A fossil-calibrated time-tree of all Australian freshwater fishes

Amy R. Tims[a,*], Peter J. Unmack[b], Simon Y. W. Ho[c]

[a] *Department of Biological Sciences, Macquarie University, Sydney, Australia*

[b] *Center for Applied Water Science, Institute for Applied Ecology, University of Canberra, Canberra, Australia*

[c] *School of Life and Environmental Sciences, University of Sydney, Sydney, Australia*

[*] Corresponding author. *E-mail address*: amy.tims@students.mq.edu.au (A. Tims).

Abstract

Australian freshwater fishes are a relatively species-poor assemblage, mostly comprising groups derived from older repeated freshwater invasions by marine ancestors, plus a small number of Gondwanan lineages. These taxa are both highly endemic and highly threatened, but a comprehensive phylogeny for Australian freshwater fishes is lacking. This has hampered efforts to study their phylogenetic diversity, distribution of extinction risk, speciation rates, and rates of trait evolution. Here, we present a comprehensive dated phylogeny of 412 Australian fishes. We include all formally recognized freshwater species plus a number of genetically distinct subpopulations, species awaiting formal description, and predominantly brackish-water species that sometimes enter fresh water. The phylogeny was inferred using maximum-likelihood analysis of a multilocus data set comprising six mitochondrial and three nuclear genes from 326 taxa. We inferred the evolutionary timescale using penalized likelihood, then used a statistical approach to add 86 taxa for which no molecular data were available. The time-tree inferred in our study will provide a useful resource for macroecological studies of Australian freshwater fishes by enabling corrections for phylogenetic non-independence in evolutionary and ecological comparative analyses.

Keywords: freshwater fishes; phylogenetic tree; fossil calibration; molecular dating; Gondwana; comparative analysis

1. Introduction

The Australian continent is home to approximately 360 fish species that are dependent on fresh water for at least part of their life cycle (Unmack, 2018). Of these species, 253 have been described and a further 62 to 107 are as yet undescribed, with recent molecular studies revealing large numbers of cryptic species (e.g., Adams et al., 2014, 2013; Hammer et al., 2007; Shelley et al., 2018). The habitats of these species span all aquatic environments, from Australia's highest to lowest elevations, and from alpine streams and lakes to large temperate and tropical floodplain rivers and smaller desert waterholes and springs (Allen et al., 2002). The umbrella term 'freshwater fishes' can also apply to a number of predominantly marine or estuarine species that frequently enter the lower reaches of freshwater rivers and are sometimes found appreciable distances upstream, though these species are not strictly dependent on freshwater environments. Even when including ~100 of these euryhaline species, Australia's freshwater fish diversity is considerably lower than that of other landmasses of similar size such as Europe (540 species; Kottelat and Freyhof, 2007), the United States (890 species; Froese and Pauly, 2000), and China (1300 species; Xing et al., 2016).

Australian freshwater fishes do not form a monophyletic group, but rather are an assemblage derived from a number of old independent freshwater invasions by marine lineages in addition to the ancient lineages of Gondwana. The fauna is notable in that it is composed predominantly of acanthopterygians rather than the ostariophysian fishes which dominate freshwater environments on other continents (Unmack, 2013). This faunal composition is an outcome of Australia's geologic and climatic history and long geographical isolation. Together with Africa, South America, India, and Antarctica, Australia was once part of the supercontinent Gondwana, which began to fragment in the Mesozoic. Australia began to separate from Antarctica around 95 Ma (Veevers, 1986) and was completely isolated by around 37 Ma (Veevers, 1984). For some reason, ostariophysian fishes appear to have been historically absent from Australia. Because

most ostariophysian fishes require fresh water for dispersal and cannot cross marine barriers, this group has been unable to invade Australia except for two families of catfishes (Siluriformes) with marine representatives: Plotosidae and Ariidae.

Australian freshwater fish lineages vary widely in age. At least two families have origins in Gondwana, evidenced by their fossil records and present-day distributions. The Queensland lungfish, *Neoceratodus forsteri*, is possibly the world's oldest surviving vertebrate species, with virtually identical fossils (based on their toothplates) dating back to the Cretaceous found in present-day Australia and South America (Cavin et al., 2007; Kemp and Molnar, 1981). The bony tongues (Osteoglossidae), of which there are two Australian species *Scleropages jardinii* and *S. leichardti*, have a fossil record dating back around 50 Ma (Cioffi et al., 2019; Zhang and Wilson, 2017) and have extant representatives in South America, Africa, and South-East Asia. This distribution suggests that the family's origins predate the breakup of Gondwana. A Gondwanan origin is possible for at least one other family, Galaxiidae, which is distributed through Australia, New Zealand, South Africa, and southern South America. However, this family also has members that spend part of their life cycle at sea and it is unclear if their distribution is related to marine dispersal or continental breakup (see Burridge et al., 2012). The paucity of Gondwanan lineages in Australia is likely to be a consequence of the increasing aridity of the continent through the Cenozoic and its increasingly hostile conditions for freshwater-dependent fauna. Instead, the majority of the Australian freshwater fishes are derived from predominantly marine families that invaded freshwater environments after the breakup of Gondwana. Indeed, Australia is home to the only freshwater representatives of a number of fish families.

In view of their evolutionary origins, it is unsurprising that Australian freshwater fishes are highly endemic. Over one-third of genera and nearly 75% of species are unique to Australia. While some families, such as gudgeons (Eleotridae), are widespread in Australia, species assemblages are generally spatially distinct. In the tropical north, they are dominated by rainbowfishes

77 (Melanotaeniidae), grunters (Terapontidae), gobies (Gobiidae), and catfishes (Ariidae and
78 Plotosidae). The southern temperate fauna, by contrast, is dominated by galaxiids (Galaxiidae), and
79 freshwater basses and cods (Percichthyidae).

80 In addition to being species-poor and highly endemic, Australian freshwater fishes are under
81 threat from invasive species, climate change, and natural system modifications. Recent assessments
82 of Australian freshwater fishes found 37% of species to be threatened (Critically Endangered,
83 Endangered, or Vulnerable) and a further 7% to be Near Threatened (IUCN; www.iucnredlist.org).
84 Almost one-third of threatened species are recently discovered, having been described only in the
85 past decade or remaining undescribed. Thus, there are likely to be undescribed species that are
86 potentially at risk or that have already gone extinct.

87 The origins and diversification of Australian freshwater fishes can be understood using a
88 phylogenetic approach. Phylogenetic trees form the basis of large-scale studies of distribution,
89 phylogenetic diversity, speciation rates, and rates of trait evolution (e.g., Jetz et al., 2012; Rabosky et
90 al., 2018). Knowledge of the phylogeny also enables quantification of evolutionary distinctiveness
91 (Isaac et al., 2007), a common metric in evaluating conservation priorities. The life-history traits of
92 many Australian freshwater fishes are poorly known; phylogenetic relationships can be used to
93 predict traits, including those that covary with extinction risk (e.g., Purvis et al., 2005), based on data
94 from related species that have been better characterized. Additionally, the integration of ecological
95 and evolutionary information allows quantification of rates of niche evolution, which is important for
96 an approach that holds promise in species-distribution modelling in the context of changes in
97 climate and land use (e.g., Quintero and Wiens, 2013). Therefore, knowledge of phylogeny helps to
98 facilitate studies of threatened status and conservation priority, and can be used to inform
99 conservation management.

100 Here, we present the most taxonomically complete phylogenetic analysis of Australian
101 freshwater fishes, including 412 fresh- and brackish-water species. Our data set includes all known

102 species of freshwater fishes, as well as a number of euryhaline relatives that are commonly found in
103 the lower reaches of freshwater rivers. Using a penalized-likelihood approach, we infer the
104 evolutionary timescale by calibrating the phylogeny with 25 fossil-based age constraints. Taxa for
105 which no molecular data are available are grafted into the tree using a probabilistic method, which
106 leverages existing taxonomic information while accounting for phylogenetic uncertainty. Our study
107 provides a set of 10,000 dated phylogenetic trees that account for uncertainties in the topology and
108 divergence times. We show that the majority of freshwater fish families were established in
109 Australia during the first half of the Cenozoic, with the divergences among a number of genera
110 having occurred in the last 10–15 Myr as the drying climate increased barriers to gene flow.

111

112 **2. Materials and methods**

113 *2.1. List of Australian freshwater fish species*

114 We compiled a comprehensive list of Australian freshwater fish species that includes: all
115 species that are strictly freshwater and never occur in estuarine or marine environments; species
116 that occur primarily in fresh water but also descend into tidal reaches and estuaries; and
117 diadromous species that spend parts of their life cycle in both freshwater and marine environments.
118 A number of cryptic species have been discovered on the basis of recent genetic work, and these are
119 denoted by the names of their parent species with Roman numerals appended. We also included a
120 number of predominantly estuarine species that are commonly found in fresh water. The inclusion
121 of estuarine species here is somewhat arbitrary, but is based primarily on Allen et al. (2002) and
122 Humphries and Walker (2013). Our list contains a total of 412 taxa.

123

124 *2.2. DNA sequence alignment*

We use Batch Entrez to download DNA sequences of six mitochondrial genes (*12S*, *16S*, *COI*, *CYTB*, *ND2*, and *ND4*) and three nuclear genes (*18S*, *RAG1*, and *RAG2*). We chose these genes based on their availability for at least 50 species in our taxon list. Of the 412 taxa examined, 301 had sequence data for at least one gene available on GenBank (Supplementary Table S1). Additional unpublished data for 109 species were obtained by the second author; 25 of these species had no records available on GenBank. The final data matrix had 49.9% occupancy, with a total of 1463 sequences being included for 9 genes from 326 taxa.

We aligned protein-coding sequences by their translations in Geneious Prime 2019.0.3 (https://www.geneious.com) using ClustalW (Thompson et al., 1994). Sequences of rRNA genes were aligned using E-INS-i settings in MAFFT (Katoh et al., 2019). All sequence alignments were checked by eye in Geneious Prime and obvious misalignments were corrected manually. We removed sections that were poorly conserved or that had a high proportion of missing data using TrimAl v1.2 (Capella-Gutiérrez et al., 2009). We selected a gap threshold based on the number of sequences in each alignment such that only sites with coverage of at least 50 species were retained, and used a similarity threshold of 0.1. This had the effect of removing portions of sequences that could not be confidently aligned due to poor coverage or extreme variability, while allowing the retention of enough sequence variation to inform phylogenetic inference. The trimmed and untrimmed sequence alignments are described in Table 1 and are available on Dryad (temporary link: https://datadryad.org/stash/share/kcIe1Cu9QH6xTNKGESQKwhpdfdWwXR_VWRTEUfENmKw). The final concatenated alignment had a length of 12,059 base pairs and sitewise coverage of 42.3%.

We estimated substitution saturation in each alignment based on fully resolved sites using the method of Xia et al. (2003) as implemented in DAMBE v7.0.39 (Xia, 2018). In the protein-coding genes, the three codon positions were tested for saturation separately due to their differential rates of evolution. For the purposes of testing saturation, five short sequences were removed from the *CYTB* alignment and eight were removed from the *RAG1* alignment to increase the number of fully

resolved sites. To explore the potential effects of saturation on the inferred tree topology, we used two approaches when concatenating the sequences: the total data set, where all genes and codon sites were included even when we found evidence of saturation; and a reduced data set, which excluded genes and codon sites that were identified as saturated.

We compared data-partitioning schemes using a greedy algorithm in PartitionFinder v2.1.1 (Lanfear et al., 2017), starting with data blocks defined by gene and codon position. The optimal partitioning schemes involved 18 data subsets for the total data set and 16 data subsets for the reduced data set.

2.3. Phylogenetic analysis

We performed preliminary unconstrained maximum-likelihood tree searches in RAxML (Stamatakis, 2014), with a separate GTR+G+I substitution model assigned to each data subset. Five independent runs were performed on each of the two data sets. The highest-likelihood trees from the analysis of the total and reduced data sets were compared to identify any topological incongruences, and also to identify taxa that were placed in an incorrect family or whose positions were resolved inconsistently (i.e., placed with different taxa on different runs). To constrain placements of rogue taxa, and to enforce some nodes that have been well supported in previous studies using large data sets (e.g., Betancur-R et al., 2013; Broughton et al., 2013; Near et al., 2012; Thacker et al., 2015), we conducted 100 maximum-likelihood tree searches in RAxML using the total data set (see Results and Discussion for justification) while implementing topological constraints (see Fig. 2). We placed the root of the tree between Cyclostomata (jawless vertebrates, represented in Australia by three species of lamprey) and Gnathostomata (jawed vertebrates, all other fishes), based on evidence from morphological and molecular studies of vertebrate phylogeny (e.g., Forey and Janvier, 1993; Heimberg et al., 2010).

We dated the 100 trees from our maximum-likelihood analyses, using penalized likelihood as implemented in treePL (Smith and O'Meara, 2012). This rapid dating method was used because the size of the data set precludes the use of more computationally intensive approaches (Tao et al., 2020). The best rate-smoothing value for each tree was determined by a cross-validation procedure. Cross-validation tested seven smoothing values, starting at 10^{-4} and incrementally increasing by one order of magnitude. Internal nodes of the phylogeny were calibrated using age constraints on 24 nodes, based on fossil data compiled from an extensive literature review (Table 2). The dating method in treePL uses hard maximum and minimum age constraints. Minimum age constraints were defined by the age of the fossil itself, while maximum age constraints corresponded to those reported in the source paper. In most instances, these maximum age constraints had been determined by other fossil or geological evidence, such as the ages of stem-group fossils. In one case (Aschliman et al., 2012), the authors used penalized likelihood in conjunction with Marshall's (2008) method for estimating maximum age constraints (in this case, we used the 95% confidence interval of the node-age distribution). The age of the root was fixed at 510 Ma, in line with previous estimates for crown vertebrates (Erwin et al., 2011). The 100 time-trees containing the 326 species with molecular data are available on Dryad (https://doi.org/10.5061/dryad.vx0k6djnz)

Species for which no molecular data were available were grafted into the 100 time-calibrated trees. We applied a taxonomic constraint (genus- or family-level monophyly) to define the set of branches onto which each species could be grafted. The set of branches contained all internal branches of the clade, as well as a length of the stem branch (below the MRCA of species with molecular data) which was selected based on the lengths of internal branches (Fig. 1a). This constrained the maximum and minimum ages of the speciation event (Fig. 1b). The age of the speciation event was simulated using the *corsim* function (Cusimano et al., 2012) in the TreeSim package (Stadler, 2010) in R v3.4.3 (R Core Team, 2017) (Fig. 1c), and the species was then grafted to a randomly selected branch in the subtree at the simulated age (Fig. 1d). This was repeated 100 times on each of the 100 trees, giving a final set of 10,000 trees. We summarized these trees in the

form of a 50% majority-rule consensus tree and a maximum-clade-credibility tree. All of the time-trees produced in our analyses are available on Dryad (https://doi.org/10.5061/dryad.vx0k6djnz).

3. Results and discussion

Our analyses have produced a set of 100 dated phylogenies incorporating 326 species with available molecular data, and 10,000 dated phylogenies containing all of the species of Australian freshwater fishes. The maximum-clade-credibility tree (Fig. 2) and the 50% majority-rule consensus tree (Supplementary Fig. S1) are broadly similar, but some of the internal branches in the latter have collapsed within clades that contain species lacking molecular data. Additionally, in the maximum-clade-credibility tree, the genus *Cairnsichthys* is resolved as the sister group to Pseudomugilidae, while in the majority-rule consensus tree, *Cairnsichthys* forms a polytomy with Pseudomugilidae and the other members of Melanotaeniidae.

Both the majority-rule consensus tree and maximum-clade-credibility trees reveal strong support for the monophyly of many widely accepted groups, with almost all of the 51 families found to be monophyletic. In concordance with a recent study of Atheriniformes (Campanella et al., 2015), we found that the family Melanotaeniidae (rainbowfishes, node A) is not monophyletic due to the placement of the genus *Cairnsichthys*. Retropinnidae (node B) is placed in the same clade as Galaxiidae (node C), which is consistent with the findings of Li et al. (2010) and Smith et al. (2016). However, this result stands in contrast with that of Near et al. (2012), who suggested that Retropinnidae are part of an earlier-diverging clade, and of Betancur-R et al. (2013), who found that Galaxiidae are an earlier-diverging lineage. Our analysis yielded support for a deep divergence in species traditionally assigned to Gobiidae (node D), which has also been described in other studies (Thacker et al., 2015; Tornabene et al., 2013), and supports taxonomic reclassification to two separate families (Gobiidae and Gobionellidae).

In the shallower parts of the phylogenetic tree, many strongly supported clades described monophyletic genera. Lower support values, where found, are often near the tips of the tree in larger groups of congeneric species. These can be variously attributed to short branches (e.g., *Hypseleotris*), high proportions of missing data (e.g., *Stiphodon*), or poor constraints on the placement of species that lacked molecular data (e.g., *Gadopsis* and *Melanotaenia*). Uncertainty in genus-level placement of several species of catfish has been highlighted previously (e.g., Betancur-R, 2009; Pusey et al., 2017; Unmack, 2013), and consistent with this we find that the families Ariidae and Plotosidae contain a number of genera that are not monophyletic in our tree (e.g., *Neosilurus*, *Porochilus*, *Tandanus*, and *Neoarius*). Other genera found to be non-monophyletic were *Galaxias* (includes *Paragalaxias*; concordant with findings of Burridge et al., 2012), *Macquaria* (includes *Guyu wujalwujalensis*; also reported by Jerry et al., 2001 and Unmack, 2013), *Hypseleotris* (includes *Kimberleyeleotris hutchinsi*), and *Pseudomugil* (includes *Scaturiginichthys vermeilipinnis*).

We found some evidence of substitution saturation in the rapidly evolving third codon sites, especially in mitochondrial genes (Supplementary Table S2). This might have a negative impact on the analyses of the total data set, which yielded some resolutions of deeper evolutionary relationships that differed from those of previous studies. Most notably, unconstrained analyses of the total data set placed the Carangimorpha (node E) as the sister lineage to Ovalentaria (node F) + Eupercaria (node G), whereas previous studies resolved Eupercaria as the sister lineage to Carangimorpha + Ovalentaria (e.g., Betancur-R et al., 2017; Hughes et al., 2018; Near et al., 2013).

Removal of the saturated subsets in our reduced data set led to instability in the inferred phylogenetic tree, as indicated by a larger number of rogue taxa, reduced monophyly at the genus level, and an increase in the number of poorly supported nodes near the tips of the tree. This was particularly pronounced within the gobionellid branch of the family Gobiidae (node D); even when the more saturated sites are included, this branch has generally lower support, probably due to a high proportion of missing data. Our findings suggest that although sequences might be saturated in

249 the deeper parts of the tree, they can still be highly informative for resolving the evolutionary
250 relationships in the shallower parts of the tree. This is particularly true when some species only have
251 sequence data for one or two genes, or when the rapidly evolving third codon sites are needed to
252 distinguish between closely related species. Therefore, sequences deemed to be saturated should
253 not be dismissed out of hand as being phylogenetically uninformative, especially when the shallower
254 nodes in the tree are of interest.

255 The inclusion of saturated sites appeared primarily to affect the deeper portions of the
256 phylogeny, where relationships among major groups are not always resolved correctly. Along the
257 deep branches in the tree, heterogeneity in the evolutionary process can lead to the substitution
258 model being misspecified. In turn, this leads to the underestimation of the lengths of deeper
259 branches (Phillips, 2009). These effects can be counteracted by including carefully chosen
260 topological constraints, given that the higher phylogenetic relationships of fishes are now relatively
261 well understood. Furthermore, there are a number of fossil calibrations that can be used to
262 constrain the node times in deeper parts of the tree. Constraining the ages of deeper nodes in the
263 tree can potentially mitigate the negative effects of saturation (Duchêne et al., 2014).

264 Our inferred time-tree shows that the majority of the Australian freshwater fish fauna is Late
265 Cretaceous or younger in age. The estimates of divergence times and their 95% HPD intervals place
266 the origins of most orders in the Late Cretaceous, and the majority of currently recognized families
267 in Australia originated in the Eocene (56–34 Ma; Supplementary Table S3). This was a time when the
268 climate had become warmer and wetter, and was less hostile to freshwater fishes than the arid
269 climate of the present day. Age estimates of the last common ancestors of many large Australian
270 clades are broadly comparable to estimates from other studies. For example, our age estimate of
271 27–37 Ma for Melanotaeniidae is similar to that of 31–46 Ma estimated by Campanella et al. (2015),
272 whereas our age estimate of 28–42 Ma for Terapontidae matches the estimate of 25–39 Ma by
273 Shelley et al. (2018). Our date estimate for Galaxiidae (78–109 Ma) is older than the 52–84 Ma

reported by Burridge et al. (2012), most likely due to differences in calibration points and the choice of outgroup taxa. A number of present-day genera diverged from each other in the late Miocene, a period during which the climate became progressively drier and sea levels dropped. Altered dispersal and gene-flow patterns relating to these changes in hydrological regime are likely to have been the cause of these evolutionary radiations.

Our study has provided a set of 10,000 dated phylogenetic trees that include the full species diversity of Australian freshwater fishes, and that take into account uncertainty in the node times and evolutionary relationships. These trees will provide a potentially useful resource for ecological studies that aim to include an evolutionary perspective. Our time-tree highlights two major periods of diversification: the initial colonization of fresh water by different families through the first half of the Cenozoic; and genus-level diversification through the last 15 Myr as the climate became more arid and sea levels dropped, altering patterns of gene flow and allowing for multiple speciation events.

Acknowledgements

We thank Guillermo Ortí and Chris Burridge for their helpful comments on our paper. This work was funded by an iMQRES scholarship to AT. SYWH was funded by the Australian Research Council (FT160100167).

Competing interests

We have no competing interests.

Authors' contributions

297 **Amy Tims:** Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing -

298 Review & Editing, Visualization

299 **Peter Unmack:** Resources, Writing - Review & Editing, Supervision

300 **Simon Ho:** Methods, Writing - Review & Editing, Supervision

301

References

Adams, M., Page, T.J., Hurwood, D.A., Hughes, J.M., 2013. A molecular assessment of species boundaries and phylogenetic affinities in *Mogurnda* (Eleotridae): A case study of cryptic biodiversity in the Australian freshwater fishes. Mar. Freshw. Res. 64, 920–931. https://doi.org/10.1071/MF12237

Adams, M., Raadik, T.A., Burridge, C.P., Georges, A., 2014. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? Syst. Biol. 63, 518–533. https://doi.org/10.1093/sysbio/syu017

Allen, G.R., Midgley, S.H., Allen, M., 2002. Field Guide to the Freshwater Fishes of Australia. Western Australian Museum, Perth, Australia.

Aschliman, N.C., Nishida, M., Miya, M., Inoue, J.G., Rosana, K.M., Naylor, G.J.P., 2012. Body plan convergence in the evolution of skates and rays (Chondrichthyes: Batoidea). Mol. Phylogenet. Evol. 63, 28–42. https://doi.org/10.1016/j.ympev.2011.12.012

Bannikov, A.F., 2014. The systematic composition of the Eocene actinopterygian fish fauna from Monte Bolca, northern Italy, as known to date. Misc. Paleontol. 12, 22–34.

Benton, M., Donoghue, P., Vinther, J., Asher, R., Friedman, M., Near, T., 2015. Constraints on the timescale of animal evolutionary history. Palaeontol. Electron. 18, 1–107. https://doi.org/10.26879/424

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

Betancur-R, R., 2009. Molecular phylogenetics and evolutionary history of ariid catfishes revisited: A

326 comprehensive sampling. BMC Evol. Biol. 9. https://doi.org/10.1186/1471-2148-9-175

327 Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., Ortí, G., 2017.
328 Phylogenetic classification of bony fishes. BMC Evol. Biol. 17, 162.
329 https://doi.org/10.1186/s12862-017-0958-3

330 Broughton, R.E., Betancur-R, R., Li, C., Arratia, G., Ortí, G., 2013. Multi-locus phylogenetic analysis
331 reveals the pattern and tempo of bony fish evolution. PLOS Curr. Tree Life April 16, 1–33.
332 https://doi.org/https://dx.doi.org/10.1371%2Fcurrents.tol.2ca8041495ffafd0c92756e7524748
333 3e

334 Burridge, C.P., Mcdowall, R.M., Craw, D., Wilson, M.V.H., Waters, J.M., 2012. Marine dispersal as a
335 pre-requisite for Gondwanan vicariance among elements of the galaxiid fish fauna. J. Biogeogr.
336 39, 306–321. https://doi.org/10.1111/j.1365-2699.2011.02600.x

337 Campanella, D., Hughes, L.C., Unmack, P.J., Bloom, D.D., Piller, K.R., Ortí, G., 2015. Multi-locus fossil-
338 calibrated phylogeny of Atheriniformes (Teleostei, Ovalentaria). Mol. Phylogenet. Evol. 86, 8–
339 23. https://doi.org/10.1016/j.ympev.2015.03.001

340 Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment
341 trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.
342 https://doi.org/10.1093/bioinformatics/btp348

343 Cavin, L., Suteethorn, V., Buffetaut, E., Tong, H., 2007. A new Thai Mesozoic lungfish (Sarcopterygii,
344 Dipnoi) with an insight into post-Palaeozoic dipnoan evolution. Zool. J. Linn. Soc. 149, 141–177.
345 https://doi.org/10.1111/j.1096-3642.2007.00238.x

346 Cioffi, M. de B., Ráb, P., Ezaz, T., Bertollo, L.A.C., Lavoué, S., de Oliveira, E.A., Sember, A., Molina,
347 W.F., de Souza, F.H.S., Majtánová, Z., Liehr, T., Al-Rikabi, A.B.H., Yano, C.F., Viana, P., Feldberg,
348 E., Unmack, P., Hatanaka, T., Tanomtong, A., Perez, M.F., 2019. Deciphering the evolutionary
349 history of arowana fishes (Teleostei, Osteoglossiformes, Osteoglossidae): insight from

comparative cytogenomics. Int. J. Mol. Sci. 20. https://doi.org/10.3390/ijms20174296

Cusimano, N., Stadler, T., Renner, S.S., 2012. A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. Syst. Biol. 61, 785–792. https://doi.org/10.1093/sysbio/sys031

Duchêne, S., Lanfear, R., Ho, S.Y.W., 2014. The impact of calibration and clock-model choice on molecular estimates of divergence times. Mol. Phylogenet. Evol. 78, 277–289. https://doi.org/10.1016/j.ympev.2014.05.032

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., Peterson, K.J., 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science. 334, 1091–1097. https://doi.org/10.1126/science.1206375

Forey, P.L., Janvier, P., 1993. Agnathans and the origin of jawed vertebrates. Nature 361, 129–134. https://doi.org/10.1038/361129a0

Froese, R., Pauly, D., 2000. FishBase [WWW Document]. URL https://www.fishbase.org (accessed 5.1.17).

Hammer, M.P., Adams, M., Unmack, P.J., Walker, K.F., 2007. A rethink on *Retropinna*: Conservation implications of new taxa and significant genetic sub-structure in Australian smelts (Pisces: Retropinnidae). Mar. Freshw. Res. 58, 327–341. https://doi.org/10.1071/MF05258

Heimberg, A.M., Cowper-Sallari, R., Sémon, M., Donoghue, P.C.J., Peterson, K.J., 2010. MicroRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. Proc. Natl. Acad. Sci. U. S. A. 107, 19379–19383. https://doi.org/10.1073/pnas.1010350107

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., Xie, B., Zhou, Z., Huang, H., Chen, S., Venkatesh, B., Shi, Q., 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii)

374 based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. U. S. A. 115, 6249–6254.
375 https://doi.org/10.1073/pnas.1719358115

376 Humphries, P., Walker, K.F. (Eds.), 2013. Ecology of Australian Freshwater Fishes. CSIRO Publishing,
377 Collingwood, Australia.

378 Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C., Baillie, J.E.M., 2007. Mammals on the EDGE:
379 conservation priorities based on threat and phylogeny. PLoS One 2, e296.
380 https://doi.org/10.1371/journal.pone.0000296

381 Jerry, D.R., Elphinstone, M.S., Baverstock, P.R., 2001. Phylogenetic relationships of Australian
382 members of the family Percichthyidae inferred from mitochondrial 12S rRNA sequence data.
383 Mol. Phylogenet. Evol. 18, 335–347. https://doi.org/10.1006/mpev.2000.0871

384 Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in
385 space and time. Nature 491, 444–448. https://doi.org/10.1038/nature11631

386 Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment,
387 interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166.
388 https://doi.org/10.1093/bib/bbx108

389 Kemp, A., Molnar, R.E., 1981. *Neoceratodus forsteri* from the Lower Cretaceous of New South Wales,
390 Australia. J. Paleontol. 55, 211–217.

391 Kottelat, M., Freyhof, J., 2007. Handbook of European freshwater fishes. Publications Kottelat,
392 Cornol, Switzerland.

393 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: New
394 methods for selecting partitioned models of evolution for molecular and morphological
395 phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. https://doi.org/10.1093/molbev/msw260

396 Lavoué, S., 2015. Testing a time hypothesis in the biogeography of the arowana genus Scleropages

(Osteoglossidae). J. Biogeogr. 42, 2427–2439. https://doi.org/10.1111/jbi.12585

Li, H., He, Y., Jiang, J., Liu, Z., Li, C., 2018. Molecular systematics and phylogenetic analysis of the Asian endemic freshwater sleepers (Gobiiformes: Odontobutidae). Mol. Phylogenet. Evol. 121, 1–11. https://doi.org/10.1016/j.ympev.2017.12.026

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

Marceniuk, A.P., Menezes, N.A., Britto, M.R., 2012. Phylogenetic analysis of the family Ariidae (Ostariophysi: Siluriformes), with a hypothesis on the monophyly and relationships of the genera. Zool. J. Linn. Soc. 165, 534–669. https://doi.org/10.1111/j.1096-3642.2012.00822.x

Marshall, C.R., 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. Am. Nat. 171, 726–742. https://doi.org/10.1086/587523

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbrink, F.T., Friedman, M., Wainwright, P.C., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci. U. S. A. 110, 12738–12743. https://doi.org/10.5061/dryad.d3mb4

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

Phillips, M.J., 2009. Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. Gene 441, 132–140. https://doi.org/10.1016/j.gene.2008.08.017

Purvis, A., Cardillo, M., Grenyer, R., Collen, B., 2005. Correlates of Extinction Risk: Phylogeny, Biology, Threat and Scale, in: Purvis, A., Gittleman, J.L., Brooks, T.M. (Eds.), Phylogeny and

421 Conservation. Cambridge University Press, Cambridge, United Kingdom, pp. 295–316.

422 Pusey, B.J., Kennard, M.J., Perna, C.N., Unmack, P.J., Allsop, Q., Hammer, M.P., 2017. Freshwater
423 fishes of northern Australia. Zootaxa 4253, 1–104. https://doi.org/10.11646/zootaxa.4253.1.1

424 Quintero, I., Wiens, J.J., 2013. Rates of projected climate change dramatically exceed past rates of
425 climatic niche evolution among vertebrate species. Ecol. Lett. 16, 1095–1103.
426 https://doi.org/10.1111/ele.12144

427 R Core Team, 2017. R: A Language and Environment for Statistical Computing.

428 Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C.,
429 Near, T.J., Coll, M., Alfaro, M.E., 2018. An inverse latitudinal gradient in speciation rate for
430 marine fishes. Nature 559, 392–395. https://doi.org/10.1038/s41586-018-0273-1

431 Santini, F., May, M.R., Carnevale, G., Moore, B.R., 2015. Bayesian inference of divergence times and
432 feeding evolution in grey mullets (Mugilidae). bioRxiv 019075. https://doi.org/10.1101/019075

433 Shelley, J.J., Swearer, S.E., Adams, M., Dempster, T., Le Feuvre, M.C., Hammer, M.P., Unmack, P.J.,
434 2018. Cryptic biodiversity in the freshwater fishes of the Kimberley endemism hotspot,
435 northwestern Australia. Mol. Phylogenet. Evol. 127, 843–858.
436 https://doi.org/10.1016/j.ympev.2018.06.032

437 Smith, S.A., O'Meara, B.C., 2012. TreePL: Divergence time estimation using penalized likelihood for
438 large phylogenies. Bioinformatics 28, 2689–2690.
439 https://doi.org/10.1093/bioinformatics/bts492

440 Smith, W.L., Stern, J.H., Girard, M.G., Davis, M.P., 2016. Evolution of venomous cartilaginous and
441 ray-finned fishes. Integr. Comp. Biol. 56, 950–961. https://doi.org/10.1093/icb/icw070

442 Stadler, T., 2010. TreeSim in R - simulating trees under the birth-death model.

443 Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large

444 phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

445 Stevens, M.I., Hicks, B.J., 2009. Mitochondrial DNA reveals monophyly of New Zealand's
446 *Gobiomorphus* (Teleostei: Eleotridae) amongst a morphological complex. Evol. Bioinforma. 11,
447 109–123.

448 Tao, Q., Tamura, K., Kumar, S., 2020. Efficient methods for dating evolutionary divergences, in: Ho,
449 S.Y.W. (Ed.), The Molecular Evolutionary Clock. Springer, Cham, Switzerland, pp. 197–219.

450 Thacker, C.E., Satoh, T.P., Katayama, E., Harrington, R.C., Eytan, R.I., Near, T.J., 2015. Molecular
451 phylogeny of Percomorpha resolves Trichonotus as the sister lineage to Gobioidei (Teleostei:
452 Gobiiformes) and confirms the polyphyly of Trachinoidei. Mol. Phylogenet. Evol. 93, 172–179.
453 https://doi.org/10.1016/j.ympev.2015.08.001

454 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of
455 progressive multiple sequence alignment through sequence weighting, position-specific gap
456 penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

457 Tornabene, L., Chen, Y.J., Pezold, F., 2013. Gobies are deeply divided: phylogenetic evidence from
458 nuclear DNA (Teleostei: Gobioidei: Gobiidae). Syst. Biodivers. 11, 345–361. https://doi.org/Doi
459 10.1080/14772000.2013.818589

460 Unmack, P.J., 2018. Australian Freshwater Fishes [WWW Document]. URL
461 http://peter.unmack.net/biogeog/fish.html (accessed 6.24.20).

462 Unmack, P.J., 2013. Biogeography, in: Humphries, P., Walker, K. (Eds.), Ecology of Australian
463 Freshwater Fishes. CSIRO Publishing, Collingwood, Australia, pp. 25–48.

464 Unmack, P.J., 2001. Biogeography of Australian Freshwater Fishes. J. Biogeogr. 28, 1053–1089.
465 https://doi.org/https://doi.org/10.1046/j.1365-2699.2001.00615.x

466 Veevers, J.J., 1986. Breakup of Australia and Antarctica estimated as mid-Cretaceous (95 ± 5 Ma)

467 from magnetic and seismic data at the continental margin. Earth Planet. Sci. Lett. 77, 91–99.
468 https://doi.org/10.1016/0012-821X(86)90135-4

469 Veevers, J.J., 1984. Phanerozoic Earth History of Australia. Clarendon Press, Oxford, United Kingdom.

470 Xia, X., 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution.
471 Mol. Biol. Evol. 35, 1550–1552. https://doi.org/10.1093/molbev/msy073

472 Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its
473 application. Mol. Phylogenet. Evol. 26, 1–7. https://doi.org/10.1016/S1055-7903(02)00326-3

474 Xing, Y., Zhang, C., Fan, E., Zhao, Y., 2016. Freshwater fishes of China: Species richness, endemism,
475 threatened species and conservation. Divers. Distrib. 22, 358–370.
476 https://doi.org/10.1111/ddi.12399

477 Zhang, J., Wilson, M.V.H., 2017. First complete fossil Scleropages (Osteoglossomorpha). Vertebr.
478 Palasiat. 55, 1–23.

479

480

Table 1. Details of nucleotide sequence data from six mitochondrial and three nuclear genes from Australian freshwater fishes.

Gene	Number of taxa	Untrimmed alignment length (bp)	Trimmed alignment length (bp)
12S	215	1129	624
16S	137	2076	1623
COI	228	1616	654
CYTB	286	1209	1141
ND2	131	1072	1043
ND4	112	1402	1381
18S	141	2038	1760
RAG1	150	3435	2928
RAG2	63	1224	905

485 **Table 2:** Age constraints used in the penalized-likelihood molecular dating analysis.

Node	Clade	Fossil/secondary	Min age (Ma)	Max age (Ma)	Reference
1	Vertebrata = root	Secondary	510	510	Erwin et al. (2011)
2	Gnathostomata	Fossil † *Elegestolepis conica*	426	519	Broughton et al. (2013)
3	Osteichthyes	Fossil † *Guiyu oneiros*	420.7	444.9	Benton et al. (2015)
4	Elopomorpha	Fossil † *Elopsomolos frickhingeri*	149	260	Broughton et al. (2013)
5	Euteleostei	Fossil † *Orthogonikleithrus hoelli*	149	260	Broughton et al. (2013)
6	Teleostei	Fossil † *Anaethalion zapporum*	151.2	252.7	Benton et al. (2015)
7	Clupeocephala	Fossil † *Leptolepides haerteisi*	150.94	235	Benton et al. (2015)
8	Otocephala	Fossil † *Tischlingerichthys viohli*	150.94	228.4	Benton et al. (2015)
9	Ostariophysi	Fossil † *Rubiesichthys gregalis*	126.3	158.3	Benton et al. (2015)
10	Elasmobranchii	Fossil †Archaeobatidae	173	310	Aschliman et al. (2012)
11	*Scleropages*	Fossil † *Scleropages cf. leichardti*	45	89	Lavoué (2015); Unmack (2001)
12	Syngnathidae	Fossil †*"Syngnathus" heckeli*	50	57.3	Near et al. (2012)
13	Gobiiformes	Fossil † *Carlomonnius quasigobius*	47.8	122	Li et al. (2018)
14	Eleotridae + Butidae	Fossil † *Paralates*	33.9	56	Li et al. (2018)
15	Butidae	Fossil †*Lepidocottus aries*	23.03	37.8	Li et al. (2018)
16	Apogonidae	Fossil † *Eosphaeramia pygopterus*	47.8	78.3	Bannikov (2014)
17	G*obiomorphus*	Fossil *Gobiomorphus*	16	37	Stevens and Hicks (2009)
18	Atherinidae	Fossil † *Hemitrichas stapfi*	23	70.5	Campanella et al. (2015)

19	Carangidae	Fossil † *Eastmanalepes primaevus*	49	79.9	Rabosky et al. (2018)
20	Beloniformes	Fossil † *Rhamphexocoetus volans*	50	57.3	Near et al. (2013)
21	Mugilidae	Fossil † *Mugil princeps*	30	84.5	Santini et al. (2015)
22	Tetraodontidae	Fossil † *Archaeotetraodon winterbottomi*	32	97	Betancur-R. et al. (2013)
23	Scatophagidae	Fossil † *Eoscatophagus frontalis*	49	67.2	Rabosky et al. (2018)
24	Maccullochella	Fossil † *Maccullochella cf. macquariensis*	13.1	30	Unmack (2001)
25	Percichthyidae	Fossil † *Macquaria antiquus*	45	70	Unmack (2001)

486

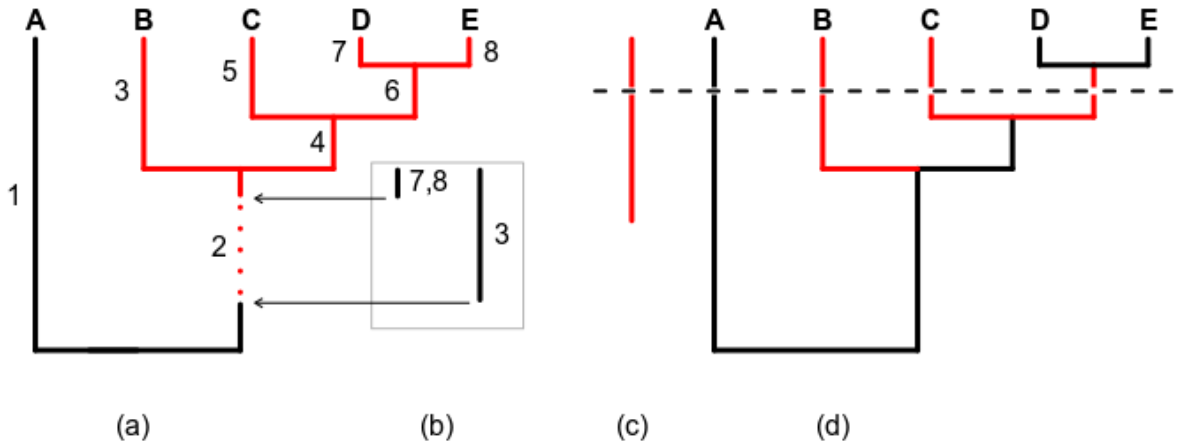
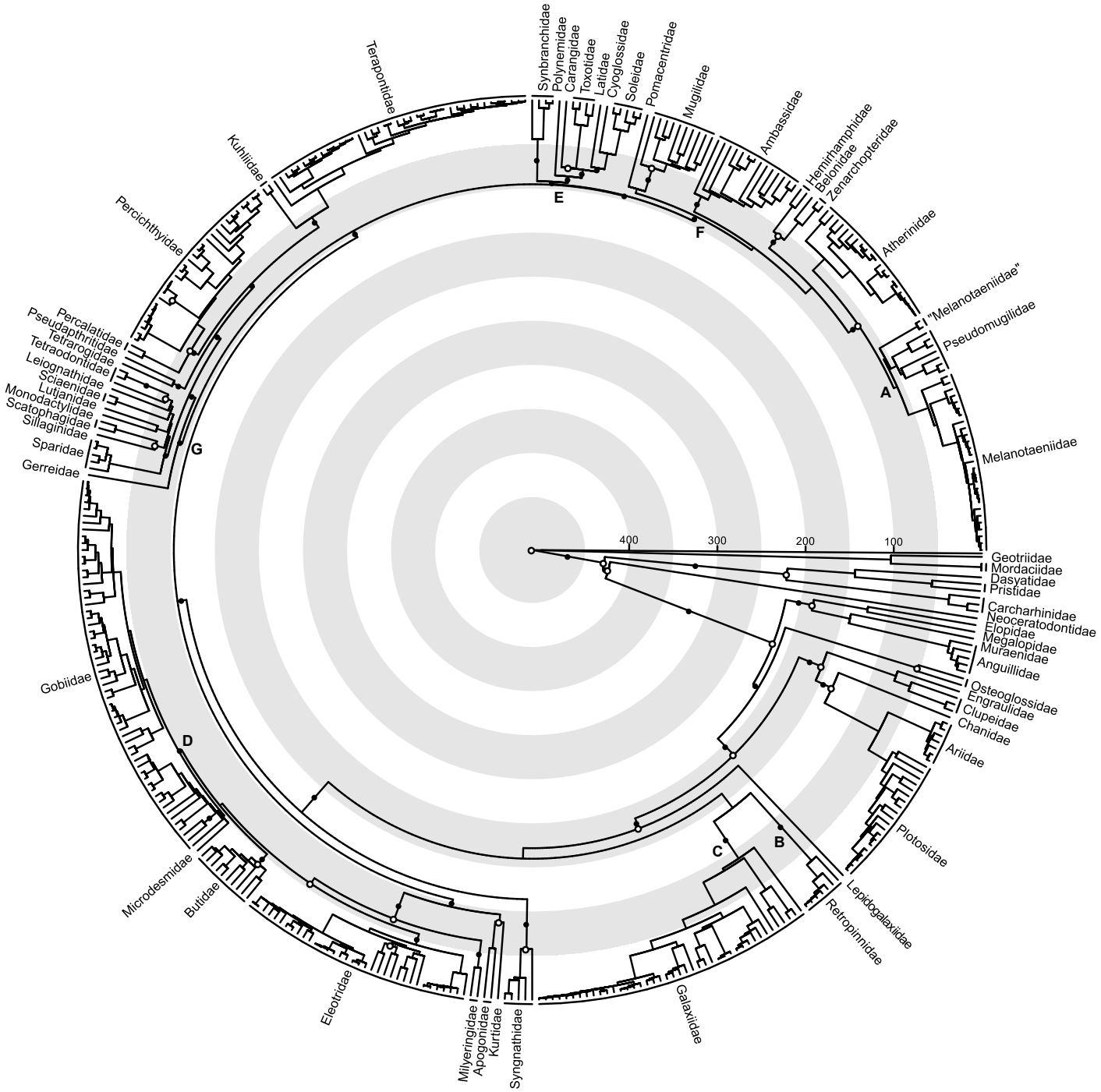


Fig. 1. Grafting procedure for species that lack sequence data. a) Identify branches of the molecular phylogeny to which species can be grafted with a taxonomic constraint, as indicated in red. This includes all internal branches of a clade (branches 3–8), as well as a portion of the stem (branch 2), allowing for a species to be grafted in below the MRCA of species with molecular data. b) Constrain the portion of the stem lineage (branch 2) to which species can be grafted based on the lengths of internal branches. The minimum length of eligible stem corresponds to the length of the shortest internal branch (branches 7 and 8), while the maximum length corresponds to the length of the longest internal branch (branch 3). The length of eligible stem for each grafting iteration was drawn from a uniform distribution between these minimum and maximum stem lengths to provide a maximum age for the speciation event. c) Simulate age of speciation event (indicated by dashed horizontal line) with *corsim* function based on maximum and minimum age constraints. The range of possible ages of a speciation event is indicated by the red line, between a minimum age of zero, and a maximum age falling somewhere in the range described in (b). d) Identify possible grafting points of simulated speciation event, indicated by the intersection of the red phylogeny branches and the horizontal dashed line representing the age of the simulated speciation event, then graft species onto a randomly selected branch.



504

505 **Fig. 2.** Time-tree of all known Australian freshwater fish species and a number of brackish-water

506 relatives. The topology is a maximum-clade-credibility summary of 10,000 dated trees. Black circles

507 represent nodes constrained *a priori* during phylogenetic analysis. White circles represent nodes

508 that had age constraints for molecular dating. Letters indicate nodes that are discussed in the main

509 text: A, Melanotaeniidae (rainbowfishes); B, Retropinnidae (southern smelts); C, Galaxiidae

510 (galaxiids) ; D, Gobiidae (gobies); E, Carangimorpha; F, Ovalentaria; and G, Eupercaria.