Confronting Sources of Systematic Error to Resolve Historically Contentious Relationships: A Case Study Using Gadiform Fishes (Teleostei, Paracanthopterygii, Gadiformes)

ADELA ROA-VARÓN[1,2,3,*], REBECCA B. DIKOW[4], GIORGIO CARNEVALE[5], LUKE TORNABENE[6,7], CAROLE C. BALDWIN[2], CHENHONG LI[8], AND ERIC J. HILTON[3]

[1]*National Systematics Laboratory of the National Oceanic Atmospheric Administration Fisheries Service, 10th St. & Constitution Ave. NW, Washington, DC 20560, USA;* [2]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th St. & Constitution Ave. NW, Washington, DC 20560, USA;* [3]*Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA 23062, USA;* [4]*Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, 10th St. & Constitution Ave. NW, Washington, DC 20560, USA;* [5]*Dipartimento di Scienze della Terra, Università degli Studi di Torino, Via Valperga Caluso, 35, 10125 Torino, Italy;* [6]*School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Seattle, WA 98105, USA;* [7]*Burke Museum of Natural History and Culture, 4300 15th Ave NE, Seattle, WA 98105, USA;* [8]*Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, 999 Hucheng Ring Rd, Pudong, Shanghai, China*
Correspondence to be sent to: National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA; E-mail: Roa-VaronA@si.edu.

Abstract.—Reliable estimation of phylogeny is central to avoid inaccuracy in downstream macroevolutionary inferences. However, limitations exist in the implementation of concatenated and summary coalescent approaches, and Bayesian and full coalescent inference methods may not yet be feasible for computation of phylogeny using complicated models and large data sets. Here, we explored methodological (e.g., optimality criteria, character sampling, model selection) and biological (e.g., heterotachy, branch length heterogeneity) sources of systematic error that can result in biased or incorrect parameter estimates when reconstructing phylogeny by using the gadiform fishes as a model clade. Gadiformes include some of the most economically important fishes in the world (e.g., Cods, Hakes, and Rattails). Despite many attempts, a robust higher-level phylogenetic framework was lacking due to limited character and taxonomic sampling, particularly from several species-poor families that have been recalcitrant to phylogenetic placement. We compiled the first phylogenomic data set, including 14,208 loci (>2.8 M bp) from 58 species representing all recognized gadiform families, to infer a time-calibrated phylogeny for the group. Data were generated with a gene-capture approach targeting coding DNA sequences from single-copy protein-coding genes. Species-tree and concatenated maximum-likelihood (ML) analyses resolved all family-level relationships within Gadiformes. While there were a few differences between topologies produced by the DNA and the amino acid data sets, most of the historically unresolved relationships among gadiform lineages were consistently well resolved with high support in our analyses regardless of the methodological and biological approaches used. However, at deeper levels, we observed inconsistency in branch support estimates between bootstrap and gene and site coefficient factors (gCF, sCF). Despite numerous short internodes, all relationships received unequivocal bootstrap support while gCF and sCF had very little support, reflecting hidden conflict across loci. Most of the gene-tree and species-tree discordance in our study is a result of short divergence times, and consequent lack of informative characters at deep levels, rather than incomplete lineage sorting. We use this phylogeny to establish a new higher-level classification of Gadiformes as a way of clarifying the evolutionary diversification of the order. We recognize 17 families in five suborders: Bregmacerotoidei, Gadoidei, Ranicipitoidei, Merluccioidei, and Macrouroidei (including two subclades). A time-calibrated analysis using 15 fossil taxa suggests that Gadiformes evolved ~79.5 Ma in the late Cretaceous, but that most extant lineages diverged after the Cretaceous–Paleogene (K-Pg) mass extinction (66 Ma). Our results reiterate the importance of examining phylogenomic analyses for evidence of systematic error that can emerge as a result of unsuitable modeling of biological factors and/or methodological issues, even when data sets are large and yield high support for phylogenetic relationships. [Branch length heterogeneity; Codfishes; commercial fish species; Cretaceous-Paleogene (K-Pg); heterotachy; systematic error; target enrichment.]

The rise of phylogenomic analyses over the last two decades has improved phylogenetic reconstruction across a wide range of divergence times and taxa (e.g., Rokas et al. 2003; Delsuc et al. 2006; O'Hara et al. 2014; Gilbert et al. 2015; Branstetter et al. 2017). Gene capture of coding and noncoding genomic loci (e.g., ultra-conserved elements, anchored phylogenetics) in conjunction with next-generation sequencing has become one of the preferred methods of subsampling genomes for phylogenomic studies (e.g., Faircloth et al. 2012; Lemmon et al. 2012; Li et al. 2013; Jiang et al. 2019; Yuan et al. 2016). The large amount of sequence data that now can be generated has addressed the stochastic or sampling error related to limited numbers of phylogenetically informative characters generated by Sanger-based approaches. However, studies at the genomic scale are prone to systematic error (systematic bias) due to the presence of nonphylogenetic signal, which in some cases may lead to strongly supported but incorrect phylogenies (e.g., Kumar et al. 2012; Philippe et al. 2017). As phylogenomic data sets are steadily growing, it is crucial to develop and employ methods to assess and understand the extent to which systematic error affects phylogenetic inference, and to explore ways of mitigating this in empirical studies (e.g., Felsenstein 2004; Philippe et al. 2005; Duchene et al. 2017).

Systematic errors result from an inadequate modeling of methodological factors (e.g., incorrect model selection, poor orthology inference, biased taxon or gene sampling; Delsuc et al. 2005, Philippe et al. 2011) and/or biological factors (e.g., compositional bias, heterotachy, rate heterogeneity, incomplete lineage sorting (ILS);

Jeffroy et al. 2006, Philippe et al. 2017), both of which can lead to biased or incorrect parameter estimates in phylogenomic analyses (e.g., Som 2014; Young and Gillung 2020). These factors can be exacerbated by both speciation events occurring closely in time (short internal branches) and by ancient events (long terminal branches with homoplastic characters) (Philippe et al. 2011). Multiple strategies have been proposed to overcome systematic error resulting from methodological factors, such as improving the quality of primary alignments, increasing taxonomic sampling, using more realistic models of sequence evolution to detect multiple substitutions, and filtering and removing paralogous loci (Hedtke et al. 2006; Lartillot et al. 2007; Philippe et al. 2017). Similarly, systematic error stemming from biological factors can be reduced by the selective removal of outgroups and long-branch taxa (e.g., Huelsenbeck and Hillis 1993), excluding third-codon positions (Sanderson et al. 2000), removing rapidly evolving genes or sites, applying the site-heterogeneous CAT model and amino acids (e.g., Philippe et al. 2005, 2017; Talavera and Vila 2011), using evolutionary models that explicitly model heterotachy (e.g., Crotty et al. 2019), and applying coalescent-based tree-building methods (e.g., Xi et al. 2014; Liu et al. 2015).

Exon capture that is explicitly designed to capture "single-copy" coding sequences across moderate to highly divergent species (Bi et al., 2012, 2013; Li et al. 2013; Bragg et al. 2016) is advantageous in that the "signal-to-noise ratio" in protein sequence alignments is better than in alignments of DNA; may be partitioned by codon position; and can be translated to amino acids to minimize artifacts from base compositional biases (e.g., Davalos et al. 2008). Applications of these advances in genome-scale data set production and phylogenetic inference offer an opportunity to improve our knowledge of the systematic relationships of organisms. In this study, we explore approaches for confronting sources of systematic error using gadiform fishes as a model clade.

Gadiformes include some of the most important commercially harvested fishes in the world (e.g., Alaskan Pollock, Atlantic and Pacific Cods, Blue Whiting, and Hake), accounting for more than a fifth of the world's catch of marine fishes (FAO 2016). However, the confused state of their higher systematics and the large number of poorly known species obscure the necessary framework for many conservation initiatives. This is particularly important considering the threats that commercial and recreational fisheries face from overfishing, physical habitat modification, ocean acidification, nanoparticles, climate change among others, and worrisome reports on the status of some stocks (FAO 2016). For example, Alaskan Pollock (*Gadus chalcogramma*) is considered fully fished in the North Pacific, Atlantic Cod (*Gadus morhua*) is overfished in the Northwest Atlantic and fully to overfished in the northeast Atlantic, and all Hake (*Merluccius* spp.) stocks are considered overfished.

Gadiformes inhabit cool waters circumglobally and are found in all portions of the water column at high latitudes but are primarily restricted to deeper layers of tropical seas. Between 11 and 14 families, 84 genera, and 613 species are currently recognized (e.g., Endo 2002; Roa-Varón et al. 2009; Nelson et al. 2016). The order has been characterized as monophyletic, even though well-defined morphological synapomorphies supporting its monophyly have yet to be established (e.g., Rosen and Patterson 1969; Patterson and Rosen 1989; Murray and Wilson 1999; Endo 2002).

Gadiformes appear to be nested within the Paracanthopterygii; however, the taxonomic content of the superorder is the subject of debate based on both morphological (e.g., Greenwood et al. 1966, Patterson and Rosen 1989, Johnson and Patterson 1993, Davesne et al. 2016) and molecular (e.g., Miya et al. 2001, 2003, 2005, 2007; Grande et al. 2013; Near et al. 2013; Betancur-R et al. 2013; Chen et al. 2014; Alfaro et al. 2018; Hughes et al. 2018) data. There is, however, a general agreement among recent molecular studies for a lineage comprising Zeiformes, Stylephoriformes, and Gadiformes. Among these orders, Gadiformes are the most species rich and phylogenetically complex. The resolution of the interrelationships of Gadiformes, therefore, is critical for better understanding the relationships and evolution of paracanthopterygians and other early divergent acanthomorph fishes.

There have been several attempts to resolve the relationships of Gadiformes based primarily on morphological data (Berg 1940; Svetovidov 1948; Fahay and Markle 1989; Howes 1989; Markle 1989; Nolf and Steurbaut 1989) (S1a-c, Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.5qfttdz2w). These studies produced largely conflicting results, with no consensus regarding the phylogeny and classification of the group (Cohen 1989). Endo (2002) published the most recent and extensive morphology-based analysis of the phylogeny and classification of Gadiformes (S1d, Supplementary Material available on Dryad). Several genetic studies have focused on the systematic relationships of subgroups (i.e., families or the phylogenetic affinities of specific taxa) of Gadiformes (e.g., Paulin 1983; Dunn and Matarese 1984; Carr et al. 1999; Roldan et al. 1999; Quinteiro et al. 2000; Grant and Leslie 2001; Møller et al. 2002; Torii et al. 2003; Teletchea et al. 2006). To date, Roa-Varón et al. (2009) published the only molecular phylogeny focused on Gadiformes based on one nuclear and two mitochondrial DNA (mtDNA) genes from 117 taxa (S1e, Supplementary Material available on Dryad). Other studies, while not directly addressing the phylogeny of Gadiformes, presented results based on substantial yet incomplete taxon sampling. In an analysis of all bony fishes, Betancur-R et al. (2013) used 21 molecular markers for 42 taxa representing nine families of Gadiformes (S1f, Supplementary Material available on Dryad).

Malstrøm et al. (2016) analyzed 567 exons representing 111 genes (71,418 bp) for 27 gadiform species from seven families. The authors reported that Gadiformes lost the major histocompatibility complex II approximately 105 Ma (S1g, Supplementary Material available on Dryad). Hughes et al. (2018) mined data from Malstrøm et al. (2016) to compile 1,105 protein-coding genes for 11 gadiform families (S1h, Supplementary Materials available on Dryad), but no consensus regarding the relationships among these families emerged from a variety of analyses (Hughes et al. 2018, Supplementary Figs. S2–S5). These prior studies have yielded little to no consensus on the family level relationships of Gadiformes. Here, we generated new genomic data for all families of the order using a custom-designed probe set targeting more than 14,000 loci in the nuclear genome. We inferred the species tree based on concatenated and multispecies coalescent methods for data sets differing in the amount of missing data and loci informativeness to investigate biological (e.g., compositional heterogeneity, heterotachy, branch length heterogeneity) and methodological (e.g., optimality criteria, character sampling, model selection) sources of systematic error to infer the most robust and comprehensive backbone phylogeny for Gadiformes to date, estimate divergence times, and highlight remaining challenges.

MATERIALS AND METHODS

Species Sampling and Molecular Techniques

We collected phylogenomic data from single-copy nuclear coding sequence markers across 58 species, including 51 gadiforms representing all currently recognized families and subfamilies (S2a, Supplementary materials available on Dryad), using a targeted sequencing approach and the bioinformatic workflows of Yuan et al. (2016), Song et al. (2017), and Jiang et al. (2019) with minor modifications (lab protocol, bioinformatics pipeline, scripts, and associated files available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.5qfttdz2w). Because the relationships among the earliest diverging acanthomorphs are unresolved (Near et al. 2013; Betancur-R et al. 2013; Chen et al. 2014; Alfaro et al. 2018), we included representatives of all putative acanthomorph orders (Lampridiformes, Percopsiformes, Polymixiiformes, Stylephoriformes, and Zeiformes) and Ophidiiformes (as representative of all remaining percomorphs) as outgroups. We generated genomic data for 43 species. For the remaining 15 species, loci of interest were extracted from published genomic data (Malstrøm et al. 2016) following a modified pipeline for harvesting loci from genomes (Faircloth 2016). Lab protocols and bioinformatics are described briefly below and detailed in S3–S5, Supplementary Materials available on Dryad.

Gene Capture and Probe Design

An initial set of 17,817 conservative single-copy nuclear coding sequence regions (CDS > 90 bp) identified by Song et al. (2017) were obtained by comparing eight well-annotated fish genomes (*Anguilla japonica*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis niloticus*, *Oryzias latipes*, and *Tetraodon nigroviridis*) using Evolmarkers (Li et al. 2013). Common metrics to assess assembly quality are scaffold and contig N50 and L50, which indicate the total number and minimum length, respectively, of all scaffolds or contigs that together account for 50% of the genome. The genomes were highly contiguous: contigs with N50 range: 11 Kb to 2.9 Mbp and L50 range 88–19 kb; scaffold with N50 range: 734 Kb–38.8 Mbp and L50 range 1–102 (S3, Table 1 Supplementary Materials available on Dryad). A set including 14,217 CDS (>120 bp) was generated and used to design baits based on the sequences of *G. morhua* (no degenerated baits were included). Bait sequences of 120 bp were tiled to obtain 2× coverage of each targeted locus (60 bp overlap between baits). Biotinylated RNA probes of bait sequences were synthesized by Arbor Bioscience (formerly MYcroarray, Ann Arbor, MI, USA). Illumina sequencing libraries (Meyer and Kircher 2010) were prepared for each sample following the "with-bead" method of Li et al. (2013).

Data Assembly, Orthology Testing, and Alignment

Data were assembled using a pipeline modified from Li et al. (2013, S5, Supplemental Materials available on Dryad). We removed adapter sequences, low-quality reads, and PCR duplicates, and parsed reads into files based on similarity to the target loci. Parsed reads were assembled into contigs and putatively orthologous genes were identified by matching contigs to the *G. morhua* genome using BLAST + (v. 2.4.0; Camacho et al. 2009). Loci were translated into amino acids (AA), and both DNA and AA loci were aligned in MAFFT (v.7.221; Katoh and Standley 2013) and concatenated with FASconCAT-G (Kück and Longo 2014).

Concatenated Phylogenetic Analyses

In this study, we used numerous approaches to tackle potential sources of systematic error from inappropriate modeling of biological and/or methodological factors (Fig. 1). Both of these sources of error can lead to biased or incorrect parameter estimates in phylogenomic analyses yet result in statistically highly supported, phylogenomic trees.

Missing Data and Loci Informativeness.—We built four matrices to account for missing data and loci informativeness. The first matrix is the original data set (DNA sequences, 58 taxa, hereafter as DNA-58T). The remaining three matrices were built with MARE v0.1.2-rc
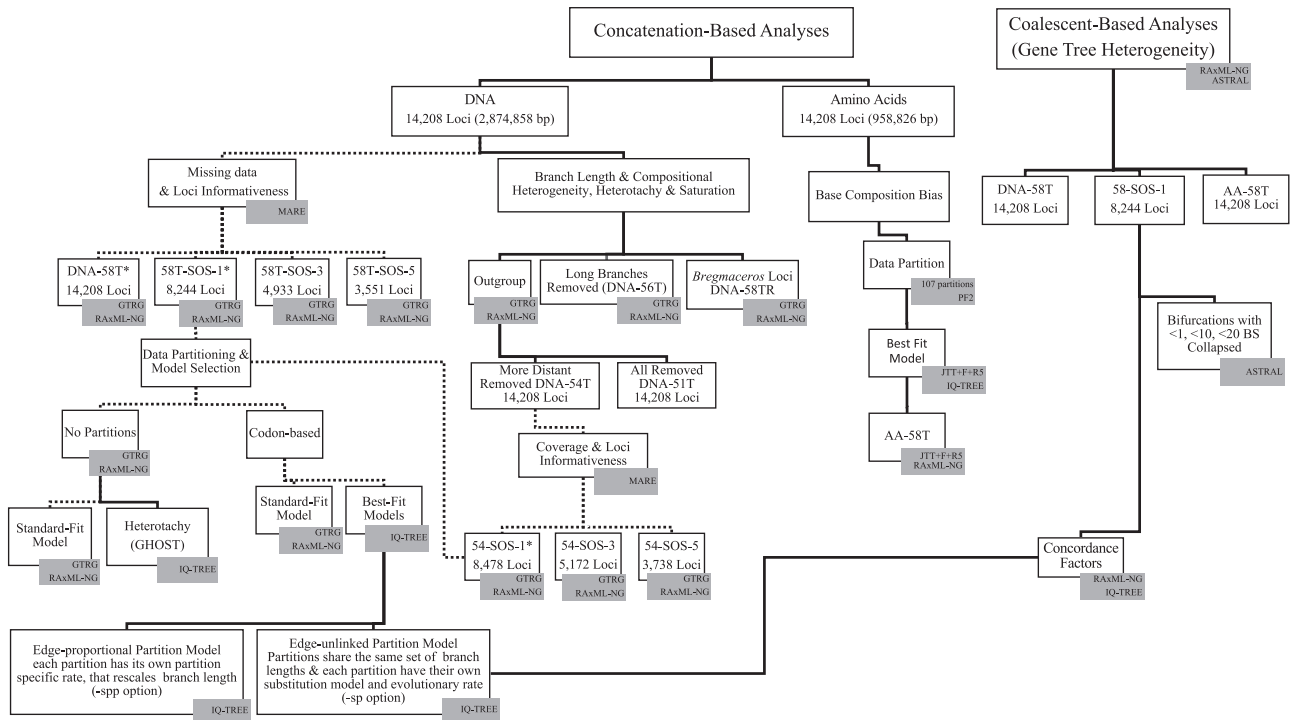
FIGURE 1. Workflow of the data analyses exploring biological (solid) and methodological (dotted) sources of systematic error that can result in biased or incorrect parameter estimates in our phylogenomic analyses. *degen applied.

(**MA**trix **RE**duction, Meyer et al. 2011) by adjusting the α value (α1 more relaxed and α5 stricter), 58T-SOS-1: α = 1; 58T-SOS-3: α = 3; and 58T-SOS-5: α = 5 to search for the highest information content while excluding as few loci as possible (Fig. 1). MARE combines potential signal (i.e., informativeness) of loci with its data coverage to generate **S**elected **O**ptimal **S**ubsets (SOS) of taxa and loci. The relative information content of each locus was calculated as the average value over all taxa including missing taxa, whereas the total average information content of the supermatrix was calculated as the sum of the relative information content of all loci in relation to the number of taxa. ML analyses were conducted using RAxML-NG v0.8.1 BETA (Kozlov 2018) under the GTR+G model. ML tree searches were conducted using 20 distinct random starting trees (10 random + 10 parsimony). Support for nodes in the ML tree was assessed with 100 slow bootstrap replicates for each of the four matrices. We define the support of the species trees as follows: strongly supported nodes have a BS >90, moderately supported nodes are between 90 and 70 BS, and poorly supported nodes are those with a BS <70. All trees were visualized with FigTree v1.4.2 (Rambaut 2009).

Data Partitioning and Model Selection.—The following two strategies were applied to the optimized data sets (58T-SOS-1) to evaluate the impact of data partitioning and model selection: (i) unpartitioned versus partitioned by codon position with the GTR model and gamma distribution assuming constant substitution rates during evolution (Fig. 1). ML tree searches were conducted in

RAxML-NG v0.8.1 using the same parameters described above; and (ii) partitioning by codon position was used as input for model selection in IQ-TREE v1.6 (Kalyaanamoorthy et al. 2017). The best-fitting model for each codon position was selected based upon Akaike Information Criteria (Akaike 1974).

Branch Length and Compositional Heterogeneity.— Outgroup taxa may attract fast-evolving species of the ingroup through long branch attraction (LBA), thereby forcing the rapidly evolving taxa to be recovered as emerging too deeply in the tree (Philippe et al. 2011; Whelan et al. 2015). This effect was assessed by: (i) removing distant outgroups (Lampridiformes and the Ophidiiformes; DNA-54T) followed by applying the same matrix reduction strategy applied to 58T data set with MARE v0.1.2-rc (**MA**trix **RE**duction, Meyer et al. 2011); (ii) removing all outgroups (DNA-51T); (iii) building a reduced matrix including only the 146 loci found in *Bregmaceros* spp. (DNA-58T-R); and (iv) removing long-branch taxa (*Bregmaceros* spp. DNA-56T). ML tree searches were conducted in RAxML-NG v0.8.1 using the same settings as described above (Fig. 1).

Heterotachy.—Because heterotachy is one of the primary sources of model misspecification (Kolaczkowski and Thornton 2004; Crotty et al. 2017), we addressed the rate of variation across sites and lineages using the following approaches: (i) edge-proportional partition model with proportional branch lengths, but each

partition has its own partition specific rate (-spp option); (ii) edge-unlinked partition model—each partition has its own set of branch lengths (-sp option); and (iii) General Heterogeneous evolution On a Single Topology (GHOST) model using GTR+FO+H4 in IQ-TREE (Nguyen et al. 2015; Crotty et al. 2019). The first two approaches were partitioned a priori by codon position, and then separate branch length and/or model parameters were inferred for each partition. However, because the partition itself is a possible source of model misspecification, the GHOST model offers the advantage of inferring heterotachous evolutionary processes without specifying partitions a priori and minimizing model assumptions (Fig. 1).

Saturation and Base Composition Bias.—Rapidly evolving synonymous nucleotide changes are mostly uninformative at deep levels and often cause analytical problems such as LBA and nucleotide compositional heterogeneity, among others (e.g., Song et al. 2010, Zwick et al. 2012). To examine the extent of saturation for the nucleotide data sets (DNA-54T, SOS-1) the degen v1.4 Perl script (Zwick et al. 2012) was used to exclude synonymous signal that can contribute to saturation. ML tree searches were conducted in RAxML-NG v0.8.1 under the GTR+G model and using the same settings as described above (Fig. 1). Because translation to amino acids can eliminate potentially problematic synonymous nucleotide changes a tree based on amino acids for AA-58T data set was inferred by estimating the number of partitions with PF2 (Lanfaear 2016) and the following parameters: "rcluster-max" to 1000 and "rcluster-percent" to 10 for the relaxed clustering algorithm. Model selection was performed in IQ-TREE followed by ML analyses using RAxML-NG v0.8.1 following settings as above (Fig. 1).

Coalescent Species Tree Analyses

Gene Tree Heterogeneity.—Species trees were inferred for the DNA-58T, 58T-SOS-1, and the AA-58T data sets using the coalescent summary method in ASTRAL-II v5.6.3 (Mirarab and Warnow 2015). The individual input gene trees were first estimated by using RAxML-NG v0.8.1 as described above. We used the best-fitting ML gene trees as an input and conducted gene/site resampling on the bootstrap replicates (Fig. 1). We collapsed to polytomies bifurcations with very low bootstrap support (<1 BS, <10 BS, <20 BS, <30 BS) within each gene tree to discern if discordant topologies are due to low informativeness of the data and to minimize the impact of estimation error (Zhang et al. 2017, 2018; Sayyari and Mirarab 2018).

Gene Tree and Species Tree Support

To assess the degree of discordance between gene trees and the species tree in our data, we applied the following approaches. First, because bootstrap or posterior probabilities do not provide a comprehensive measure of the underlying agreement or disagreement among sites and genes for any topology, we estimated the gene concordance factor (gCF) and the site concordance factor (sCF) for each 58T-SOS data set using IQ-TREE v1.6.10 (Kalyaanamoorthy et al. 2017). For every branch of the tree, gCF is the percentage of "decisive" gene trees containing that branch and accounting for variable taxon coverage among gene trees and sCF is the percentage of "decisive" sites supporting a branch in the tree. sCF was calculated based on the concatenated alignment and gCF based on the individual gene trees (Fig. 1). Both gCF and sCF complement bootstrap branch support values by providing a full description of underlying disagreement among loci and sites (Minh et al. 2020). Bootstrap, gCF, and sCF statistics for each branch were plotted in R (R Core Team 2013). Third, to test if the discordance among gene trees and sites come from ILS and/or poorly estimated gene trees, we calculated the probability that the data can reject equal frequencies for genes (gEFp) and for sites (sEFp) using a chi-square test (assuming that sites are not linked within single loci) to see if they are significantly different. If they were not significant, discordance among gene trees and among sites was likely due to ILS (script available here: http://www.robertlanfear.com/blog/files/concordance_factors.html). Second, to test the significance of the differences between phylogenies derived from our analyses, we employed the approximately unbiased (AU) test assuming independence of the sites (Shimodaira (2002) included in IQ-TREE v1.6.10 (Kalyaanamoorthy et al. 2017)). We may reject the possibility that a tree is the most likely tree among all candidates when the AU *p*-value <0.05.

Divergence Time Analyses

Divergence-time estimation was conducted using the program RelTime (Tamura et al. 2012), a fast, *ad hoc* approach for large molecular phylogenies implemented in MEGA v. 7.0.20 (Kumar et al. 2012, 2016b). The basis of the RelTime approach is a relative rate framework that combines comparisons of evolutionary rates in sister lineages with the principle of minimum rate change between evolutionary lineages and their respective descendants (Sanderson 1997; Tamura et al. 2012, 2018). This approach relaxes the molecular-clock assumption and smooths the local rates in a tree (Kumar and Hedges 2016a). Other ML-based approaches (e.g., TreePL) are computationally efficient, requiring only a fraction of the time and resources demanded by Bayesian approaches and generate accurate time estimates (Tamura et al. 2012, 2018). However, they are limited by the lack of reliable calculation of the uncertainty surrounding divergence times, which are represented by confidence intervals (CIs) implemented in RelTime (Tao et al. 2019). In total, we used 15 calibration points across the best ML tree hypothesis using minimum and/or maximum boundaries under the GTR+G model and using all sites.

Justification and description of the fossils used for the calibration can be found in S6, Supplementary Material available on Dryad.

<center>RESULTS AND DISCUSSION</center>

Gene Capture Data Collection

We analyzed sequences for 14,208 loci from 58 taxa: 51 species representing all families and subfamilies of Gadiformes and seven outgroups representing all earliest diverging acanthomorph orders and other percomorphs. On average, 36.5 % of loci were recovered (95% CI 5.5–60.8), locus length was 193.7 bp (95% CI 67.4–216.5), there were 285,599 parsimony-informative sites (95% CI 18,557–390,582), and there were 31.4% variable sites (95% CI 25.7–38.4). The number of sequences generated, accession numbers, number of raw reads, contigs and values to assess the quality of the assembled loci, can be found in S2a, Supplementary Materials available on Dryad.

The percentage of capture efficiency by family ranged from 12.3% to 76.0%, except for Bregmacerotidae, which registered the lowest capture efficiency of 4.8%. The composition and properties of the primary matrices are summarized in S2b, Supplementary Materials available on Dryad. Data available from NCBI (http://www.ncbi.nlm.nih.gov/bioproject/609091) and TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S27320).

Phylogenomic Analyses

Phylogenetic hypotheses were inferred using ML and coalescent summary methods based on 12 different data sets to assess how methodological and biological sources of systematic error may affect phylogenetic inference (S2b, Supplementary Materials available on Dryad).

Missing Data and Loci Informativeness.—Four distinct supermatrices were assembled. The complete unreduced matrix (DNA-58T) was based on all filtered alignments comprising at least four sequences. Three "MARE" matrices were obtained by removing relatively uninformative and low-coverage loci from the DNA-58T matrix by adjusting the alpha variable: α1: 58T-SOS-1 (8,244 loci; 1,784,739 sites), α3: 58T-SOS-3 (4,933 loci; 1,052,865 sites), and α5: 58T-SOS-5 (3,551 loci; 715,188 sites), while keeping the number of taxa fixed (deleting organisms was allowed, but none was excluded; S7, Supplementary Materials available on Dryad). The ML phylogenetic trees were inferred using a single GTR+G model across the alignment for all four matrices (Fig. 2a–d).

A fully congruent topology was resolved for DNA-58T and 58T-SOS-1 (97% and 99% of the nodes with more than 95% BS values respectively; Fig. 2a,b). Relationships among the earliest diverging taxa are highly supported, as are all clades corresponding to families and subfamilies. The tree was rooted on Ophidiiformes, and Lampridiformes were recovered as the sister group to (Polymixiiformes, (Percopsiformes, (Zeiformes, (Stylephoriformes, Gadiformes)))). Among the taxa sampled *Stylephorus chordatus* was recovered as the closest extant relative of Gadiformes. Within Gadiformes, six main lineages were found based on the concatenated ML topology (2.5–24.4 gCF, 33.2–41.6 sCF, and 99–100% BS) representing 17 families (45.6–83.0 gCF; except Lotidae with 11.2; 51.4–89.6 sCF; and 100% BS). Bregmacerotidae was resolved on a long branch as the sister group of all other gadiforms, which were divided into two well-supported clades (Fig. 2a–d). The first clade, Gadoidei, includes (Phycidae, (Gaidropsaridae, (Lotidae, Gadidae))). Within the second clade, the monotypic Ranicipitidae and Merlucciidae were successive sister clades to the remaining Gadiformes, which are represented here by Macrouroidei with two subclades. The first of these subclades comprises (Trachyrincidae, (Euclichthyidae, (Melanonidae, Muraenolepididae))) and the second (Bathygadidae, (Macruronidae, Lyconidae)), (Steindachneriidae, Macrouridae), with Moridae as its sister group. Both ML topologies (58T-SOS-3 and 58T-SOS-5) were largely consistent with those obtained from the ML analysis of the DNA-58T and 58T-SOS-1 and differed only in one clade in the second lineage that registered the lowest gCF, sCF, and bootstrap support values: (Muraenolepididae, Merlucciidae) was found to be sister to the remaining gadiforms in 58T-SOS-3 (2.3/35.1/53, respectively; Fig. 2c; S2d, Supplementary Materials available on Dryad) or nested within a clade including (Euclichthyidae, (Trachyrincidae, (Melanonidae, (Muraenolepididae, Merlucciidae)))) in the 58T-SOS-5 matrix (0.3/37.2/21, respectively; Fig. 2d; S2d, Supplementary Materials available on Dryad). The drop in the concordance factors and bootstrap support values in the 58T-SOS-3 and 58T-SOS-5 matrices, mainly in the second lineage are possibly due to the decrease of information content. This corroborates the importance of exploring different α values when running MARE to identify the matrix with the highest information content while excluding as few loci as possible. The preferred topology among the four matrices in terms of missing data and loci informativeness is the topology resulting from the 58T-SOS-1. The matrix is about 58% of the size of the original, the information content increased from 0.276 to 0.402, and was supported by the combination of concordance factors and bootstrap support values (S2d, Supplementary Materials available on Dryad).

Selected Removal of Outgroup and Long-Branch Taxa.—A topology that is fully congruent with DNA-58T was found after: (i) removing the most distant outgroup taxa (DNA-54T; Fig. 3a,b); (ii) applying a similar strategy with MARE (54T-SOS-1; S8, Supplementary Materials available on Dryad); (iii) removing the long branches (two *Bregmaceros* spp.- DNA-56T; Fig. 3d); and (iv) removing all the outgroups (DNA-51T; Fig. 3c). Nodes with less than 95% BS increased from 2.8% to 3.7%
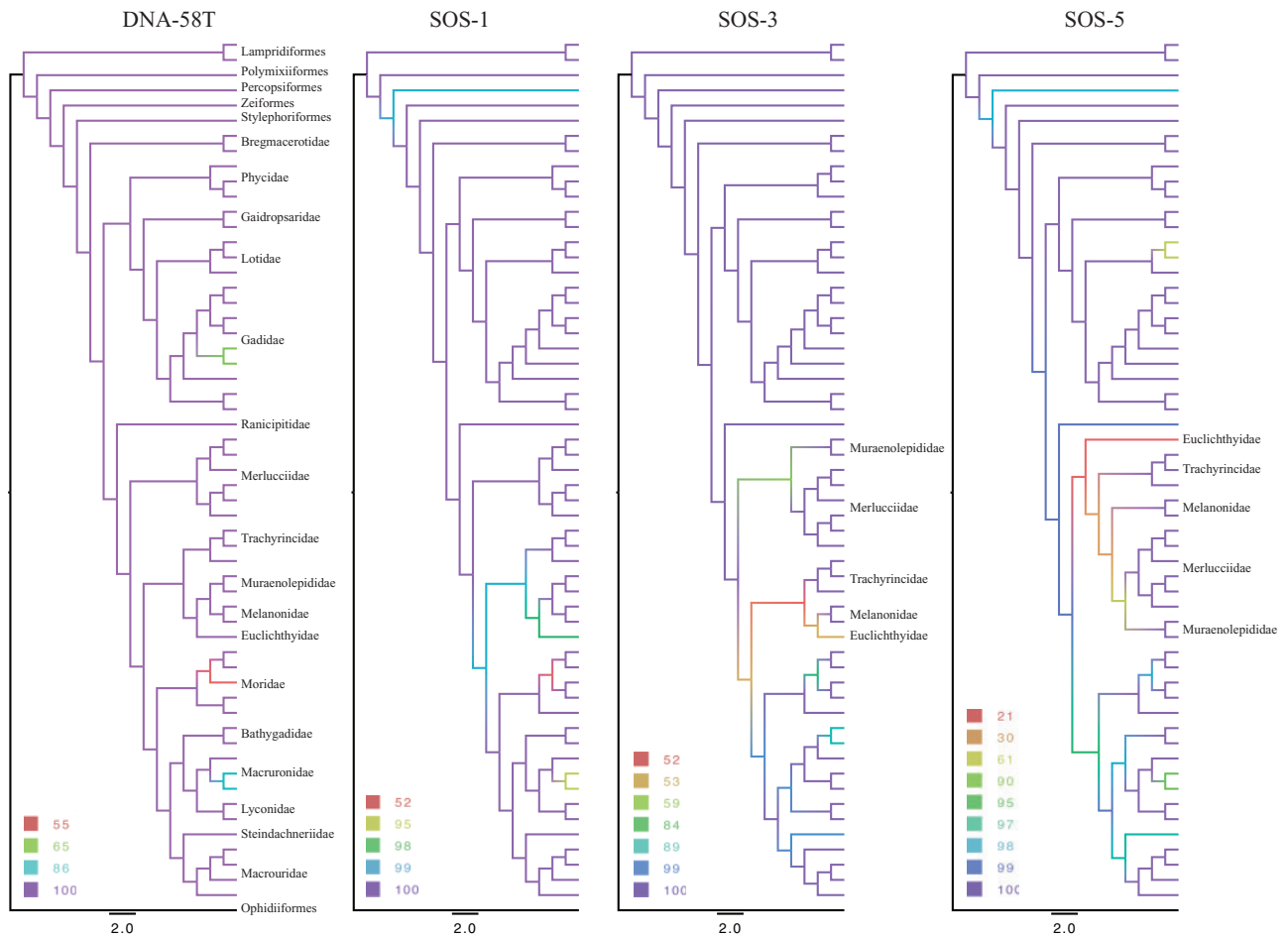
FIGURE 2.   Data coverage and loci information content. Phylogeny of 58 gadiform taxa inferred with maximum likelihood inference using concatenated data sets: DNA-58T (14,208 loci/2.874,858 bp); SOS-1 (8,244 loci/1,784,738 bp); SOS-3 (4,933 loci/1,052,865 bp); SOS-5 (3,551 loci/715,188 bp) and a GTRGAMMA global evolutionary model. Bootstrap support values are color coded.

and 4.7% in the last two approaches, respectively. Both increase and reduction in branch support have been reported after removal of taxa, suggesting other sources for systematic error such as improper modeling of biological phenomena (e.g., compositional bias, rate heterogeneity, heterotachy) and/or methodological issues (e.g., poor model selection, biased taxon or gene sampling, poor orthology inference) (e.g., Philippe et al. 2011; Whelan et al. 2015). We attempted to break up the long branches by increasing taxon sampling within Bregmacerotidae through the inclusion of loci from *B. cantori* that were harvested from Malstrøm et al. (2016, Supplementary Figure 1; a; long branch leading to Bregmacerotidae was also observed in their study). However, very few loci were captured for *B. cantori* (0.2% capture efficiency representing only 25 of our 14,208 loci). We did not include the species in the analyses because it would introduce more noise than phylogenetic signal and increase computing time due to the amount of missing data.

The topology recovered through analysis of the reduced data set, which included only those loci present in Bregmacerotidae (146 loci recovered; 83,025

bp;18,557 parsimony informative sites), changed the topology drastically (S9, Supplementary Materials available on Dryad). For example, with the reduced data set, Melanonidae, Moridae, Euclichthyidae, and Trachyrincidae form a series of successive sister taxa to all the remaining gadiforms, and Merlucciidae and Ranicipitidae are nested within the Gadoidei. Additionally, bootstrap support dropped at many nodes, suggesting significant loss of phylogenetic signal. This is not unexpected because the most consistently recovered loci across taxa are those that are more conserved and therefore less informative.

Codon-based Partitioning, Substitution and Heterotachy Models.—As a result of the optimization process, the 58T-SOS-1 and 54T-SOS-1 matrices were used in the downstream phylogenetic analyses to evaluate the impact of data partitioning and model selection. For each of the following strategies, we found a fully congruent topology: (a) unpartitioned with GTR+G model; (b) partitioning scheme by codon position with the GTR+G model; and (c) partitioning scheme with substitution
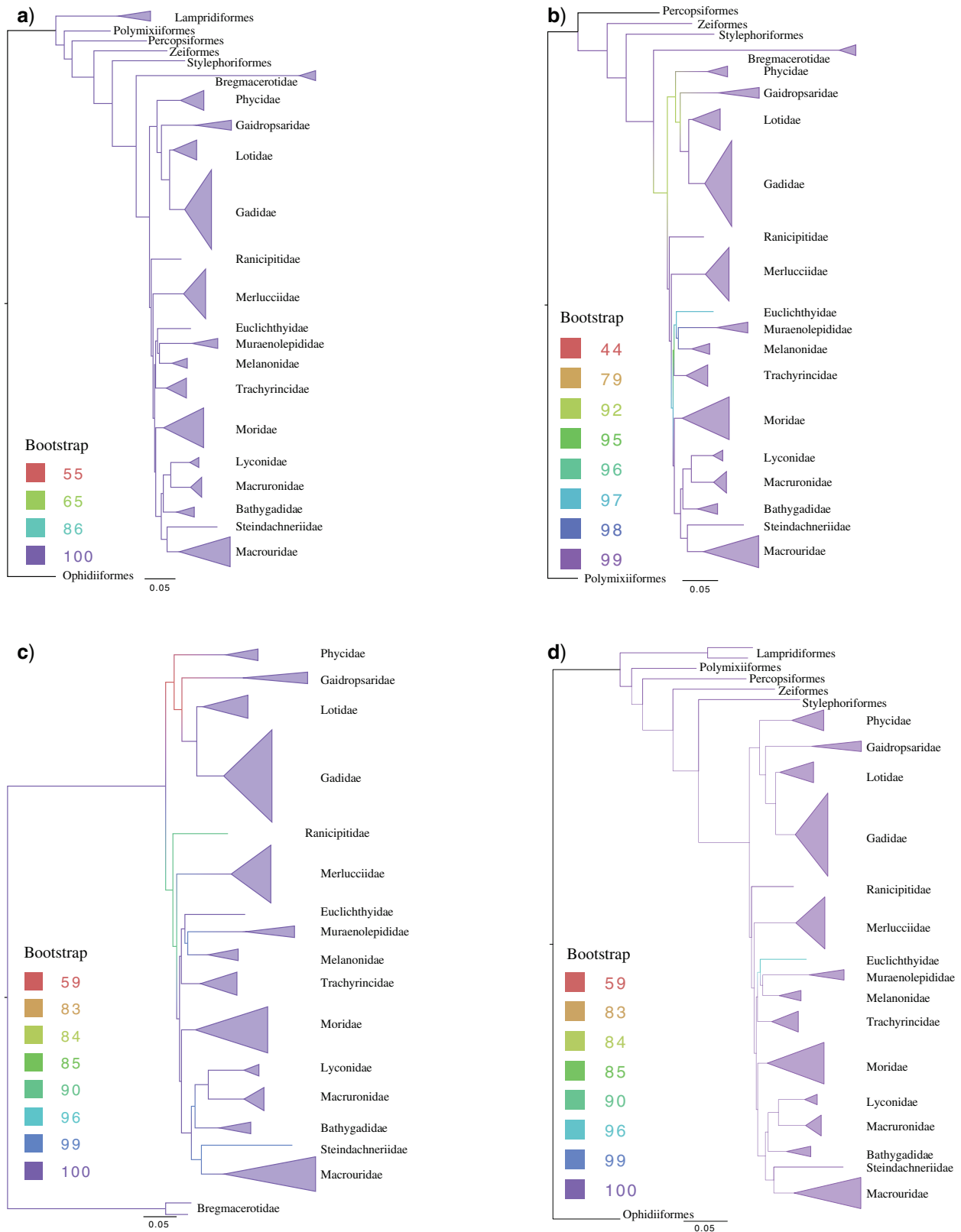
FIGURE 3. Branch length and compositional heterogeneity. Maximum likelihood inference using concatenated data sets: a) DNA-58T (14,208 loci/2,874,858 bp); b) DNA-54T-SOS-1 (removing the most distant outgroups); c) DNA-51T (all outgroups removed); d) DNA-56T (matrix where the long-branch taxa were removed) Bootstrap support values are color coded.

model and rate of evolution by codon position estimated with IQ-TREE v1.6 (P1: GTR+F+R4; P2: GTR+F+R5, P3: GTR+F+R7) (S10, Supplementary Materials available on Dryad). Our results suggest that partitioning by codon position and choosing the best-fitting model by codon position had little impact on the resulting tree topology. The models estimated by IQ-TREE for each codon position were the most parameter rich and very similar to GTR+G (58T-SOS-1−P1: GTR+F+R4; P2: GTR+F+R5; P3: GTR+F+R7 and 54T-SOS-1−P1: GTR+R4; P2 and P3: GTR+R6). This could result in consistent topologies with marginal differences in bootstrap support values.

Additional analyses to accommodate heterotachy resulted in the same topology and LBA artifact for the 58T-SOS-1 independent of the approach used (edge-proportional, edge-unlinked, and GHOST models) (S11a-c, Supplementary Materials available on Dryad). In contrast, the phylogenetic placement of three families changed depending on the model used for the 54T-SOS-1 matrix (S11d–f, Supplementary Materials available on Dryad). The edge-unlinked partition model suggested Muraenolepididae + Merlucciidae (sister to the rest of Gadiformes except Bregmacerotidae), Trachyrincidae within Macrouroidei—clade 2), and Ranicipitidae within Gadoidei; edge-proportional partition model recovered Muraenolepididae (sister to the rest of Gadiformes except Bregmacerotidae); and the GHOST model, included Ranicipitidae within Gadoidei (S11d–f, Supplementary Materials available on Dryad). The different topologies between the data sets suggest that the number of taxa included in the outgroup is affecting the performance of the models addressing the rate of variation across sites and lineages and/or that those models of molecular evolution are insufficient to deal with all the phenomena present in these genome-scale matrices.

Saturation and Base Composition Bias.—The topology resulting from the DNA-54T + degen after removing synonymous signal was exactly the same as DNA-54T and DNA-54T-SOS-1. In this analysis, no saturation was detected suggesting that the concatenated data were suitable for phylogenetic analysis (S12, Supplementary Materials available on Dryad). Analyses using the concatenated amino acid alignment for addressing problematic synonymous nucleotide changes consisted of 956,264 sites and 107 subsets resulting from Partition Finder 2 (PF2, Lanfear 2016) using the JTT+F+R5 model. ML analysis generated phylogenetic hypotheses in which the majority of nodes were well supported, with some exceptions (S13, Supplementary Materials available on Dryad). We observed topological conflict between the amino acid (AA-58T) and DNA ML trees (58T-SOS-1) at only three nodes: (i) the placement of Merlucciidae as the sister group of all gadiforms except Bregmacerotidae with high support (99% BS); (ii) Ranicipitidae within the gadoids clade with moderately high support (90% BS); and (iii) Euclichthyidae sister to ((Melanonidae, Muraenolepididae), Trachyrincidae)

with relatively low support (77% BS) in the AA-58T data set. Recovering Bregmacerotidae, Merlucciidae and Muraenolepididae diverging so deeply in the amino acid-based topology suggests low amino acid variation and/or that LBA could be biasing the topology. However, using amino acids instead of nucleotides has been suggested to alleviate LBA (Inagaki et al. 2004; Mathews et al. 2010; Talavera and Vila 2011). Therefore, we favor the explanation of low phylogenetic signal in the amino acid data set.

Gene Tree Heterogeneity.—The species trees inferred using ASTRAL (DNA-58T and 58T-SOS-1) resulted in the same topology and recovered monophyly of all gadiform families (except Lotidae) as in the concatenated analyses DNA-58T, 58T-SOS-1, DNA-54T, and the 54T-SOS-1 with high support (S14, Supplementary Materials available on Dryad). The nonmonophyly of Lotidae had low support compared with the ML tree estimated for the 58T-SOS-1 (gCF = 11.16%; sCF = 45.67% and 100% BS). After collapsing branches with low support (<1% and <10% support) within each gene tree the monophyly of Lotidae still was not supported. However, when branches with <20% and <30% support were collapsed the monophyly of Lotidae was recovered with high support suggesting an increase in accuracy by reducing noise (S15, Supplementary Materials available on Dryad). In these trees, the families Merlucciidae and Muraenolepididae were recovered sister to the Gadoidei and Macrouroidei lineages, as in the topologies recovered in the amino acid-based (S13, Supplementary Materials available on Dryad), and applying the edge-unlinked partition model from the 54T-SOS-1 data set (S11, Supplementary Materials available on Dryad). The topology resulting from the amino acids data set (AA-58T) included one clade of species, representing different families, that correspond to the sequences harvested from Malstrøm et al. (2016), suggesting issues with the reading frame of these data (S16, Supplementary Materials available on Dryad). Additionally, it is likely that the individual amino acid trees contain insufficient phylogenetic information as a result of their short length.

Gene Tree and Species Tree Support

Concordance factors.—Our estimates of sCF and gCF were correlated across the species tree of gadiforms, but we noted that both of these measures fell well below those of bootstrap support measures (S2d, S17; Supplementary Materials available on Dryad). The 58T-SOS-1 concatenated ML tree had 95% of the bootstrap values in the tree >99%, and the gCF and sCF values range from 2.5% to 83.0% (median: 22.0) and 18.5% to 94.6% (median: 50.0%) divergence, respectively. The gCF and sCF were higher and consistent across matrices at shallow nodes supporting the monophyly of all families, but discordance at deeper nodes (e.g., suborders) increase. Short internodes connecting long branches at deep levels are suggestive of ancient divergences

that gave rise to speciation events over relatively short periods of time (S2d, S17; Supplementary Materials available on Dryad).

The phylogenetic placement of Bregmacerotidae did not change in any of our analyses regardless of the methodological and biological approaches used to break the long branch (Figs. 2–5, S9–S16). For example, if this branch was being artificially attracted toward outgroups, a different placement for the family would be expected when different outgroups are included. Applying models of sequence evolution accounting for base compositional heterogeneity and heterotachy in partitioned and unpartitioned data sets recovered identical relationships, indicating that model choice did not bias results (S10, Supplementary Materials available on Dryad). Overall, findings presented here support Bregmacerotidae as sister to all remaining gadiform fishes (23.5 gCF; 49.9 sCF; 100 BS) and provide strong evidence that LBA artifacts are not biasing the placement of the family. However, support values can be explained by both good support and alternatively due to systematic error consistently affecting all loci. Additionally, the low capture efficiency (4.8%) for the family can exacerbate systematic error leading to LBA artifacts by reducing the ability to detect multiple substitutions, along these long branches. This limited our conclusions and future investigations including additional taxa will be needed to more fully assess its placement.

The historically unresolved relationships among gadiform lineages were consistently well resolved in our analyses regardless of the methodological and biological approaches used. Only three analyses rendered different placements for Merlucciidae, Muraenolepididae, and Ranicipitidae: (i) coverage and loci informativeness analysis, in which Merlucciidae and Muraenolepididae were sister taxa in 58T-SOS-3 and 58T-SOS-5, but this hypothesis was poorly supported (5.5/2.9 gCF; 34.3/34.2 sCF; 59/30 BS, respectively; Fig. 2c,d; S2d, Supplementary Materials available on Dryad); (ii) amino acid-based analysis, in which Merlucciidae was recovered as the sister group of all gadiforms except Bregmacerotidae (S13, Supplementary Material available on Dryad); and (iii) heterotachy analysis in 54T-SOS-1, in which Muraenolepididae and Merlucciidae were recovered sister to the Gadoidei and Macrouroidei lineages (edge-proportional partition); Muraenolepididae sister to all gadiforms except Bregmacerotidae (edge-unlinked partition model); and Ranicipitidae was the earliest branch within Gadoidei according to GHOST (S11f, Supplementary Materials available on Dryad). The first two could be a result of loss of phylogenetically informative characters, while the third one suggests that theoretical work is needed to explore the relationship between outgroup rooting and LBA, especially with regard to how different outgroup taxon sampling strategies affect the probability of LBA and other systematic errors. Finally, most of the gene-to-species-tree discordance were not described by ILS, where 75.0% of nodes show significant differences between the two discordant topologies (S2e,

Supplementary Material available on Dryad). Similarly, sCF shows 36.0% of nodes have significance difference among discordance sites. Hence, most of the gene-tree and species-tree discordance in our data set is likely due to short times between divergences and consequent lack of informative characters at deep levels rather than a result of ILS.

Gene Tree and Species Tree Incongruence

Approximately unbiased test.—AU *p*-values represent how strongly competing topological hypotheses conflict with a preferred hypothesis, with a canonical value of 0.05 reflecting strong support for rejection. Alternative hypotheses to the most frequent best-fit tree were all rejected by AU test and therefore the phylogeny presented in Fig. 4 remains unfalsified (S18; Supplementary Materials available on Dryad).

Phylogeny and Evolution of Gadiformes (Figs. 4 and 5)

Recent molecular phylogenetic studies have produced conflicting hypotheses of relationships among early diverging acanthomorphs (e.g., Lampridiformes, Polymixiiformes, Percopsiformes, Zeiiformes, Stylephoriformes, and Gadiformes). While resolving the relationships among acanthomorphs is beyond the scope of this study, as noted above, our results support a clade (Fig. 4) that includes the orders (Lampridiformes, (Polymixiiformes, (Percopsiformes, (Zeiformes, (Stylephoriformes, Gadiformes))))) and corroborate the placement of *Stylephorus chordatus* as the closest extant relative of Gadiformes (Miya et al. 2007; Chen et al. 2014; Alfaro et al. 2018).

The ML phylogeny calibrated with 15 fossils suggests that Gadiformes likely originated during the Late Cretaceous, at ~79.5 Ma (95% CI 98.0–61.0 Ma, Fig. 5; S2c), more recent than predicted by Malstrøm et al. (2016). Initial divergence of crown-group Gadiformes appears to have occurred rapidly near the Cretaceous–Paleogene (K-Pg) boundary, generating three major lineages in a span of 15 million years. The earliest diverging lineage is Bregmacerotoidei (Bregmacerotidae) at ~78.5 Ma (CI 98.0–61.0 Ma) followed by the divergence of two large lineages, which diverged at ~71.7 Ma (CI 90.7–61.0). We tested if the recovered relationships within Gadiformes were influenced by the long branches leading to Bregmacerotidae. However, Bregmacerotidae was always recovered as the earliest branch and the ingroup relationships did not change even when the outgroups were removed, and the tree was rooted with the family. This suggests that the placement of Bregmacerotidae as sister to all remaining gadiforms is not sensitive to the systematic errors explored.

Our results consistently support four lineages following the divergence of Bregmacerotidae from other gadiforms. Gadoidei comprises four monophyletic families: Phycidae; Gaidropsaridae; Lotidae; and
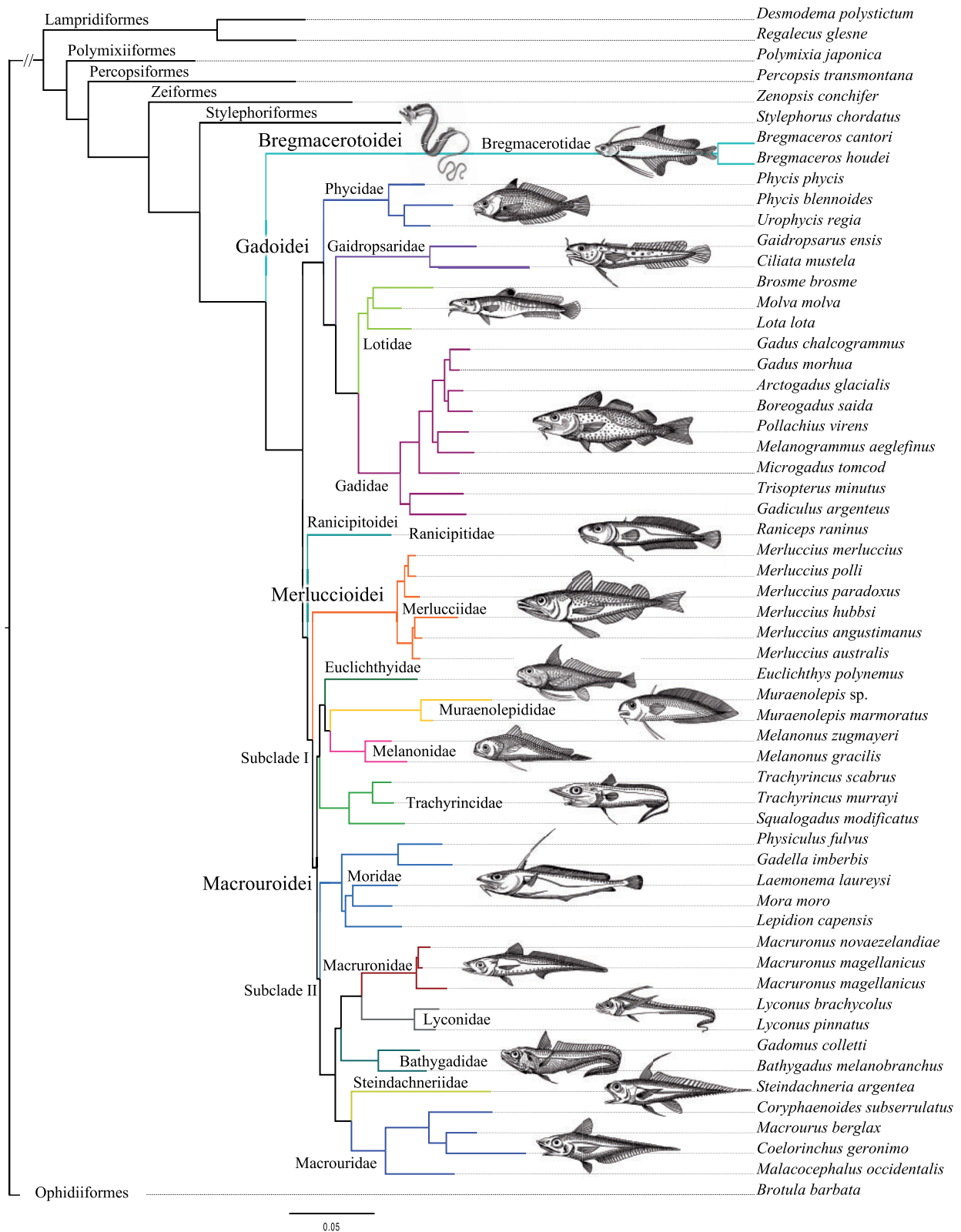
FIGURE 4.   Relationships of Gadiformes based on a concatenated alignment including 58 taxa (58T-SOS-1) using a partitioned data set by codon position and best evolutionary models (P1: GTR+F+R4; P2: GTR+F+R5; P3: GTR+F+R7) with their own branch lengths to account for heterotachy. All nodes had 100% bootstrap values except when noted.
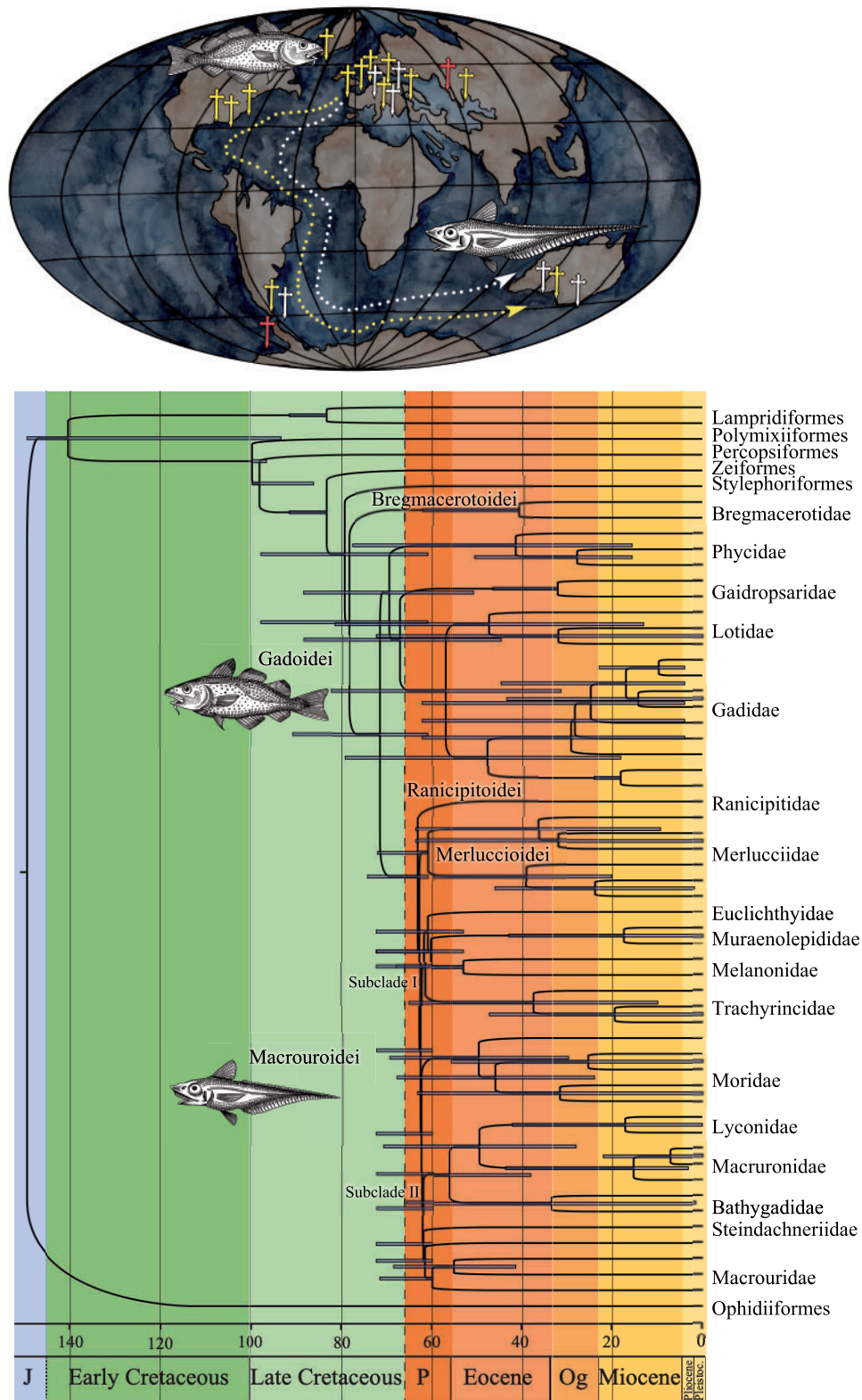
FIGURE 5. Timescale for the evolution of gadiform fishes. Phylogeny inferred for 58 taxa based on Maximum likelihood analyses of 8,244 loci (1,784,738 bp) and using a partitioned data set by codon position and best evolutionary models (P1: GTR+F+R4; P2: GTR+F+R5; P3: GTR+F+R7). Paleobiogeographic distribution of gadiform occurrences in the Paleocene and Eocene with potential migration routes of gadoids (yellow), macrouroids (white), and unknown (red) gadiform fishes according to Schwarzhans (1985, 2012), Nolf and Dockery (1993), Kriwet and Hecht (2008), Schwarzhans and Bratishko (2011), Marramá et al. 2019, and G. Carnevale (personal communication).

Gadidae. Extant members of Gadoidei, began to diverge during the Late Cretaceous to mid Eocene (69.7 Ma; CI 88.6–50.7 Ma). This clade includes one of the most intensively studied and exploited group of fishes—the commercially important Codfishes. The relationships of these fishes have not been resolved in previous morphological (e.g., Cohen 1984; Howes 1989; Endo 2002; Teletchea et al. 2006; Gaemers 2017; Gaemers and Poulsen 2017) and molecular studies (e.g., von der Heyden and Matthee 2008; Roa-Varón et al. 2009, Malstrøm et al. 2016). The divergence of Ranicipitoidei, including only the monotypic Ranicipitidae (*Raniceps raninus*), also occurred in the Late Cretaceous to mid Paleocene (~63.2 Ma; 74.5–61 Ma). Its placement within Gadoidei based on the amino acid data set and the edge-unlinked and GHOST models should be explored by further targeted analyses. The composition of Merluccioidei is restricted to Merlucciidae (including only the species of *Merluccius*), corroborating the original morphological concept of the family (Howes l991, Endo 2002).

The final gadiform lineage, Macrouroidei, is composed of two clades. The first clade in our preferred hypothesis (SOS-1; Macrouroidei - Subclade I) is ((Euclichthyidae, (Muraenolepididae, Melanonidae)), Trachyrincidae) and has strong support (100% BS). Bootstrap values drop in SOS-3 and SOS-5 (52% and 21% BS, respectively), and the gCF and sCF percentages of "decisive" gene trees and sites supporting the clade were lower and the relationships changed (S2d, S17; Supplementary Material available on Dryad). Differences between the SOS-1 tree and the amino acid and summary coalescent (ASTRAL) topologies are observed in the placement of Euclichthyidae and Muraenolepididae, but relationships among these families were weakly supported in both cases (S13, S14, Supplementary Material available on Dryad). Within the second clade, *Lyconus* and *Macrouronus* are the sole members of the Lyconidae and Macruronidae, respectively, and those families form the sister clade to Bathygadidae. *Steindachneria argentea*, as the only member of the Steindachneriidae, is the sister group of Macrouridae. The explosive diversification within Macrouridae, which gave rise to more than half of the extant gadiform species, occurred in the Late Cretaceous to mid Paleocene (~61.6 Ma; 72.5–60.0 Ma). The presence of the oldest fossil gadiforms from the Danian and Selandian of Europe and South Australia suggest a bipolar distribution by the early Paleogene, with gadoids reported in the Paleocene of North Atlantic area and North Sea Basin (Schwarzhans 2003, 2004; Kriwet and Hecht 2008; Schwarzhans and Bratishko 2011; Schwarzhans 2012; Marramá et al. 2019) and macrouroids reported in both the North Sea Basin and the Southern Ocean (Schwarzhans 1985; Kriwet and Hecht 2008; G. Carnevale, personal communication). Macrouridae traditionally have been recognized as comprising four subfamilies: Macrourinae, Bathygadinae, Trachyrincinae, and Macrouroidinae (Marshall 1965; Cohen 1984; Iwamoto 1989; Nolf and Steurbaut 1989; Cohen et al. 1990; Endo 2002). Hypotheses about the interrelationships among these families often have been contentious, and the family has been regarded as likely a paraphyletic assemblage within gadiforms (Howes 1989; Roa-Varón et al. 2009). In contrast, Okamura (1989) suggested a close relationship among macrourines, trachyrincids, bathygadines, and euclichthyids. Our study supports the recognition of three macrourid subfamilies as families: Macrourinae, Bathygadinae, and Trachyrincinae (taxa included historically in both Trachyrincinae and Macrouroidinae).

This is the first study including complete taxonomic sampling at the family level and has resolved the relationships among families of Gadiformes. This phylogeny will serve as the basis for mapping molecular and morphological adaptations that might have facilitated the transition(s) between freshwater and marine environments and between shallow and deep-sea habitats. This will result in a better understanding of geographic and ecological patterns of diversification. Furthermore, this phylogenetic framework will allow the assessment of modes of colonization of deep-water habitats by fishes, which has long been debated (e.g., Andriyashev 1953, White 1988).

Conclusions

Our phylogenomic analyses of the teleostean order Gadiformes extensively explored potential biological and methodological sources of systematic error. We demonstrated that discarding too many loci to reduce the relative amount of missing data without considering the actual signal in the data can be detrimental. Consequently, effective gene filtering that generates optimal data matrices in terms of potential phylogenetic signal requires finding a balance between data quantity and quality. Strategies used to improve the signal-to-noise ratio (e.g., partitioning the data, estimating the most realistic model of sequence evolution, and using models addressing the rate of variation across sites and lineages) had little impact on the resulting tree topology or were insufficient to deal with all the phenomena present in our genome-scale matrices.

Changes in the tree topology for three families were observed after removing the most divergent outgroups (54T-SOS-1) and modeling rate of variation across sites and lineages (Edge-proportional partition model, Edge-unlinked partition model, and GHOST), suggesting that theoretical work is needed to explore the relationship between outgroup rooting and LBA, especially how different outgroup taxon sampling strategies affect the probability of LBA.

Using multiple approaches for assessing phylogenetic confidence, such as gCF and sCF analyses, serves to complement measures such as bootstrap values by capturing variance in phylogenetic signal and identifying hidden conflict at different evolutionary scales. Bootstrap contraction thresholds (e.g., 20%)

recovered the monophyly of Lotidae and lead to greater accuracy in species tree inference using ASTRAL-III. Most of the gene-tree and species-tree discordance is a result of short times between divergences and consequent lack of informative characters at deep levels rather than a result of ILS.

This study provides a higher-level classification that is of operative value for a wide range of comparative studies of multiple traits in an intensively harvested group of fishes. The highly resolved backbone of our phylogeny alters historical perceptions of evolution within the group. Changes in classification are necessary and a revised classification comprising five suborders and 17 families is proposed. While not the final effort on inferring higher-level gadiform relationships, our analyses have full representation at the family level, including a number of difficult-to-obtain specimens (e.g., Euclichthyidae, Lyconidae, Ranicipitidae) that have not previously been included in a molecular study. In addition, this is the first explicitly time-calibrated tree using 15 fossil calibration points for Gadiformes, and the dating analyses indicate that the order probably originated in the North Atlantic and diversified rapidly in the late Cretaceous, a result confirmed by recent paleontological discoveries (G. Carnevale, personal communication).

The large amount of congruence across analyses and sequence-bias testing increases confidence in the results across different evolutionary depths. Future efforts will focus on refining the bait set to mask over-sequenced regions (evening read coverage across loci) and increasing taxonomic sampling within each family. Finally, our results reiterate the importance of examining phylogenomic data sets for evidence of systematic error that can emerge as a result of unsuitable modeling of biological and/or methodological issues, even when data sets are large and yield high support for phylogenetic relationships.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.5qfttdz2w.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

AUTHOR'S CONTRIBUTIONS

A.R-V. designed the experiments, carried out the lab work, designed and performed the data analyses, made all figures, and wrote the manuscript. R.B.D. participated in the data analyses and reviewed the final manuscript. G.C. selected appropriate calibration points and reviewed the final manuscript. L.T. provided support at an early stage of the Trinity assembly and reviewed the final manuscript. C.C.B. reviewed and edited multiple versions of the manuscript and supported A.R-V.'s pre-doctoral fellowship at the Smithsonian Institution, NMNH. C.L. provided training in the molecular approach to A.R-V. in his lab in Shanghai, pipeline and scripts associated with it, and reviewed the final manuscript. E.J.H. directed this study, provided funding, discussed results and approaches to study, and revised the manuscript at multiple stages. All authors read and approved the final manuscript.

REFERENCES

Andriyashev A.P. 1953. Ancient deep-water and secondary deep-water fishes and their importance in zoogeographical analysis. In: Notes on special problems in ichthyology. Akad Nauk SSSR Ikhtiol Kom, Moscow-Leningrad. [Translation by A.R. Gosline, Bureau of Commercial Fisheries, Ichthyological Laboratory, US National Museum, Washington, DC].

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19:716–723. doi:10.1109/TAC.1974.1100705.

Alfaro M.E., Faircloth B.C. Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Èerný D., Near T.J.? 2018. Explosive

diversification of marine fishes at the Cretaceous–Palaeogene boundary. Nat. Ecol. Evol. 2:688–696. doi:10.1038/s41559-018-0494-6.

Berg L.S. 1940. Classification of fishes both recent and fossil. Trav. Inst. Zool. Acad. Sci., USSR 5(2):517.

Bergsten J. 2005. A review of long-branch attraction. Cladistics. 21:163–193. doi:10.1111/j.1096-0031.2005.00059.x.

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-Varón A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia T., Ortí G. 2013. The tree of life and a new classification of bony fishes. PLoS Curr. 5:ecurrents.tol.53ba26640df0ccaee75bb165c8c26288. doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Bragg J.G., Potter S., Bi K., Moritz C. 2016. Exon capture phylogenetics: efficacy across scales of divergence. Mol. Ecol. Resour. 16:1059–1068. doi: 10.1111/1755-0998.12449.

Branstetter M.G., Danforth B.N., Pitts J.P., Faircloth B.C., Ward P.S., Buffington M.L., Gates M.W., Kula R.R., Brady S.G. 2017. Phylogenomics and improved taxon sampling resolve relationships among ants, bees, and stinging wasps. Curr. Biol. 27:1019–1024. doi: 10.1016/j.cub.2017.03.027.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: architecture and applications. BMC Bioinform. 10:42. doi: 10.1186/1471-2105-10-421.

Carr S.M., Kivlichan D.S., Pepin P., Crutcher D.C. 1999. Molecular systematics of gadid fishes: implications for the biogeographic origins of Pacific species. Can. J. Zool. 77:19–26.

Chen W-J., Santini F., Carnevale G., Chen J-N., Liu S-H., Lavoué S., Mayden R.L. 2014. New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). Front. Mar. Sci. 1:1–17. doi: 10.3389/fmars.2014.00053.

Cohen D.M. 1984. Gadiformes: overview. In: Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall A.W. Jr., Richardson S.L., editors. Ontogeny and systematics of fishes. Lawrence, KS: American Society of Ichthyologists and Herpetologists. p. 259–265.

Cohen D.M., editor. 1989. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. 262 p.

Cohen D.M., Inada T., Iwamoto T., Scialabba N. 1990. FAO species catalogue: Vol. 10 Gadiform fishes of the world. Rome: Food and Agriculture Organization of the United Nations. 442 p.

Crotty S.M., Minh B.Q., Bean N.G., Holland B.R., Tuke J., Jermiin L.S., von Haeseler A. 2019. GHOST: Recovering historical signal from heterotachously evolved sequence alignments. Syst. Biol. 69(2):249–264. doi: 10.1093/sysbio/syz051.

Davesne D., Gallut C., Barriel V., Janvier P., Lecointre G., Otero O. 2016. The phylogenetic intrarelationships of spiny-rayed fishes (Acanthomorpha, Teleostei, Actinopterygii): fossil taxa increase the congruence of morphology with molecular data. Front. Ecol. Evol. 4:129. doi: 10.3389/fevo.2016.00129.

Delsuc F., Brinkmann H., Chourrout D., Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 439:965–968. doi: 10.1038/nature04336.

Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Gen. 6:361–375. doi: 10.1038/nrg1603.

Dunn J.R., Matarese A.C. 1984. Gadidae: development and relationships. In: Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall A.W. Jr., Richardson S.L., editors. Ontogeny and systematics of fishes. Lawrence, KS: American Society of Ichthyologists and Herpetologists, Special Publication. p. 283–299.

Dutheil J.Y., Figuet E. 2015. Optimization of sequence alignments according to the number of sequences vs. number of sites trade-off. BMC Bioinformatics 16:e190. doi: 10.1186/s12859-015-0619-8.

Endo H. 2002. Phylogeny of the order Gadiformes (Teleostei, Paracanthopterygii). Mem. Grad. Sch. Fish. Sci. Hokkaido Univ. 49:75–149.

Fahay M.P., Markle D.F. 1984. Gadiformes: development and relationships. In: Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall A.W. Jr., Richardson S.L., editors. Ontogeny and systematics of fishes. Lawrence, KS: American Society of Ichthyologist and herpetologists Special Publication. p. 265–283.

Faircloth B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32(5):768–788.

Faircloth B.C., McCormack J.E., Crawford N.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers for target enrichment spanning multiple evolutionary timescales, Syst. Biol. 61(5):717–726. doi: 10.1093/sysbio/sys004.

FAO. 2016. The state of world fisheries and aquaculture (SOFIA). Contributing to food security and nutrition for all. Food and Agriculture Organization of the United Nations. Rome. 204 p.

Felsenstein J. 2004. Inferring phylogenies. Sunderland (MA): Sinauer Associates, Inc.

Foster P.G. 2004. Modeling compositional heterogeneity. Syst. Biol. 53(3):485–495. doi: 10.1080/10635150490445779.

Gaemers P.A.M. 2017. Taxonomy, distribution and evolution of Trisopterine Gadidae by means of otoliths and other characteristics. Fishes. 1:18–52. doi: 10.3390/fishes1010018.

Gaemers P.A.M., Poulsen J.Y. 2017. Recognition and distribution of two North Atlantic *Gadiculus* Species, *G. argenteus* and *G. thori* (Gadidae), Based on Otolith Morphology, Larval Pigmentation, Molecular Evidence, Morphometrics and Meristics. Fishes. 2(3):15. doi: 10.3390/fishes2030015.

Gilbert P.S., Chang J., Pan C., Sobel E.M., Sinsheimer J.S., Faircloth B.C., Alfaro M.E. 2015. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. Mol. Phylogen. Evol. 92:140–146.

Grande T., Borden W.C., Smith L. 2013. Limits and relationships of Paracanthopterygii: A molecular framework for evaluating past morphological hypotheses. In: Arratia G., Schultze H.-P., Wilson M.V.H., editors. Mesozoic Fishes 5—Global Diversity and Evolution. 6 figs., 4 apps. by Verlag Dr. Friedrich Pfeil, München, Germany. p. 385–418. ISBN 978-3-89937-159-8.

Grant W.S., Leslie R.W. 2001. Inter-ocean dispersal is an important mechanism in the zoogeography of hakes (Pisces: *Merluccius* spp.). J. Biogeogr. 28:699–721.

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. Bull. Am. Mus. Nat. Hist. 131:341–455.

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

Howes G.J. 1989. Phylogenetic relationships of macrourid and gadoid fishes based on cranial myology and arthrology. In: Cohen D.M., editor. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. p. 113–128.

Howes G.J. l991 Biogeography of gadoid fishes. J. Biogeogr. 18:595–622.

Huelsenbeck J.P., Hillis D.M. 1993. Success of phylogenetic methods in the 4-taxon case. Syst. Biol. 42(3):247–264. doi: 10.1093/sysbio/42.3.247.

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) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. USA. (24):6249–6254. doi: 10.1073/pnas.1719358115.

Inagaki Y., Simpson A.G.B., Dacks J.B., Roger A.J. 2004. Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study. Syst. Biol. 53:582–593.

Iwamoto T. 1989. Phylogeny of grenadiers (suborder Macrouroidei): another interpretation. In: Cohen D.M., editor. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. p. 159–173.

Jiang J., Yuan H., Zheng X., Wang Q., Kuang T., Li J., Liu J., Song S., Wang W., Cheng F., Li H., Huang J., Li C. 2019. Gene markers for exon capture and phylogenomics in ray-finned fishes. Ecol. Evol. 9(7):3973–3983. doi: 10.1002/ece3.5026.

Johnson G.D., Patterson C. 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. Bull. Mar. Sci. 52(1):554–626.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods, 14:587–589. doi: 10.1038/nmeth.4285.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30(4):772–780. doi: 10.1093/molbev/mst010 (v.7.221).

Kozlov A. (2018, June 16). amkozlov/raxml-ng: RAxML-NG v0.6.0 BETA (Version 0.6.0). Zenodo. http://doi.org/10.5281/zenodo.1291478.

Kolaczkowski B., Thornton J.W. (2004). Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature. 431(7011):980–984.

Kriwet J., Hecht T. 2008. A review of early gadiform evolution and diversification: first record of a rattail fish skull (Gadiformes, Macrouridae) from the Eocene of Antarctica, with otoliths preserved in situ. Naturwissenschaften. 95:899–907. doi: 10.1007/s00114-008-0409-5.

Kück P., Longo G.C. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11:81. doi: 10.1186/s12983-014-0081-x.

Kumar S., Hedges S.B. 2016a. Advances in time estimation methods for molecular data. Mol. Biol. Evol. 33(4):863–869.

Kumar S., Stecher G., Peterson D., Tamura K. 2012. M EGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinform. 28:2685–2686.

Kumar S., Tamura M.D., Sanderford V., Gray V., Kumar S. 2016b. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7):245–254.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34(3):772–773. https://doi.org/10.1093/molbev/msw260.

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61(5):727–744.

Li C., Hofreiter M., Straube N., Corrigan S., Naylor G.J.P. 2013. Capturing protein-coding genes across highly divergent species. BioTechniques. 54:321–326. doi: 10.2144/000114039.

Liu L., Xi Z., Davis C.C. 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32:791–805.

Malstrøm M., Matschiner M., Tørresen O.K., Star B., Snipen L.G., Hansen T.F., Baalsrud H.T., Nederbragt A.J., Hanel R., Salzburger W., Stenseth N.C., Jakobsen K.S., Jentoft S. 2016. Evolution of the immune system influences speciation rates in teleost fishes. Nat. genet. 48:1204–1210. doi: 10.1038/ng.3645.

Markle D.F. 1989. Aspects of character homology and phylogeny of the Gadiformes. In: Cohen D.M., editor. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. p. 59–88.

Marramà G., Carnevale G., Smirnov P.V., Trubin Y.S., Kriwet J. 2019. First report of Eocene gadiform fishes from the Trans-Urals (Sverdlovsk and Tyumen regions, Russia). J. Paleont. 93(5):1001–1009. doi: 10.1017/jpa.2019.15.

Marshall N.B. 1965. Systematic and biological studies of the macrourid fishes (Anacanthini-Teleosteii). Deep-Sea Res. 12:299–322.

Mathews S., Clements M.D., Beilstein M.A. 2010. A duplicate gene rooting of seed plants and the phylogenetic position of flowering plants. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365:383–395.

Meyer B., Meusemann K., Misof B. 2011. MARE v0.1.2-rc. https://www.zfmk.de.

Meyer M., Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Hard Protoc. 2010(6):pdb.prot5448. doi: 10.1101/pdb.prot5448. PMID: 20516186.

Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37(9):2727–2733. doi: 10.1093/molbev/msaa106.

Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics (ISMB special issue). 31(12):i44–i52. doi: 10.1093/bioinformatics/btv234.

Miya M., Holcroft N.I., Satoh T.P., Yamaguchi M., Nishida M., Wiley E. 2007. Mitochondrial genome and a nuclear gene indicate a novel phylogenetic position of deep-sea tube-eye fish (Stylephoridae). Ichthyological Research, 54(4):323–332. doi: 10.1007/s10228-007-0408-0.

Miya M., Kawaguchi A., Nishida M. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18:1993–2009.

Miya M., Satoh T.R., Nishida M. 2005. The phylogenetic position of toadfishes (Order Batrachoidiformes) in the higher ray-finned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. Biol. J. Linn. Soc. 85:289–306.

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

Møller P., Jordan A.D., Gravlund P., Steffensen J.F. 2002. Phylogenetic position of the cryopelagic codfish genus *Arctogadus* Drjagin, 1932 based on partial mitochondrial cytochrome b sequences. Polar Biol. 25:342–349.

Murray A.M., Wilson M.V.H. 1999. Contributions of fossils to the phylogenetic relationships of the percopsiform fishes (Teleostei: Paracanthopterygii): order restored. In: Arratia G., Schultze H.- P., editors. Mesozoic Fishes. Systematics and Fossil Record. Munich: Verlag Dr. Friedrich Pfeil. p. 397–411.

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. USA. 110(31):12738–12743.

Nelson J.S., Grande T.C., Wilson M.V.H. 2016. Fishes of the World. 8th ed. Hoboken (NJ): John Wiley and Sons.

Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32:268–274. doi: 10.1093/molbev/msu300.

Nolf D., Dockery D. 1993. Fish otoliths from the Matthews Landing Marl member (Porters Creek Formation), Paleocene of Alabama. Mississippi Geol. 14:24–39.

Nolf D., Steurbaut E. 1989. Evidence from otoliths for establishing relationships between gadiforms and other groups. In: Cohen D.M., editor. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. p. 37–45.

O'Hara T.D., Hugall A.F., Thuy B., Moussalli A. 2014. Phylogenomic resolution of the class Ophiuroidea unlocks a global microfossil record. Curr. Biol. 24:1874–1879.

Okamura O. 1989. Relationships of the suborder Macrouroidei and related groups, with comments on Merlucciidae and *Steindachneria*. In: Cohen D.M., editor. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. p. 129–142.

Patterson C., Rosen D.E. 1989. The Paracanthopterygii revisited: order and disorder. In: Cohen D.M., editor. Papers on the systematics of gadiform fishes. Science Series No. 32, Natural History Museum of Los Angeles County. p. 5–36.

Paulin C.D. 1983. A revision of the family Moridae (Pisces: Anacanthini) within the New Zealand region. Rec. Nat Mus MZ. 2:81–126.

Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9(3):e1000602. doi:10.1371/journal.pbio.1000602.

Philippe H., Delsuc F., Brinkmann H., Lartillot N. 2005. Phylogenomics. Annu. Rev. Ecol. Evol. Syst. 36(1):541–562. doi: 10.1146/annurev.ecolsys.35.112202.130205.

Philippe H., de Vienne D.M., Ranwez V., Roure B., Baurain D., Delsuc. F. 2017. Pitfalls in supermatrix phylogenomics. Eur. J. Taxon. 283:1–25. http://dx.doi.org/10.5852/ejt.2017.283.

Quinteiro J., Vidal R., Rey-Mendez M. 2000. Phylogeny and biogeographic history of hake (genus *Merluccius*), inferred from mitochondrial DNA control-region sequences. Mar. Biol. 136:163–174.

Rambaut A. 2012. Figtree v 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/.

Roa-Varón A., Ortí G. 2009. Phylogenetic relationships among families of Gadiformes (Teleostei, Paracanthopterygii) based on nuclear and mitochondrial data. Mol. Phylogen. Evol. 52:688–704.

Rokas A., King N., Finnerty J., Carroll S. B. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. Evol. Dev. 5:346–359.

Roldan M.I., Garcia-Marin J.L., Utter F.M., Pla C. 1999. Population genetic structure of European hake, *Merluccius merluccius*. Heredity. 81:327–334.

Rosen D.E., Patterson C. 1969. The structure and relationships of the paracanthopterygian fishes. Bull. Am. Mus. Nat. Hist. 141:357–474.

Roure B., Baurain D., Philippe H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol. 30:197–214. doi: 10.1093/molbev/mss208.

Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14(12):1218–1231.

Sanderson M.J., Wojciechowski M.F., Hu J.M., Khan T.S., Brady S.G. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. Mol. Biol. Evol. 17(5):782–797. doi: 10.1093/oxfordjournals.molbev.a026357.

Schwarzhans W. 1985. Tertiäre Otolithen aus South Australia und Victoria (Australien). Palaeo. Ichthyol. 3:1–60.

Schwarzhans W. 2003. Fish otoliths from the Paleocene of Denmark. Geol. Surv. Den. Greenl. Bull. 2:1–94.

Schwarzhans W. 2004. Fish otoliths from the Paleocene (Selandian) of West Greenland. Medd. Grønl. Geoscience. 42:1–32.

Schwarzhans W. 2012. Fish otoliths from the Paleocene of Bavaria (Kressenberg) and Austria (Kroisbach and Oiching-Graben). Palaeo. Ichthyol. 12:1–88.

Schwarzhans W., Bratishko A. 2011. The otoliths from the middle Paleocene of Luzanivka (Cherkasy district, Ukraine): Neues Jahrbuch fur Geologie und Palä ontologie, Abhandlungen, v. 261:83–110.

Schwarzhans W., Stringer G. L. 2020. Fish otoliths from the late Maastrichtian Kemp Clay (Texas, USA) and the early Danian Clayton Formation (Arkansas, USA) and an assessment of extinction and survival of teleost lineages across the K-Pg boundary based on otoliths. Riv. It. Paleont. Strat. 126:395–446.

Song S., Zhao J., Li C. 2017. Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences. Mol. Phylogenet. Evol. 111:44–55. doi: 10.1016/j.ympev.2017.03.014.

Svetovidov A.N. 1948. Treskoobraznye (Gadiformes). Fauna SSSR, Zoologicheskii Institut Akademi Nauk SSSR (n.s) 34, Ryby (fishes) 9(4):1–222 (in Russian; English translation, 1962, Jerusalem: Israel Program for Scientific Translations, 304).

Talavera G., Vila R. 2011. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evol. Biol. 11:315. doi: 10.1186/1471-2148-11-315.

Tamura K., Battistuzzi F.U., Billing-Ross P., Murillo O., Filipski A., Kular S. 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. USA. 109(47):19333–9338. https://doi.org/10.1073/pnas.1213199109.

Tamura K., Tao Q., Kumar S. 2018. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. Mol Biol Evol. 357:1770–1782.

Tao Q., Tamura K., MelloB., Kumar S. 2019. Reliable confidence intervals for RelTime estimates of evolutionary divergence times. Mol. Biol. Evol. 37(1):280–290. https://doi.org/10.1093/molbev/msz236.

Teletchea F., Laudet V., Hanni C. 2006. Phylogeny of the Gadidae (sensu Svetovidov, 1948) based on their morphology and two mitochondrial genes. Mol. Phylogenet. Evol. 38:189–199.

Torii A., Harold A.S., Ozawa T. 2003. Redescription of type specimens of three *Bregmaceros* species (Gadiformes: Bregmacerotidae): *B. bathymaster*, *B. rarisquamosus*, and *B. cayorum*. Memoirs of the Faculty of Fisheries. Kagoshima University. 52:23–32.

von der Heyden S., Matthee C.A. 2008. Towards resolving familial relationships within the Gadiformes and the resurrection of the Lyconidae. Mol. Phylogenet. Evol. 48:764–769.

Whelan N., Kocot K.M., Moroz L.L., Halanych K.M. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. Proc. Natl. Acad. Sci. USA. 112(18):5773–5778. https://doi.org/10.1073/pnas.1503453112.

White B.N. 1988. Oceanic anoxic events and allopatric speciation in the deep sea. Biol. Oceanog. 5(4):243–259. doi: 10.1080/01965581.1987.10749516.

Young A.D., Gillung J.P. 2020. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. Syst. Entom. 45:225–247. doi: 10.1111/syen.12406.

Yuan H., Jiang J., Jimenez F.A., Hoberg E.P., Cook J.A., Galbreath K.E., Li C. 2016. Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. Mol. Ecol. Resour. 16(5):1095–1106.