 0)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

# Under BAYAREALIKE+J, the max of "j" should be 1, not 3 (as is default in DEC+J) or 2 (as in
DIVALIKE+J)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","max"] = 0.99999

# Adjust linkage between parameters
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "1-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/1"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "1-j"

# Only sympatric/range-copying (y) events allowed, and with
# exact copying (both descendants always the same size as the ancestor)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","init"] = 0.9999
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","est"] = 0.9999

# NOTE (NJM, 2014-04): BAYAREALIKE+J seems to crash on some computers, usually Windows
# machines. I can't replicate this on my Mac machines, but it is almost certainly
# just some precision under-run issue, when optim/optimx tries some parameter value
# just below zero.  The "min" and "max" options on each parameter are supposed to
# prevent this, but apparently optim/optimx sometimes go slightly beyond
# these limits.  Anyway, if you get a crash, try raising "min" and lowering "max"
# slightly for each parameter:
#BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","min"] = 0.0000001
#BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","max"] = 4.9999999

#BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","min"] = 0.0000001
#BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","max"] = 4.9999999
```

```
#BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","min"] = 0.00001
#BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","max"] = 0.99999


# Check the inputs
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR OWN
runslow = T
resfn = "OsteoMarineAsTrait_BAYAREALIKEJ.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2   =   as.data.frame(matrix(data=c("BAYAREALIKE+J",   res$total_loglikelihood,   txtout1),
nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                      =                     c("model_num",                      "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resBAYAREALIKEj = res
} else {
  # Loads to "res"
  load(resfn)
  resBAYAREALIKEj = res
}

pdffn = "OsteoMarineAsTrait_BAYAREALIKEJ.pdf"
pdf(pdffn, width=8.5, height=11)

######################################################
# Plot ancestral states - BAYAREALIKE
######################################################
analysis_titletxt ="BioGeoBEARS BAYAREALIKE on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resBAYAREALIKE
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

######################################################
# Plot ancestral states - BAYAREALIKE+J
######################################################
analysis_titletxt ="BioGeoBEARS BAYAREALIKE+J on Osteoglossomorpha - MarineAsTrait strategy"

# Setup
results_object = resBAYAREALIKEj
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)
```

```
# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

dev.off()
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr)


######################################################
# Run BAYAREALIKE +t12 + t21 + m1
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE             # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up BAYAREALIKE model
# No subset sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

# No vicariance
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","est"] = 0.0

# No jump dispersal/founder-event speciation
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = 0.01
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = 0.01

# Adjust linkage between parameters
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "1-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/1"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "1-j"

# Only sympatric/range-copying (y) events allowed, and with
# exact copying (both descendants always the same size as the ancestor)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","init"] = 0.9999
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","est"] = 0.9999
```

```
# Set up BAYAREALIKE + t12 + t21 + m1 model
# Get the ML parameter values from the BAYAREALIKE model
# (speeds up ML calculations) (AC)

dstart = resBAYAREALIKE$outputs@params_table["d","est"]
estart = resBAYAREALIKE$outputs@params_table["e","est"]

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

######################################################################################################

trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)

######################################################
# Manual modifications of trait-based model
######################################################
# Edit t12 and t21 rates
t12_start = 0.01
t21_start = 0.01
m1_start = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = round(max(t12_start,
t21_start)*10, 3)

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = round(max(t12_start,
t21_start)*10, 3)


# Set 0/1 multipliers on dispersal rate
# Multiplier fixed at 1 for marine (m2)
# For freshwater (m1), max multiplier is 5, min multiplier is 0.0001, and estimated
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "min"] = 0.01
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "min"] = 0.0001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "max"] = 5


# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table
```

```
check_BioGeoBEARS_run(BioGeoBEARS_run_object)

#CHANGE FILE NAMES TO MATCH YOUR OWN
runslow = T
resfn = "OsteoMarineAsTrait_BAYAREALIKEm1.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2    =    as.data.frame(matrix(data=c("BAYAREALIKE+t12+t21+m1",    res$total_loglikelihood,
txtout1), nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                        =                    c("model_num",                     "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resBAYAREALIKEm1 = res
} else {
  # Loads to "res"
  load(resfn)
  resBAYAREALIKEm1 = res
}


######################################################
# Run BAYAREALIKE+J+t12+t21+m1
######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size
BioGeoBEARS_run_object$min_branchlength = 0.000001    # Min to treat tip as a direct ancestor (no
speciation event)
BioGeoBEARS_run_object$include_null_range = TRUE     # set to FALSE for e.g. DEC* model, DEC*+J,
etc.

# Speed options and multicore processing if desired
BioGeoBEARS_run_object$speedup = TRUE              # shorcuts to speed ML search; use FALSE if
worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = "GenSA"     # if FALSE, use optim() instead of optimx()
BioGeoBEARS_run_object$num_cores_to_use = 4
BioGeoBEARS_run_object$on_NaN_error = -1e50     # returns very low lnL if parameters produce NaN
error (underflow check)
BioGeoBEARS_run_object$force_sparse=FALSE     # sparse=FALSE causes pathology & isn't much faster
at this scale

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE    # get ancestral states from optim run

# Set up BAYAREALIKE+J+t12+t21+m1 model
# Get the ML parameter values from the 5-parameter nested model
# (this will ensure that the 6-parameter model always does at least as good)
dstart = resBAYAREALIKEm1$outputs@params_table["d","est"]
estart = resBAYAREALIKEm1$outputs@params_table["e","est"]
jstart = 0.0001

# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# No subset sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0

# No vicariance
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","est"] = 0.0

# *DO* allow jump dispersal/founder-event speciation (set the starting value close to 0)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

# Under BAYAREALIKE+J, the max of "j" should be 1, not 3 (as is default in DEC+J) or 2 (as in
DIVALIKE+J)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","max"] = 0.99999

# Adjust linkage between parameters
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "1-j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] = "ysv*1/1"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] = "1-j"

# Only sympatric/range-copying (y) events allowed, and with
# exact copying (both descendants always the same size as the ancestor)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","type"] = "fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","init"] = 0.9999
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01y","est"] = 0.9999


######################################################################################

trait_fn = "traits.data_NoRogues.txt"
trait_values = getranges_from_LagrangePHYLIP(lgdata_fn=trait_fn)
trait_values

# Add the traits data and model
BioGeoBEARS_run_object       =       add_trait_to_BioGeoBEARS_run_object(BioGeoBEARS_run_object,
trait_fn=trait_fn)

#####################################################
# Manual modifications of trait-based model
#####################################################
# Edit t12 and t21 rates
t12_start = resBAYAREALIKEm1$outputs@params_table["t12","est"]
t21_start = resBAYAREALIKEm1$outputs@params_table["t21","est"]
m1_start = resBAYAREALIKEm1$outputs@params_table["m1","est"]

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "init"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "est"] = t12_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t12", "max"] = round(max(t12_start,
t21_start)*10, 3)

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "init"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "est"] = t21_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "min"] = 0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["t21", "max"] = round(max(t12_start,
t21_start)*10, 3)


# Set 0/1 multipliers on dispersal rate
# Multiplier fixed at 1 for marine (m2)
# For freshwater (m1), max multiplier is 5, min multiplier is 0.0001, and estimated
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "type"] = "fixed"
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "init"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "est"] = 1
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "min"] = 0.01
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m2", "max"] = 1

BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "init"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "est"] = m1_start
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "min"] = 0.0001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["m1", "max"] = 5


# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

check_BioGeoBEARS_run(BioGeoBEARS_run_object)
#CHANGE FILE NAMES TO MATCH YOUR OWN
runslow = T
resfn = "OsteoMarineAsTrait_BAYAREALIKEJm1.Rdata"
if (runslow)
{
  # Calculate the lnL for the parameters, and store in text file
  res = bears_optim_run(BioGeoBEARS_run_object, skip_optim=F, skip_optim_option="return_all")
  res$total_loglikelihood
  txtout1 = res$output@params_table$est[res$output@params_table$type=="free"]
  txtout2   =   as.data.frame(matrix(data=c("BAYAREALIKE+J+t12+t21+m1",   res$total_loglikelihood,
txtout1), nrow=1), stringsAsFactors=FALSE)
  names(txtout2)                    =                   c("model_num",                    "lnL",
row.names(res$output@params_table[res$output@params_table$type=="free",]))
  suppressWarnings(write.table(x=txtout2,
file="D:/Documenti/OsteoglossomorphaPhylogenyPaper/BiogeographicAnalysis/ML_models_MarineAsTrait.
txt", sep="\t", append=TRUE, row.names=FALSE, quote=FALSE, col.names=TRUE))

  save(res, file=resfn)

  resBAYAREALIKEjm1 = res
} else {
  # Loads to "res"
  load(resfn)
  resBAYAREALIKEjm1 = res
}


########################################
#############################
# Plotting ancestral areas after trait-based dispersal model (AC)

# Collapse the geog+traits probabilities to just traits
resBAYAREALIKE_wTraits = resBAYAREALIKEm1
resBAYAREALIKEj_wTraits = resBAYAREALIKEjm1

numcols_yTrait = ncol(resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node)
numcols_nTrait = ncol(resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node) /
2

resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node                        =
resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,1:numcols_nTrait]     +
resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,(numcols_nTrait+1):numc
ols_yTrait]

resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node                  =
resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,1:numcols_nTrait]
+
resBAYAREALIKE_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,(numcols_nTrait+1
):numcols_yTrait]


resBAYAREALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node                       =
resBAYAREALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,1:numcols_nTrait]    +
resBAYAREALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node[,(numcols_nTrait+1):num
cols_yTrait]
```

```
resBAYAREALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node                 =
resBAYAREALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,1:numcols_nTrait
]                                                                                               +
resBAYAREALIKEj_wTraits$ML_marginal_prob_each_state_at_branch_bottom_below_node[,(numcols_nTrait+
1):numcols_yTrait]



######################################################
# PDF plots
######################################################
pdffn = "OsteoMarineAsTrait_BAYAREALIKEJ_m1.pdf"
pdf(pdffn, width=8.5, height=11)

######################################################
# Plot ancestral states - BAYAREALIKE+t12+t21+m1
######################################################
analysis_titletxt ="BioGeoBEARS BAYAREALIKE+t12+t21+m1 on Osteoglossomorpha - MarineAsTrait
strategy"

# Setup
results_object = resBAYAREALIKE_wTraits
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)


######################################################
# Plot ancestral states - BAYAREALIKEJ+t12+t21+m1
######################################################
analysis_titletxt ="BioGeoBEARS BAYAREALIKE+J+t12+t21+m1 on Osteoglossomorpha - MarineAsTrait
strategy"

# Setup
results_object = resBAYAREALIKEj_wTraits
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1   =   plot_BioGeoBEARS_results(results_object,   analysis_titletxt,   addl_params=list("j"),
plotwhat="text",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object,         analysis_titletxt,         addl_params=list("j"),
plotwhat="pie",   label.offset=0.45,   tipcex=0.7,   statecex=0.7,   splitcex=0.6,   titlecex=0.8,
plotsplits=FALSE,        cornercoords_loc=scriptdir,        include_null_range=TRUE,        tr=tr,
tipranges=tipranges)

dev.off()  # Turn off PDF
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr) # Plot it

# Plot binary trait

results_object = resBAYAREALIKEm1

allowed_geog_states                                                                             <-
length(results_object$inputs$all_geog_states_list_usually_inferred_from_areas_maxareas)
```

```
results_marginal <- results_object$ML_marginal_prob_each_state_at_branch_top_AT_node
trait_marginal <- data.frame(matrix(ncol = 2, nrow = nrow(results_marginal)))
trait_marginal[,1] <- rowSums(results_marginal[, 1:allowed_geog_states])
trait_marginal[,2] <- rowSums(results_marginal[, (allowed_geog_states+1):ncol(results_marginal)])

tipStates <- trait_marginal[1:(tree$Nnode+1), ]
nodeStates <- trait_marginal[(tree$Nnode+2):nrow(trait_marginal), ]

plotTree(tree)
tiplabels(pie=unlist(matrix(tipStates)), cex = 0.3)
nodelabels(pie = unlist(matrix(nodeStates)), cex=0.5)


######################################################################
######################################################################
# Summarize results of all analyses (AC)

param_names = c("lnL", "d", "e", "j", "t12", "t21", "m1", "m2")

Trait_2rates_results = c(
  resTrait_2rates$total_loglikelihood,
  resTrait_2rates$output@params_table["d", "est"],
  resTrait_2rates$output@params_table["e", "est"],
  resTrait_2rates$output@params_table["j", "est"],
  resTrait_2rates$output@params_table["t12", "est"],
  resTrait_2rates$output@params_table["t21", "est"],
  resTrait_2rates$output@params_table["m1", "est"],
  resTrait_2rates$output@params_table["m2", "est"]
)
names(Trait_2rates_results) = paste("Trait_2rates_", param_names, sep="")

DEC_results = c(
  resDEC$total_loglikelihood,
  resDEC$output@params_table["d", "est"],
  resDEC$output@params_table["e", "est"],
  resDEC$output@params_table["j", "est"],
  resDEC$output@params_table["t12", "est"],
  resDEC$output@params_table["t21", "est"],
  resDEC$output@params_table["m1", "est"],
  resDEC$output@params_table["m2", "est"]
)
names(DEC_results) = paste("DEC_", param_names, sep="")


DECj_results = c(
  resDECj$total_loglikelihood,
  resDECj$output@params_table["d", "est"],
  resDECj$output@params_table["e", "est"],
  resDECj$output@params_table["j", "est"],
  resDECj$output@params_table["t12", "est"],
  resDECj$output@params_table["t21", "est"],
  resDECj$output@params_table["m1", "est"],
  resDECj$output@params_table["m2", "est"]
)
names(DECj_results) = paste("DECj_", param_names, sep="")

DEC_t12_t21_m2_results = c(
  resDECm1$total_loglikelihood,
  resDECm1$output@params_table["d", "est"],
  resDECm1$output@params_table["e", "est"],
  resDECm1$output@params_table["j", "est"],
  resDECm1$output@params_table["t12", "est"],
  resDECm1$output@params_table["t21", "est"],
  resDECm1$output@params_table["m1", "est"],
  resDECm1$output@params_table["m2", "est"]
)
names(DEC_t12_t21_m2_results) = paste("DEC_t12_t21_m2_", param_names, sep="")

DECj_t12_t21_m2_results = c(
  resDECJm1$total_loglikelihood,
  resDECJm1$output@params_table["d", "est"],
```

```
  resDECJm1$output@params_table["e", "est"],
  resDECJm1$output@params_table["j", "est"],
  resDECJm1$output@params_table["t12", "est"],
  resDECJm1$output@params_table["t21", "est"],
  resDECJm1$output@params_table["m1", "est"],
  resDECJm1$output@params_table["m2", "est"]
)
names(DECj_t12_t21_m2_results) = paste("DECj_t12_t21_m2_", param_names, sep="")


DIVALIKE_results = c(
  resDIVALIKE$total_loglikelihood,
  resDIVALIKE$output@params_table["d", "est"],
  resDIVALIKE$output@params_table["e", "est"],
  resDIVALIKE$output@params_table["j", "est"],
  resDIVALIKE$output@params_table["t12", "est"],
  resDIVALIKE$output@params_table["t21", "est"],
  resDIVALIKE$output@params_table["m1", "est"],
  resDIVALIKE$output@params_table["m2", "est"]
)
names(DIVALIKE_results) = paste("DIVALIKE_", param_names, sep="")


DIVALIKEj_results = c(
  resDIVALIKEj$total_loglikelihood,
  resDIVALIKEj$output@params_table["d", "est"],
  resDIVALIKEj$output@params_table["e", "est"],
  resDIVALIKEj$output@params_table["j", "est"],
  resDIVALIKEj$output@params_table["t12", "est"],
  resDIVALIKEj$output@params_table["t21", "est"],
  resDIVALIKEj$output@params_table["m1", "est"],
  resDIVALIKEj$output@params_table["m2", "est"]
)
names(DIVALIKEj_results) = paste("DIVALIKEj_", param_names, sep="")

DIVALIKE_t12_t21_m2_results = c(
  resDIVALIKEm1$total_loglikelihood,
  resDIVALIKEm1$output@params_table["d", "est"],
  resDIVALIKEm1$output@params_table["e", "est"],
  resDIVALIKEm1$output@params_table["j", "est"],
  resDIVALIKEm1$output@params_table["t12", "est"],
  resDIVALIKEm1$output@params_table["t21", "est"],
  resDIVALIKEm1$output@params_table["m1", "est"],
  resDIVALIKEm1$output@params_table["m2", "est"]
)
names(DIVALIKE_t12_t21_m2_results) = paste("DIVALIKE_t12_t21_m2_", param_names, sep="")

DIVALIKEj_t12_t21_m2_results = c(
  resDIVALIKEjm1$total_loglikelihood,
  resDIVALIKEjm1$output@params_table["d", "est"],
  resDIVALIKEjm1$output@params_table["e", "est"],
  resDIVALIKEjm1$output@params_table["j", "est"],
  resDIVALIKEjm1$output@params_table["t12", "est"],
  resDIVALIKEjm1$output@params_table["t21", "est"],
  resDIVALIKEjm1$output@params_table["m1", "est"],
  resDIVALIKEjm1$output@params_table["m2", "est"]
)
names(DIVALIKEj_t12_t21_m2_results) = paste("DIVALIKEj_t12_t21_m2_", param_names, sep="")

BAYAREALIKE_results = c(
  resBAYAREALIKE$total_loglikelihood,
  resBAYAREALIKE$output@params_table["d", "est"],
  resBAYAREALIKE$output@params_table["e", "est"],
  resBAYAREALIKE$output@params_table["j", "est"],
  resBAYAREALIKE$output@params_table["t12", "est"],
  resBAYAREALIKE$output@params_table["t21", "est"],
  resBAYAREALIKE$output@params_table["m1", "est"],
  resBAYAREALIKE$output@params_table["m2", "est"]
)
names(BAYAREALIKE_results) = paste("BAYAREALIKE_", param_names, sep="")
```

```
BAYAREALIKEj_results = c(
  resBAYAREALIKEj$total_loglikelihood,
  resBAYAREALIKEj$output@params_table["d", "est"],
  resBAYAREALIKEj$output@params_table["e", "est"],
  resBAYAREALIKEj$output@params_table["j", "est"],
  resBAYAREALIKEj$output@params_table["t12", "est"],
  resBAYAREALIKEj$output@params_table["t21", "est"],
  resBAYAREALIKEj$output@params_table["m1", "est"],
  resBAYAREALIKEj$output@params_table["m2", "est"]
)
names(BAYAREALIKEj_results) = paste("BAYAREALIKEj_", param_names, sep="")

BAYAREALIKE_t12_t21_m2_results = c(
  resBAYAREALIKEm1$total_loglikelihood,
  resBAYAREALIKEm1$output@params_table["d", "est"],
  resBAYAREALIKEm1$output@params_table["e", "est"],
  resBAYAREALIKEm1$output@params_table["j", "est"],
  resBAYAREALIKEm1$output@params_table["t12", "est"],
  resBAYAREALIKEm1$output@params_table["t21", "est"],
  resBAYAREALIKEm1$output@params_table["m1", "est"],
  resBAYAREALIKEm1$output@params_table["m2", "est"]
)
names(BAYAREALIKE_t12_t21_m2_results) = paste("BAYAREALIKE_t12_t21_m2_", param_names, sep="")

BAYAREALIKEj_t12_t21_m2_results = c(
  resBAYAREALIKEjm1$total_loglikelihood,
  resBAYAREALIKEjm1$output@params_table["d", "est"],
  resBAYAREALIKEjm1$output@params_table["e", "est"],
  resBAYAREALIKEjm1$output@params_table["j", "est"],
  resBAYAREALIKEjm1$output@params_table["t12", "est"],
  resBAYAREALIKEjm1$output@params_table["t21", "est"],
  resBAYAREALIKEjm1$output@params_table["m1", "est"],
  resBAYAREALIKEjm1$output@params_table["m2", "est"]
)
names(BAYAREALIKEj_t12_t21_m2_results) = paste("BAYAREALIKEj_t12_t21_m2_", param_names, sep="")



tmp_results   =   c(Trait_2rates_results,   DEC_results,   DECj_results,   DEC_t12_t21_m2_results,
DECj_t12_t21_m2_results,
               DIVALIKE_results,         DIVALIKEj_results,         DIVALIKE_t12_t21_m2_results,
DIVALIKEj_t12_t21_m2_results,
               BAYAREALIKE_results,     BAYAREALIKEj_results,     BAYAREALIKE_t12_t21_m2_results,
BAYAREALIKEj_t12_t21_m2_results)
tmp_results_mat = matrix(tmp_results, nrow = 13, ncol=length(param_names), byrow=T)
tmp_results_mat = data.frame(tmp_results_mat, stringsAsFactors=FALSE)
colnames(tmp_results_mat) <- param_names
rownames(tmp_results_mat)     <-     c("Trait_2rates",     "DEC",     "DECj",     "DEC_t12_t21_m1",
"DECj_t12_t21_m2",   "DIVALIKE",   "DIVALIKEj",   "DIVALIKE_t12_t21_m2",   "DIVALIKEj_t12_t21_m2",
"BAYAREALIKE", "BAYAREALIKEj", "BAYAREALIKE_t12_t21_m2", "BAYAREALIKEj_t12_t21_m2")

outfn = "params_inferred.txt"
write.table(x=tmp_results_mat,    file=outfn,    append=FALSE,    row.names=T,    col.names=TRUE,
quote=FALSE, sep="\t")
#moref(outfn)



######################################################################
######################################################################
######################################################################
######################################################################
#
# CALCULATE SUMMARY STATISTICS TO COMPARE
# DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J
#
######################################################################
######################################################################
######################################################################
```

```
######################################################################

######################################################################
######################################################################
# REQUIRED READING:
#
# Practical advice / notes / basic principles on statistical model
#    comparison in general, and in BioGeoBEARS:
# http://phylo.wikidot.com/advice-on-statistical-model-comparison-in-biogeobears
######################################################################
######################################################################

# Set up empty tables to hold the statistical results
restable = NULL
teststable = NULL

#######################################################
# Statistics -- 2-rates trait model

Lnl_Trait = get_LnL_from_BioGeoBEARS_results_object(resTrait_2rates)



#######################################################
# Statistics -- DEC vs. DEC+J
#######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDEC) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDECj) + Lnl_Trait

numparams1 = 5
numparams2 = 4
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DEC, null model for Likelihood Ratio Test (LRT)
res2 = extract_params_from_BioGeoBEARS_results_object(results_object=resDEC, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)
# DEC+J, alternative model for Likelihood Ratio Test (LRT)
res1 = extract_params_from_BioGeoBEARS_results_object(results_object=resDECj, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)

# The null hypothesis for a Likelihood Ratio Test (LRT) is that two models
# confer the same likelihood on the data. See: Brian O'Meara's webpage:
# http://www.brianomeara.info/tutorials/aic
# ...for an intro to LRT, AIC, and AICc

rbind(res2, res1)
tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)


#######################################################
# Statistics -- DIVALIKE vs. DIVALIKE+J
#######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKE) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKEj) + Lnl_Trait

numparams1 = 5
numparams2 = 4
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DIVALIKE, null model for Likelihood Ratio Test (LRT)
res2         =         extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
```

```
# DIVALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1         =         extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)

#######################################################
# Statistics -- BAYAREALIKE vs. BAYAREALIKE+J
#######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKE) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKEj) + Lnl_Trait

numparams1 = 5
numparams2 = 4
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# BAYAREALIKE, null model for Likelihood Ratio Test (LRT)
res2        =        extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# BAYAREALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1       =       extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)

#######################################################################
# RESULTS: DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J
#######################################################################
teststable$alt = c("DEC+J", "DIVALIKE+J", "BAYAREALIKE+J")
teststable$null = c("DEC", "DIVALIKE", "BAYAREALIKE")
row.names(restable) = c("DEC", "DEC+J", "DIVALIKE", "DIVALIKE+J", "BAYAREALIKE", "BAYAREALIKE+J")

# Look at the results!!
restable
teststable


#######################################################
# Statistics -- DECm vs. DECm+J
#######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDECm1)
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDECJm1)

numparams1 = 6
numparams2 = 5
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DEC, null model for Likelihood Ratio Test (LRT)
res2 = extract_params_from_BioGeoBEARS_results_object(results_object=resDEC, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)
# DEC+J, alternative model for Likelihood Ratio Test (LRT)
res1 = extract_params_from_BioGeoBEARS_results_object(results_object=resDECj, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)
```

```
# The null hypothesis for a Likelihood Ratio Test (LRT) is that two models
# confer the same likelihood on the data. See: Brian O'Meara's webpage:
# http://www.brianomeara.info/tutorials/aic
# ...for an intro to LRT, AIC, and AICc

rbind(res2, res1)
tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)


######################################################
# Statistics -- DIVALIKEm1 vs. DIVALIKEm1+J
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKEm1)
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKEjm1)

numparams1 = 6
numparams2 = 5
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DIVALIKE, null model for Likelihood Ratio Test (LRT)
res2          =          extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# DIVALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1         =         extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)

######################################################
# Statistics -- BAYAREALIKEm1 vs. BAYAREALIKEm1+J
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKEm1)
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKEjm1)

numparams1 = 6
numparams2 = 5
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# BAYAREALIKE, null model for Likelihood Ratio Test (LRT)
res2        =        extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# BAYAREALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1       =       extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)

######################################################################
# RESULTS: DECm1, DECm1+J, DIVALIKEm1, DIVALIKEm1+J, BAYAREALIKEm1, BAYAREALIKEm1+J
```

```
#######################################################################
teststable$alt = c("DEC+m1+J", "DIVALIKE+m1+J", "BAYAREALIKE+m1+J")
teststable$null = c("DEC+m1", "DIVALIKE+m1", "BAYAREALIKE+m1")
row.names(restable) = c("DEC+m1", "DEC+m1+J", "DIVALIKE+m1", "DIVALIKE+m1+J", "BAYAREALIKE+m1",
"BAYAREALIKE+m1+J")

# Look at the results!!
restable
teststable


######################################################
# Statistics -- DEC vs. DECm1
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDEC) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDECm1)

numparams1 = 5
numparams2 = 4
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DEC, null model for Likelihood Ratio Test (LRT)
res2 = extract_params_from_BioGeoBEARS_results_object(results_object=resDEC, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)
# DEC+J, alternative model for Likelihood Ratio Test (LRT)
res1 = extract_params_from_BioGeoBEARS_results_object(results_object=resDECj, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)

# The null hypothesis for a Likelihood Ratio Test (LRT) is that two models
# confer the same likelihood on the data. See: Brian O'Meara's webpage:
# http://www.brianomeara.info/tutorials/aic
# ...for an intro to LRT, AIC, and AICc

rbind(res2, res1)
tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)

######################################################
# Statistics -- DEC+J vs. DECm1+J
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDECj) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDECJm1)

numparams1 = 6
numparams2 = 5
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DEC, null model for Likelihood Ratio Test (LRT)
res2 = extract_params_from_BioGeoBEARS_results_object(results_object=resDEC, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)
# DEC+J, alternative model for Likelihood Ratio Test (LRT)
res1 = extract_params_from_BioGeoBEARS_results_object(results_object=resDECj, returnwhat="table",
addl_params=c("j"), paramsstr_digits=4)

# The null hypothesis for a Likelihood Ratio Test (LRT) is that two models
# confer the same likelihood on the data. See: Brian O'Meara's webpage:
# http://www.brianomeara.info/tutorials/aic
# ...for an intro to LRT, AIC, and AICc

rbind(res2, res1)
tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
```

```
teststable = rbind(teststable, tmp_tests)

######################################################
# Statistics -- DIVALIKE vs. DIVALIKEm1
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKE) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKEm1)

numparams1 = 5
numparams2 = 4
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DIVALIKE, null model for Likelihood Ratio Test (LRT)
res2          =          extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# DIVALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1         =         extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)

######################################################
# Statistics -- DIVALIKE+J vs. DIVALIKEm1+J
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKEj) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resDIVALIKEjm1)

numparams1 = 6
numparams2 = 5
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# DIVALIKE, null model for Likelihood Ratio Test (LRT)
res2          =          extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# DIVALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1         =         extract_params_from_BioGeoBEARS_results_object(results_object=resDIVALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)



######################################################
# Statistics -- BAYAREALIKE vs. BAYAREALIKEm1
######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKE) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKEm1)

numparams1 = 5
numparams2 = 4
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
```

```
stats

# BAYAREALIKE, null model for Likelihood Ratio Test (LRT)
res2        =        extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# BAYAREALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1        =       extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)


#######################################################
# Statistics -- BAYAREALIKE+J vs. BAYAREALIKEm1+J
#######################################################
# We have to extract the log-likelihood differently, depending on the
# version of optim/optimx
LnL_2 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKEj) + Lnl_Trait
LnL_1 = get_LnL_from_BioGeoBEARS_results_object(resBAYAREALIKEjm1)

numparams1 = 6
numparams2 = 5
stats = AICstats_2models(LnL_1, LnL_2, numparams1, numparams2)
stats

# BAYAREALIKE, null model for Likelihood Ratio Test (LRT)
res2        =        extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKE,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)
# BAYAREALIKE+J, alternative model for Likelihood Ratio Test (LRT)
res1        =       extract_params_from_BioGeoBEARS_results_object(results_object=resBAYAREALIKEj,
returnwhat="table", addl_params=c("j"), paramsstr_digits=4)

rbind(res2, res1)
conditional_format_table(stats)

tmp_tests = conditional_format_table(stats)

restable = rbind(restable, res2, res1)
teststable = rbind(teststable, tmp_tests)


#######################################################
# Save the results tables for later -- check for e.g.
# convergence issues
#######################################################

# Loads to "restable"
save(restable, file="restable_v1.Rdata")
load(file="restable_v1.Rdata")

# Loads to "teststable"
save(teststable, file="teststable_v1.Rdata")
load(file="teststable_v1.Rdata")

# Also save to text files
write.table(restable, file="restable.txt", quote=FALSE, sep="\t")
write.table(unlist_df(teststable), file="teststable.txt", quote=FALSE, sep="\t")



#######################################################
# Model weights of all twelve models
#######################################################
```

```
restable2 <- matrix(nrow = 12, ncol = 5)
rownames(restable2)    <-    c("DEC",    "DEC+j",    "DIVALIKE",    "DIVALIKE+j",    "BAYAREALIKE",
"BAYAREALIKE+j",   "DEC+m1",   "DEC+m1+j",   "DIVALIKE+m1",   "DIVALIKE+m1+j",   "BAYAREALIKE+m1",
"BAYAREALIKE+m1+j")
colnames(restable2) <- c("LnL", "numparams", "d", "e", "j")

restable2[1, 1:2] <- c(-283.621788451647, 4)
restable2[2,  1:2] <- c(-126.924979277634, 5)
restable2[3,  1:2] <- c(-285.108501410914, 4)
restable2[4,  1:2] <- c(-128.841851060496, 5)
restable2[5,  1:2] <- c(-301.625076988781, 4)
restable2[6,  1:2] <- c(-129.595506197404, 5)
restable2[7,  1:2] <- c(-168.30089969136, 5)
restable2[8,  1:2] <- c(-109.218367236974, 6)
restable2[9,  1:2] <- c(-131.722992279787, 5)
restable2[10,  1:2] <- c(-110.679687067305, 6)
restable2[11,  1:2] <- c(-172.262757133283, 5)
restable2[12,  1:2] <- c(-110.847283523589, 6)

restable2 <- as.data.frame(restable2)

# With AICs:
AICtable = calc_AIC_column(LnL_vals=restable2$LnL, nparam_vals=restable2$numparams)
restable2 = cbind(restable2, AICtable)
restable_AIC_rellike = AkaikeWeights_on_summary_table(restable=restable2, colname_to_use="AIC")
restable_AIC_rellike

# With AICcs -- factors in sample size
samplesize = length(tr$tip.label)
AICtable     =     calc_AICc_column(LnL_vals=restable2$LnL,     nparam_vals=restable2$numparams,
samplesize=samplesize)
restable2 = cbind(restable2, AICtable)
restable_AICc_rellike = AkaikeWeights_on_summary_table(restable=restable2, colname_to_use="AIC")
restable_AICc_rellike
free_params = row.names(resDECj$output@params_table[resDECj$output@params_table$type=="free",])
names(restable_AICc_rellike) = c("LnL", "numparams", free_params, "AICc", "AICc_wt")

# Also save to text files
write.table(restable_AIC_rellike, file="restable_AIC_rellike.txt", quote=FALSE, sep="\t")
write.table(restable_AICc_rellike, file="restable_AICc_rellike.txt", quote=FALSE, sep="\t")

# Save with nice conditional formatting
write.table(conditional_format_table(restable_AIC_rellike),
file="restable_AIC_rellike_formatted.txt", quote=FALSE, sep="\t")
write.table(conditional_format_table(restable_AICc_rellike),
file="restable_AICc_rellike_formatted.txt", quote=FALSE, sep="\t")


######################################################
# Extract marginal probabilities for key nodes from the best-fitting model (in this case,
DEC+j+t12+t21+m1)

MLAnc_Nodes <- resDECj_wTraits$ML_marginal_prob_each_state_at_branch_top_AT_node
colnames(MLAnc_Nodes) <- ranges_list

tree <- read.tree(trfn)


########
# Total-group Osteoglossomorpha (root node)

node.number <- length(tree$tip.label)+1
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]


########
```

```
# Crown Osteoglossomorpha (Hiodon_alosoides + Osteoglossum_bicirrhosum)

node.number <- findMRCA(tree, c("Hiodon_alosoides", "Osteoglossum_bicirrhosum"), type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]

########
# Crown Osteoglossiformes (Pantodon_buccholzi + Osteoglossum_bicirrhosum)

node.number <- findMRCA(tree, c("Pantodon_buchholzi", "Osteoglossum_bicirrhosum"), type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]

########
# Crown Osteoglossidae (Arapaima_gigas + Osteoglossum_bicirrhosum)

node.number <- findMRCA(tree, c("Arapaima_gigas", "Osteoglossum_bicirrhosum"), type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]

########
# Crown Arapaiminae (Arapaima_gigas + Heterotis_niloticus)

node.number <- findMRCA(tree, c("Arapaima_gigas", "Heterotis_niloticus"), type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]


########
# Crown Osteoglossinae (Osteoglossum_bicirrhosum + Scleropages_leichardti)

node.number    <-    findMRCA(tree,    c("Osteoglossum_bicirrhosum",    "Scleropages_leichardti"),
type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]


########
# Crown Notopteridae (Notopterus_notopterus + Papyrocranus_afer)

node.number <- findMRCA(tree, c("Notopterus_notopterus", "Papyrocranus_afer"), type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]

########
# Crown Notopteroidei (Notopterus_notopterus + Mormyrus_ovis)
```

```
node.number <- findMRCA(tree, c("Notopterus_notopterus", "Mormyrus_ovis"), type="node")
MLProbAnc <- MLAnc_Nodes[node.number, ]
names(MLProbAnc) <- colnames(MLAnc_Nodes)

sum(MLProbAnc)    # CHECK: This should always be 1

MLProbAnc[MLProbAnc > 0.01]
```

APPENDIX D

Results from IQTREE2 and BioGeoBEARS Analyses from Chapter 5

Fig. D.1. Maximum likelihood phylogenetic tree of Osteoglossomorpha, obtained with IQTREE2 before pruning rogue taxa. Node numbers indicate ultrafast bootstrap support.

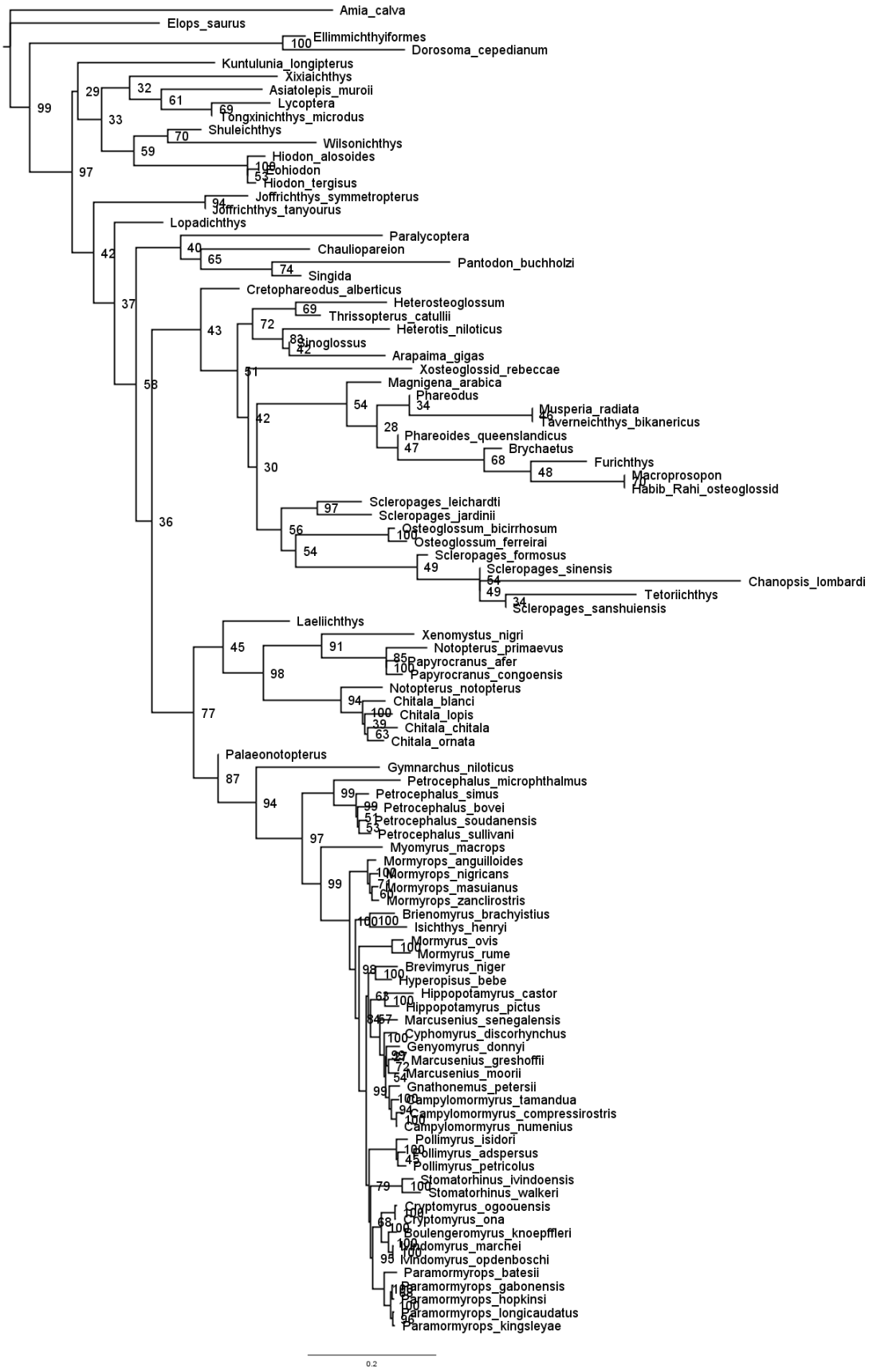
Amia_calva
Elops_saurus
Ellimmichthyiformes
Dorosoma_cepedianum
Kuntulunia_longipterus
Xixiaichthys
Asiatolepis_muroii
Lycoptera
Tongxinichthys_microdus
Shuleichthys
Wilsonichthys
Hiodon_alosoides
Eohiodon
Hiodon_tergisus
Joffrichthys_symmetropterus
Joffrichthys_tanyourus
Lopadichthys
Paralycoptera
Chauliopareion
Pantodon_buchholzi
Singida
Cretophareodus_alberticus
Heterosteoglossum
Thrissopterus_catullii
Heterotis_niloticus
Arapaima_gigas
Xosteoglossid_rebeccae
Magnigena_arabica
Phareodus
Musperia_radiata
Taverneichthys_bikanericus
Phareoides_queenslandicus
Brychaetus
Furichthys
Macroprosopon
Habib_Rahi_osteoglossid
Scleropages_leichardti
Scleropages_jardinii
Osteoglossum_bicirrhosum
Osteoglossum_ferreirai
Scleropages_formosus
Scleropages_sinensis
Chanopsis_lombardi
Tetoriichthys
Scleropages_sanshuiensis
Laeliichthys
Xenomystus_nigri
Notopterus_primaevus
Papyrocranus_afer
Papyrocranus_congoensis
Notopterus_notopterus
Chitala_blanci
Chitala_lopis
Chitala_chitala
Chitala_ornata
Palaeonotopterus
Gymnarchus_niloticus
Petrocephalus_microphthalmus
Petrocephalus_simus
Petrocephalus_bovei
Petrocephalus_soudanensis
Petrocephalus_sullivani
Myomyrus_macrops
Mormyrops_anguilloides
Mormyrops_nigricans
Mormyrops_masuianus
Mormyrops_zanclirostris
Brienomyrus_brachyistius
Isichthys_henryi
Mormyrus_ovis
Mormyrus_rume
Brevimyrus_niger
Hyperopisus_bebe
Hippopotamyrus_castor
Hippopotamyrus_pictus
Marcusenius_senegalensis
Cyphomyrus_discorhynchus
Genyomyrus_donnyi
Marcusenius_greshoffii
Marcusenius_moorii
Gnathonemus_petersii
Campylomormyrus_tamandua
Campylomormyrus_compressirostris
Campylomormyrus_numenius
Pollimyrus_isidori
Pollimyrus_adspersus
Pollimyrus_petricolus
Stomatorhinus_ivindoensis
Stomatorhinus_walkeri
Cryptomyrus_ogooensis
Cryptomyrus_ona
Boulengeromyrus_knoepffleri
Ivindomyrus_marchei
Ivindomyrus_opdenboschi
Paramormyrops_batesii
Paramormyrops_gabonensis
Paramormyrops_hopkinsi
Paramormyrops_longicaudatus
Paramormyrops_kingsleyae
0.2

Fig. D.2. Ancestral geographic range estimates under the best-fitting model (DEC+*j)* with the 'MarineAsArea' strategy, applied to the Bayesian consensus tree with minimum node ages.

BioGeoBEARS DEC+J on Osteoglossomorpha – MarineAsArea strategy – MinimumNodeAges
ancstates: global optim, 3 areas max. d=0; e=0; j=0.0195; LnL=–94.15

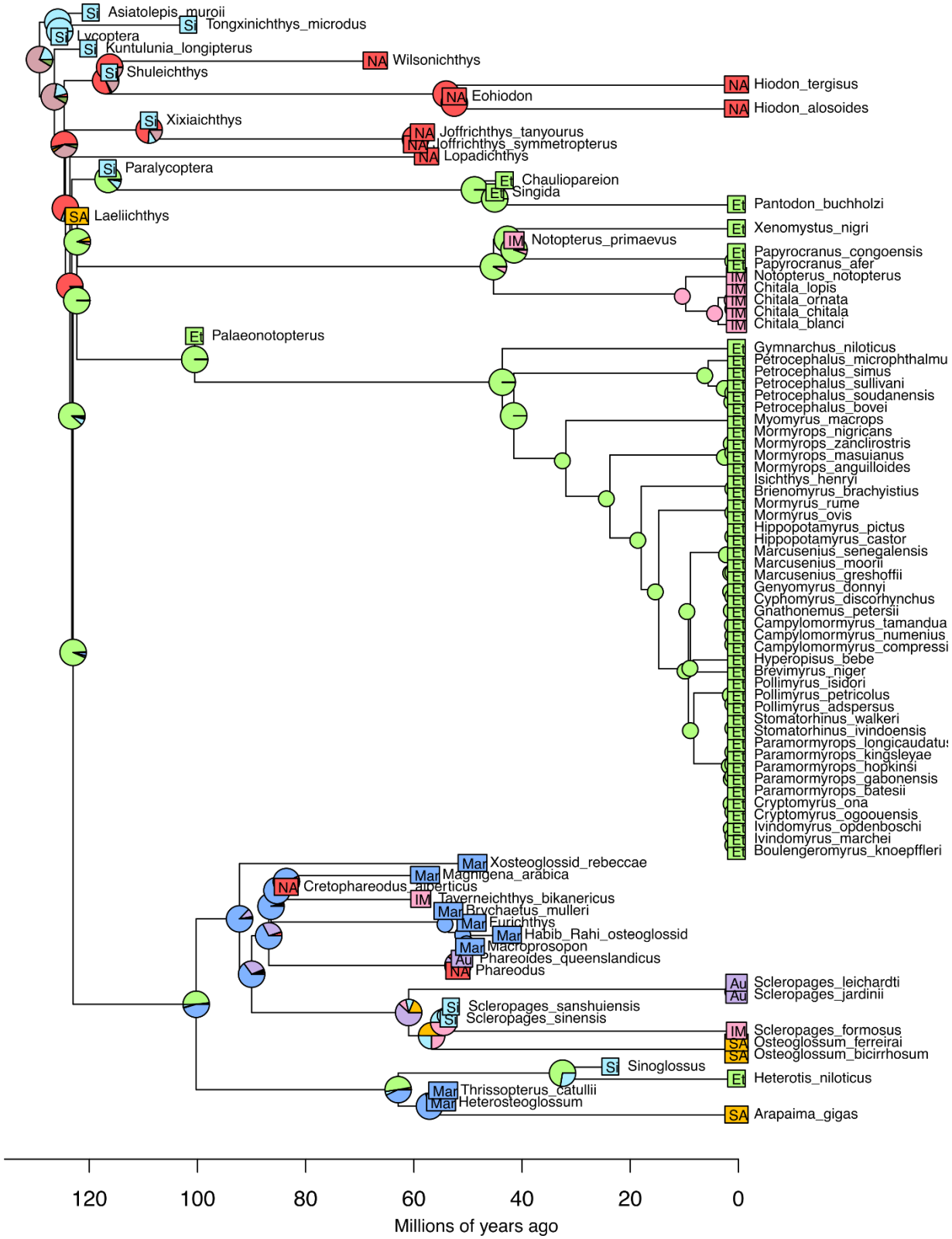

Fig. D.3. Ancestral geographic range estimates under the best-fitting model (DEC+*j)* with the 'MarineAsArea' strategy, applied to the Bayesian consensus tree with maximum node ages.

BioGeoBEARS DEC+J on Osteoglossomorpha – MarineAsArea strategy – MaximumNodeAges
ancstates: global optim, 3 areas max. d=0; e=0; j=0.0195; LnL=–94.15