Molecular Phylogenetics and Evolution 110 (2017) 27–38

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset

Alex Dornburg [a,*], Jeffrey P. Townsend [b,c,d], Willa Brooks [a], Elizabeth Spriggs [b], Ron I. Eytan [e], Jon A. Moore [f,g], Peter C. Wainwright [h], Alan Lemmon [i], Emily Moriarty Lemmon [j], Thomas J. Near [b,k]

[a] North Carolina Museum of Natural Sciences, Raleigh, NC, USA
[b] Department of Ecology & Evolutionary Biology and Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA
[c] Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA
[d] Department of Biostatistics, Yale University, New Haven, CT 06510, USA
[e] Marine Biology Department, Texas A&M University at Galveston, Galveston, TX 77554, USA
[f] Florida Atlantic University, Wilkes Honors College, Jupiter, FL 33458, USA
[g] Florida Atlantic University, Harbor Branch Oceanographic Institution, Fort Pierce, FL 34946, USA
[h] Department of Evolution & Ecology, University of California, Davis, CA 95616, USA
[i] Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306, USA
[j] Department of Biological Science, Florida State University, 319 Stadium Drive, Tallahassee, FL 32306, USA
[k] Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA

ARTICLE INFO

*Article history:*
Received 12 April 2016
Revised 22 February 2017
Accepted 25 February 2017
Available online 27 February 2017

*Keywords:*
Homoplasy
Nucleotide saturation
Beryciformes
Codon bias
GC3
Anchored phylogenomics
Hybrid enrichment

ABSTRACT

Percomorph fishes represent over 17,100 species, including several model organisms and species of economic importance. Despite continuous advances in the resolution of the percomorph Tree of Life, resolution of the sister lineage to Percomorpha remains inconsistent but restricted to a small number of candidate lineages. Here we use an anchored hybrid enrichment (AHE) dataset of 132 loci with over 99,000 base pairs to identify the sister lineage of percomorph fishes. Initial analyses of this dataset failed to recover a strongly supported sister clade to Percomorpha, however, scrutiny of the AHE dataset revealed a bias towards high GC content at fast-evolving codon partitions (GC bias). By combining several existing approaches aimed at mitigating the impacts of convergence in GC bias, including RY coding and analyses of amino acids, we consistently recovered a strongly supported clade comprised of Holocentridae (squirrelfishes), Berycidae (Alfonsinos), Melamphaidae (bigscale fishes), Cetomimidae (flabby whalefishes), and Rondeletiidae (redmouth whalefishes) as the sister lineage to Percomorpha. Additionally, implementing phylogenetic informativeness (PI) based metrics as a filtration method yielded this same topology, suggesting PI based approaches will preferentially filter these fast-evolving regions and act in a manner consistent with other phylogenetic approaches aimed at mitigating GC bias. Our results provide a new perspective on a key issue for studies investigating the evolutionary history of more than one quarter of all living species of vertebrates.



## 1. Introduction

With more than 17,100 species, percomorph fishes comprise one out of every four of the world's living species of vertebrates (Eschmeyer and Fricke, 2015), including the majority of commercially important fishes (Rosenberg et al., 2005; Sethi et al., 2010; Rhyne et al., 2012), essential components in aquatic food webs (Carpenter et al., 1987; Winemiller, 1990; La Mesa et al., 2004; Rutschmann et al., 2011; Near et al., 2012a), and several model organisms in scientific studies (Brenner et al., 1993; Shapiro et al., 2004; Seehausen, 2006). Concomitant with this staggering diversity have come some of the most vexing problems in vertebrate phylogenetics. The last decade has nevertheless yielded unprecedented progress in achieving phylogenetic resolution of percomorphs, both among closely related species (Santini et al., 2013; Hundt et al., 2014; Santini and Carnevale, 2015; Thacker et al., 2015) and between major clades (Broughton et al., 2013; Chanet et al., 2013; Faircloth et al., 2013; Lautredou et al., 2013; Near et al., 2015; Sanciangco et al., 2015). While researchers have been honing in on a robust percomorph tree of life, resolution

* Corresponding author at: North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA.
*E-mail address:* alex.dornburg@naturalsciences.org (A. Dornburg).

http://dx.doi.org/10.1016/j.ympev.2017.02.017
1055-7903/

of the sister lineage to this remarkable radiation has remains elusive.

Recent molecular studies have consistently included several fish families as candidate percomorph sister lineages. However, the resolution of the relationships between these families and percomorphs remain divergent between studies. Over the last decade, molecular studies generally resolve one of the following three clades as the percomorph sister lineage: (1) the Holocentridae (squirrelfishes) (Betancur-R et al., 2013a); (2) a clade that includes Holocentridae, Berycidae (alfonsinos), and several families of enigmatic deep sea fishes including Melamphaidae (ridgeheads) and various whalefishes such as Cetomimidae and Rondeletiidae (Miya et al., 2003); or (3) a clade containing all of the previously mentioned lineages along with Trachichthyidae (roughies), Diretmidae (dories), Monocentridae (pinecone fish), Anoplogastridae (fangfishes), and Anomalopidae (flashlightfishes) (Miya et al., 2005; Smith and Wheeler, 2006; Near et al., 2012b, 2013); (Fig. 1). Identifying potential factors underlying this topological conflict is critical if we are to consistently resolve the sister lineage to percomorphs. Such efforts will serve to not only place studies of morphological and genomic evolution of teleost fishes into a broader context, but also provide a comparative perspective to investigate the processes that have generated a quarter of living vertebrate biodiversity.

The lack of resolution concerning the sister lineage of Percomorpha exemplifies an emerging problem in phylogenomics. Even with technological improvements in DNA sequencing (Schuster, 2007; Mardis, 2008; Davey et al., 2011; Faircloth et al., 2012) that have enabled a growing trend towards larger multi-locus and genomic scale datasets (Crawford et al., 2012b, 2015; Jarvis et al., 2014; Crawford et al., 2015; Prum et al., 2015), some nodes still continue to defy resolution (Jeffroy et al., 2006; Galtier and Daubin, 2008; Philippe et al., 2011; Morgan et al., 2013; Romiguier et al., 2013). This lack of topological resolution—one of several potential pitfalls of phylogenomics—is often attributed to large numbers of sequence characters whose state represents convergence and not shared evolutionary history, masking the phylogenetic signal of relationships (Jeffroy et al., 2006; Rokas and Chatzimanolis, 2008). Correspondingly, detecting and filtering homoplasious sites has become a critical step in phylogenomic inference (Townsend et al., 2012; Lin et al., 2013; Salichos and Rokas, 2013; Doyle et al., 2015).

Homoplasy is certainly not a problem unique to phylogenomics and has been a core issue since the advent of evolutionary thinking. Decades of literature within phylogenetics can attest to the challenges of discerning whether a morphological character represents a single origin, or is in fact the culmination of multiple independent origins (Goodman, 1967; Forey, 1973; Archie, 1989; Johnson and Patterson, 1993). Likewise, molecular data are not infallible and convergence in nucleotide state is common in rapidly evolving genomic regions found in clades as diverse as mammals (Parker et al., 2013), fishes (Betancur-R et al., 2013b), and plants (Cox et al., 2014). Despite the recognition of homoplasy as a problem, determining which portions of an alignment contribute to erroneous inference is not straightforward. Removing fast-evolving regions from analyses can certainly improve phylogenetic accuracy by eliminating homoplasious sites (Delsuc et al., 2005; Dornburg et al., 2014b). However, removing sites that are evolving at only a moderately fast rate can degrade accuracy by eliminating sites containing essential phylogenetic information (Pisani, 2004). This contrast suggests that the assessment of homoplasy requires a defined relationship between rates of character change and correct resolution of phylogenetic problems.

Integrating phylogenetic informativeness (PI) profiles (Townsend, 2007) with theoretical models of phylogenetic signal and noise (Townsend et al., 2012; Su et al., 2014) presents an intriguing approach for evaluating potential homoplasy in data sets. PI profiles will by their nature depict the severity to which phylogenetic information content has decayed over time (Townsend and Leuenberger, 2011), and investigators have readily coopted the method to test for the accumulation of homoplasious sites in an alignment (Crawley and Hilu, 2012; Dornburg et al., 2014b; Hilu et al., 2014; Gilbert et al., 2015). Recently, Prum et al. (2015) combined PI profiles with signal and noise models to assess whether the rate of decay was predicted to impact nodes of varying lengths and depths. Such an approach accounts for heterogeneity in the effect of homoplasy based on inter-node length. However, it remains to be determined whether this approach could be used as a dataset filtration metric.

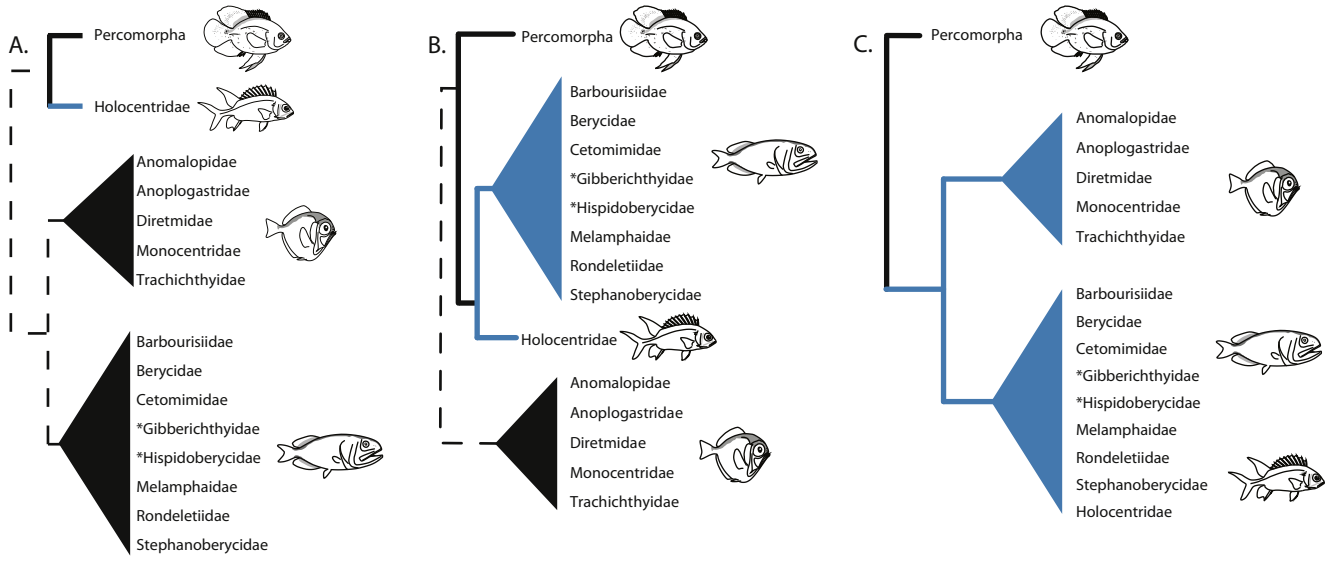


**Fig. 1.** Examples of previous molecular based studies resolving the sister lineage of Percomorpha: A. Betancur-R et al. (2013a); B. Miya et al. (2003); C. Miya et al. (2005), Near et al. (2012b). Highlighted branches indicate the inferred sister clade to Percomorpha, dashed lines indicate divergences prior to the most recent common ancestor of Percomorpha and its sister lineage. Lineages marked with a "*" indicate lineages not sampled in the referenced studies. Placement of these lineages in the tree is based on Nelson et al. (2016).

Here we assess whether we can reconcile incongruent phylogenetic hypotheses, or diagnose a lack of phylogenetic resolution, regarding the sister lineage of percomorph fishes using analyses of next-generation sequence data. With this aim, we sequenced 132 loci for species representing lineages that span the candidate pool of major lineages proposed as sister to Percomorpha as well as several percomorph, and outgroup taxa using anchored hybrid enrichment (AHE; Lemmon et al., 2012; Lemmon and Lemmon, 2013). We conducted a series of maximum likelihood (ML) phylogenetic inferences using the nucleotide sequence alignments in combination with a data filtration approach that combines signal and noise approaches (Townsend et al., 2012) with PI profiling (Townsend, 2007). We compared these analyses with a Bayesian analysis of the amino acid translations, a coalescent based species-tree analysis, and a maximum likelihood analysis of the data transformed using an RY coding scheme. Our analyses result in consistent phylogenetic resolution of a clade containing Holocentridae, Rondeletiidae, Cetomimidae, Berycidae, and Melamphaidae as the sister lineage to Percomorpha. These findings highlight that PI based filtration offers an additional tool in the phylogenomics toolbox for addressing homoplasy and elucidate a potential source of error that is increasingly ubiquitous in loci that are used to investigate phylogenetic relationships at moderate to deep timescales.

## 2. Methods

### 2.1. Sequence data acquisition

We assembled a phylogenomic dataset that spans most major lineages in the candidate pool of lineages sister to Percomorpha. Our sampling includes a representative species of each of the following lineages: Anomalopidae, Anoplogastridae, Monocentridae, Diretmidae, Trachichthyidae, Holocentridae, Rondeletiidae, Berycidae, Melamphaidae, and Cetomimidae, as well as percomorph species of Ovalentaria and Notothenioidei. We included three species to serve as outgroups in our phylogenomic analyses: *Zenopsis conchifera*, *Macrourus whitsoni*, and *Percopsis omiscomycus* that are consistently considered as more deeply diverging in both molecular and morphological studies (Johnson and Patterson, 1993; Moore, 1993b, 1993a; Colgan et al., 2000; Miya et al., 2003; Alfaro et al., 2009; Santini et al., 2009; Near et al., 2012b, 2013; Betancur-R et al., 2013a). A tissue sample of a holocentrid (*Myripristis jacobus*) was field collected in Curaçao and deposited in the fish tissue collection of the Yale Peabody Museum (YFTC). One individual per major lineage was additionally obtained through the YFTC (Table 1). Muscle tissue biopsies were stored in 95% ethanol or RNAlater (Qiagen, Valencia, CA) and DNA was extracted using Qiagen DNAeasy Tissue Extraction Kits (Qiagen, Valencia, CA). Following extraction, DNA yields were quantified using a Thermo Scientific NanoDrop. Library preparation, enrichment, and sequencing protocols followed the vertebrate protocol of Lemmon et al. (2012). Data were collected at the Center for Anchored Phylogenomics (www.anchoredphylogeny.com) at Florida State University. After extraction, genomic DNA was sonicated to a fragment size of ~150–400 bp using a Covaris E220 Focused-ultrasonicator with Covaris microTUBES. Subsequently, library preparation and indexing were performed by hand following a protocol modified from Meyer and Kircher (2010). Indexed samples were then pooled at equal quantities, and then the library pool was enriched using the Vertebrate v.1 kit (Agilent Technologies Custom SureSelect XTd; enrichment probes available from Lemmon et al., 2012). This enrichment kit targets 512 conserved regions, primarily located in exons. The enriched library pool was then sequenced on 1 PE100 Illumina HiSeq 2000 lane at the Trans-

**Table 1**
Summary statistics of the anchored hybrid enrichment capture protocol prior to dataset curation by species.

| Species | Family | YFTC Number | Reads HiSeq PE100 | Reads MiSeq PE150 | Contigs | Bases in Contigs | Reads in Contigs | Enrichment Efficiency | Loci | Average Locus Length | Percent reads in assembly | Coverage per locus | Coverage per contig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Macrourus whitsoni* | Macrouridae | 21825 | 11368266 | 522870 | 763 | 31789208 | 345563 | 213 | 262 | 853.45 | 2.91 | 1318.94 | 41663.44 |
| *Zenopsis conchifera* | Zeidae | 13730 | 14557502 | 683468 | 700 | 27281383 | 295723 | 139 | 260 | 803.43 | 1.94 | 1137.40 | 38973.40 |
| *Percopsis omiscomycus* | Percopsidae | 21744 | 13350502 | 653310 | 722 | 34499559 | 374487 | 192 | 270 | 954.98 | 2.67 | 1386.99 | 47783.32 |
| *Poromitra curilensis* | Melamphaidae | 22142 | 12745170 | 514984 | 686 | 21026106 | 242710 | 131 | 247 | 770.80 | 1.83 | 982.63 | 30650.30 |
| *Myripristis jacobus* | Holocentridae | 13625 | 15086364 | 638348 | 853 | 40386706 | 414328 | 189 | 274 | 787.97 | 2.63 | 1512.15 | 47346.67 |
| *Anomalops katoptron* | Anomalopidae | 22123 | 14956914 | 576002 | 685 | 33672444 | 282845 | 130 | 274 | 831.07 | 1.82 | 1032.28 | 49156.85 |
| *Gyrinomimus bruuni* | Cetomimidae | 13820 | 11147082 | 477296 | 902 | 39478823 | 416341 | 256 | 282 | 833.15 | 3.58 | 1476.39 | 43768.10 |
| *Diretmoides pauciradiatus* | Diretmidae | 14534 | 10673908 | 445574 | 849 | 56790239 | 600890 | 264 | 293 | 949.75 | 5.40 | 2050.82 | 66890.74 |
| *Anoplogaster cornuta* | Anoplogasteridae | 14555 | 10183850 | 443340 | 837 | 55589697 | 549827 | 371 | 271 | 843.21 | 5.17 | 2028.88 | 66415.41 |
| *Monocentrus reidi* | Monocentridae | 21753 | 16872564 | 658744 | 809 | 42132578 | 440262 | 296 | 280 | 872.54 | 2.51 | 1572.36 | 52079.82 |
| *Aulotrachichthys sajademalensis* | Trachichthyidae | 14889 | 10349390 | 437792 | 769 | 35469461 | 384082 | 232 | 275 | 864.86 | 3.56 | 1396.66 | 46124.14 |
| *Rondeletia loricata* | Rondeletiidae | 13628 | 10160056 | 470604 | 1012 | 61904146 | 640383 | 588 | 299 | 983.68 | 6.02 | 2141.75 | 61170.10 |
| *Beryx decadactylus* | Berycidae | 11163 | 12798910 | 503622 | 741 | 48430525 | 462089 | 240 | 250 | 923.58 | 3.47 | 1848.36 | 65358.33 |
| *Menidia menidia* | Atherinopsidae | 17777 | 9616306 | 371352 | 968 | 46934439 | 500858 | 396 | 265 | 920.52 | 5.01 | 1890.03 | 48485.99 |
| *Bovichtus diacanthus* | Bovichtidae | 3477 | 10696248 | 509238 | 899 | 50186840 | 529780 | 500 | 305 | 970.93 | 4.73 | 1736.98 | 55825.18 |
| | Average | | 12304202.13 | 527102.93 | 813 | 41704810.27 | 432011.2 | 275.8 | 273.8 | 877.59 | 3.55 | 1567.50 | 50779.45 |

Coverage represents the maximum number of reads overlapping a site within an assembled locus, efficiency represents the percent of reads that assembled to a locus post data filtration (Lemmon et al., 2012) divided by the percentage that would have been obtained if locus fragments were randomly distributed across a given genome.

lational Science Laboratory in the College of Medicine at Florida State University.

Loci captured using AHE were assessed for paralogs using the AHE pipeline developed by Lemmon et al. (2012) and recently outlined by Eytan et al. (2015). Briefly, this approach works by selecting a reference individual based on capture efficiency. Consensus sequences for each locus were aligned to the reference taxon's sequences. Homolog sets are then constructed algorithmically for each locus as follows: the first homolog set represents the first homolog identified for the first individual and the sequences from each other individual that had the greatest sequence similarity to the reference sequence. This set is then removed from the candidate pool of sequences. A second homolog set is then constructed from the next reference sequence, that may represent a second copy of a duplicated gene, and the corresponding sequences from each individual that best align to that reference sequence. These sequences are then removed, and the iterative process continues until all potential homolog sets are constructed. Non-exon targets were filtered and genes were edited and aligned in Geneious v7.0.6 (Drummond et al., 2010), with the resulting nucleotide alignment consisting of 99,948 base pairs. All raw data, consensus sequences, and alignments are available on Zenodo (DOI: 105281/zenodo.321684).

### 2.2. Phylogenetic information and filtration

Although AHE is based on sequencing loci that are conserved across all vertebrates (Lemmon et al., 2012), unequal rates of nucleotide substitution, in particular saturation of third codon positions, have been observed in other phylogenomic datasets over deep timescales (Jeffroy et al., 2006; Dávalos and Perkins, 2008; Lin et al., 2013) and can create discordance between amino acid and nucleotide inferred topologies (Rota-Stabelli et al., 2013). Thus, we quantified site-specific rates of substitution for each locus and codon position using the program Hyphy (Pond and Muse, 2005) in the PhyDesign web interface (Lopez-Giraldez and Townsend, 2011) with a pruned chronogram from Near et al. (2013). To assess general trends of information content over the temporal history of the taxa in this study, profiles of phylogenetic informativeness (2007) for the concatenated alignment, each locus, and each codon position were generated from the site-specific rates using the R package PhyInformR (Dornburg et al., 2016; available on CRAN and at https://github.com/carolinafishes/PhyInformR).

As PI profiles provide no direct prediction of how homoplasious site patterns will influence phylogenetic resolution (Klopfstein et al., 2010; Townsend et al., 2012), we also evaluated "signal" versus "noise" probabilities (Townsend et al., 2012) for the phylogenetic quartet representing the most recent common ancestor of Percomorpha and "Beryciformes" (sensu Near et al., 2013). We used the quartet internode homoplasy probability (QIHP), a quantification of "noise" specified in Townsend et al. (2012). Under a set of asymptotic assumptions, QIHP represents the probability of having greater strength of support at a given internode for an incorrect rather than correct quartet topology. Calculation of QIHP requires specification of heights for nodes of interest. Because these internode lengths are typically imprecisely known, PI profiles were initially used to guide selection of codon partitions and internode lengths that would maximize the range of the QIHP value distribution from the posterior distribution of trees. Since declines in PI profiles are correlated with increasing instances of homoplasy in an alignment (Townsend and Leuenberger, 2011), this usage facilitates the "worst-case" quantification of QIHP values, as the resulting internodes will include the steepest declines in PI and the correspondingly highest QIHP values.

Forty-five codon partitions from the genes that span the variation in PI profile shapes (Fig. 2) were used to calculate the QIHP across the posterior distribution of branch lengths that represent the most recent common ancestor (MRCA) of Percomorpha and its sister lineage from the Near et al. (2012b) study. The resulting distribution of QIHP facilitated the identification of the tree depth and internode distances that maximized the interval of QIHP values, as outlined in the approach above. With the identified combination of branch lengths, we calculated the QIHP for first, second, and third position sites within codons for each gene, and retained those partitions up to the locus-median change point of QIHP using the R package changepoint (Killick and Eckley, 2014). This filtration approach was repeated nine additional times using random draws of branch lengths from the posterior distribution of guide trees yielding a total of 10 datasets that contained between 69,6555 and 87,397 nucleotides. To assess the influence of the guide tree on rate estimates and the retention of loci, this entire process was also conducted using a guide tree that forced the Holocentridae to be the sister lineage to Percomorpha. Base frequencies were also compared between the filtered and unfiltered datasets, to facilitate identification of deviations from stationarity.

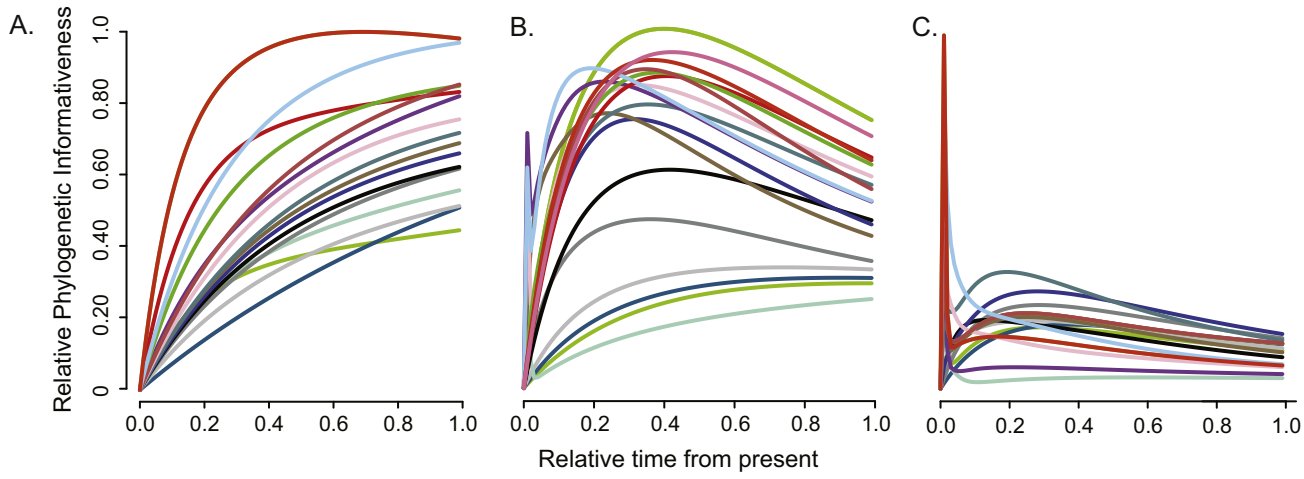


**Fig. 2.** Sample of relative PI profiles from the anchored hybrid enrichment based dataset that span the variance of all visualized PI profiles. A-C represent the three bin sets that span the variance in profile shape. The X-axis reflects time from present and the y-axis reflects phylogenetic informativeness. Line's within each plot correspond to the PI profile of individual codon partitions. PI profiles are scaled to the maximum height in each subset.

### 2.3. Phylogenomic analysis

We assembled four permutations of the AHE nucleotide dataset for ML analyses: (1) a concatenated analysis of the nucleotide dataset containing all loci partitioned by codon position, (2) a concatenated analysis of the nucleotide dataset containing only the first and second codon positions partitioned by codon position, (3) a concatenated analysis of the nucleotide dataset, partitioned by codon position, with third codon positions RY coded (Phillips and Penny, 2003; Phillips et al., 2004), and (4) a series of concatenated analyses of the nucleotide codon partitions with values below the locus-median change point of QIHP (10 analyses per initial guide tree).

For all four datasets, ML analyses were conducted using the Randomized Accelerated Maximum Likelihood (RAxML) software v. 7.3.0 (Stamatakis, 2006). Nucleotide analyses were conducted under a GTR + Γ model of nucleotide substitution. For each analysis, 200 ML searches were conducted using a random starting tree (-d setting) and a random search convergence criterion based on Robinson-Foulds distances (-D setting). One thousand bootstrap replicates were conducted using the thorough bootstrap search (option –f i) for each dataset and confidence values were mapped onto the tree topology with the highest ML inferred from the random searches.

We additionally translated the nucleotide sequences into their corresponding amino acid sequences using Geneious v7.0.6 (Drummond et al., 2010). A Bayesian analysis of the amino acid data was conducted using the site heterogeneous CAT-GTR + Γ4 (Lartillot and Philippe, 2004) model in Phylobayes MPI 1.5 (Lartillot et al., 2013). Constant sites were removed to prevent biasing the CAT model profiles using the –dc option (Lartillot et al., 2013). We ran four independent Monte Carlo Markov chains for 20,000 iterations, discarded a burn-in of the first 2000 iterations, then sampled parameter values every 10 iterations to compose an approximation of their joint posterior distribution.

To assess the impact of potential gene-tree discordance on our concatenated analysis, we conducted a gene-tree species-tree analysis using ASTRAL, which has been found to be more accurate than other leading methods for disentangling incomplete lineage sorting from phylogenomic scale analyses (Mirarab et al., 2014). First ML analyses were conducted on each individual locus as well as loci that were RY coded, and filtered using the procedure outlined above. Species-tree inferences were then conducted using the resulting ML gene trees for each protocol and the exact algorithm in ASTRAL (Mirarab et al., 2014) using 1000 multi-locus bootstrap replicates to account for topological uncertainty in gene-tree inference (Seo, 2008).

## 3. Results

### 3.1. Phylogenomic data capture

Anchored hybrid enrichment resulted in an average of 12,304,202 reads per species using the HiSeq and 527,102 on the MiSeq (Table 1), read numbers that are consistent with the expectations of Lemmon et al. (2012) on both Illumina HiSeq 2000 and MiSeq sequencers. Contig assembly produced an average of 813 contigs and an averaged total of 41,704,810 base pairs per species (Table 1). Enrichment resulted in an average capture of 275 loci spanning approximately 877 base pairs per locus, with a minimum of 130 loci for the flashlight fish *Anomalops katoptron* and a maximum of 305 loci for the notothenioid *Bovichtus diacanthus* (Table 1). These values of loci are in line with the expectations (Lemmon et al., 2012) given that the probe design model, *Danio*, is more than 260 million years divergent from our focal group (Lemmon et al., 2012; Near et al., 2012b). Limiting our dataset to only loci that were captured for more than 90% of all taxa and include greater than 450 base pairs, we assembled a data matrix containing 132 loci with 99,948 base pairs, after removing trailing edges from the individual gene alignments (Table 2).

### 3.2. Maximum likelihood analysis

Maximum likelihood analysis of the concatenated dataset resulted in a tree topology that features a clade containing all candidate lineages as the sister lineage to Percomorpha (Fig. 3A). Monophyly of this inclusive clade was not strongly supported [Bootstrap support (BSS) < 70], though monophyly of all subclades was strongly supported (BSS > 95). One of these clades comprises Diretmidae as sister to a clade of two additional clades, respectively containing Anoplogasteridae and Trachichthyidae, and Monocentridae and Anomalopidae, with strong support present for all nodes (Fig. 3A; BSS = 100). In the other clade Holocentridae is resolved as the sister lineage to a clade which includes two lineages, whalefishes (Cetomimidae and Rondeletiidae), and Melamphaidae and Berycidae with strong support for all of the nodes (BSS = 100; Fig. 3A). Visualizations of the PI profiles for the combined codon positions of the AHE data suggest that third position sites decline in informativeness prior to the origin of Percomorpha (Fig. 4). Despite this decline in signal, inclusion of third positions did not at first appear to contribute significant noise: analysis of the nucleotides with (Fig. 3A) and without third codon positions (Fig. 3B) resulted in identical topologies, with both sets of analyses strongly supporting the same relationship among candidate lineages (BSS > 95; Fig. 3A and B), while providing weak support for

**Table 2**
Results of anchored hybrid enrichment capture by species following dataset curation. N indicates number of ambiguous base reads and missing indicates missing data.

| Species | Family | Number of loci | N/% | Missing/% | GC% | GC 1st Codon | GC 2nd Codon | GC 3rd Codon |
|---|---|---|---|---|---|---|---|---|
| *Macrourus whitsoni* | Macrouridae | 122 | 235/0.3 | 20,429/20.5 | 47.8 | 56.8 | 42.6 | 80.7 |
| *Zenopsis conchifera* | Zeidae | 129 | 158/0.2 | 8722/8.7 | 51.9 | 54.8 | 42.8 | 73 |
| *Percopsis omiscomycus* | Percopsidae | 129 | 62/0.1 | 9195/9.2 | 52.7 | 50.3 | 39 | 68.5 |
| *Poromitra curilensis* | Melamphaidae | 129 | 362/0.4 | 14,194/14.2 | 49.3 | 54.6 | 42.9 | 75.1 |
| *Myripristis jacobus* | Holocentridae | 131 | 208/0.2 | 5205/5.2 | 51.1 | 53.7 | 42.8 | 65.3 |
| *Anomalops katoptron* | Anomalopidae | 128 | 63/0.1 | 1748/0.7 | 52.9 | 52.8 | 41.8 | 63.8 |
| *Gyrinomimus bruuni* | Cetomimidae | 129 | 379/0.4 | 15,132/15.3 | 47.4 | 54.3 | 43.1 | 70.8 |
| *Diretmoides pauciradiatus* | Diretmidae | 132 | 324/0.4 | 10,064/10.1 | 49.2 | 53.9 | 42.8 | 67.6 |
| *Anoplogaster cornuta* | Anoplogasteridae | 128 | 293/0.3 | 10,155/10.2 | 49.9 | 54.2 | 42.5 | 70 |
| *Monocentrus reidi* | Monocentridae | 132 | 61/0.1 | 2374/2.4 | 52.8 | 52.5 | 41.6 | 63.8 |
| *Aulotrachichthys sajademalensis* | Trachichthyidae | 132 | 113/0.1 | 1921/1.9 | 54.6 | 53.9 | 42.6 | 66.9 |
| *Rondeletia loricata* | Rondeletiidae | 131 | 243/0.3 | 9418/9.4 | 49.8 | 53.8 | 42.7 | 68.4 |
| *Beryx decadactylus* | Berycidae | 132 | 204/0.2 | 4259/4.3 | 53.4 | 54.1 | 42.9 | 70 |
| *Menidia menidia* | Atherinopsidae | 122 | 180/0.2 | 14,894/14.9 | 46.8 | 53.9 | 43 | 68.3 |
| *Bovichtus diacanthus* | Bovichtidae | 130 | 188/0.2 | 5980/6.0 | 50.3 | 53.4 | 42.5 | 64.5 |
| | **Total**: | 132 | 3073/0.2 | 133,870/8.9 | 55.7 | 53.8 | 42.3 | 69.1 |

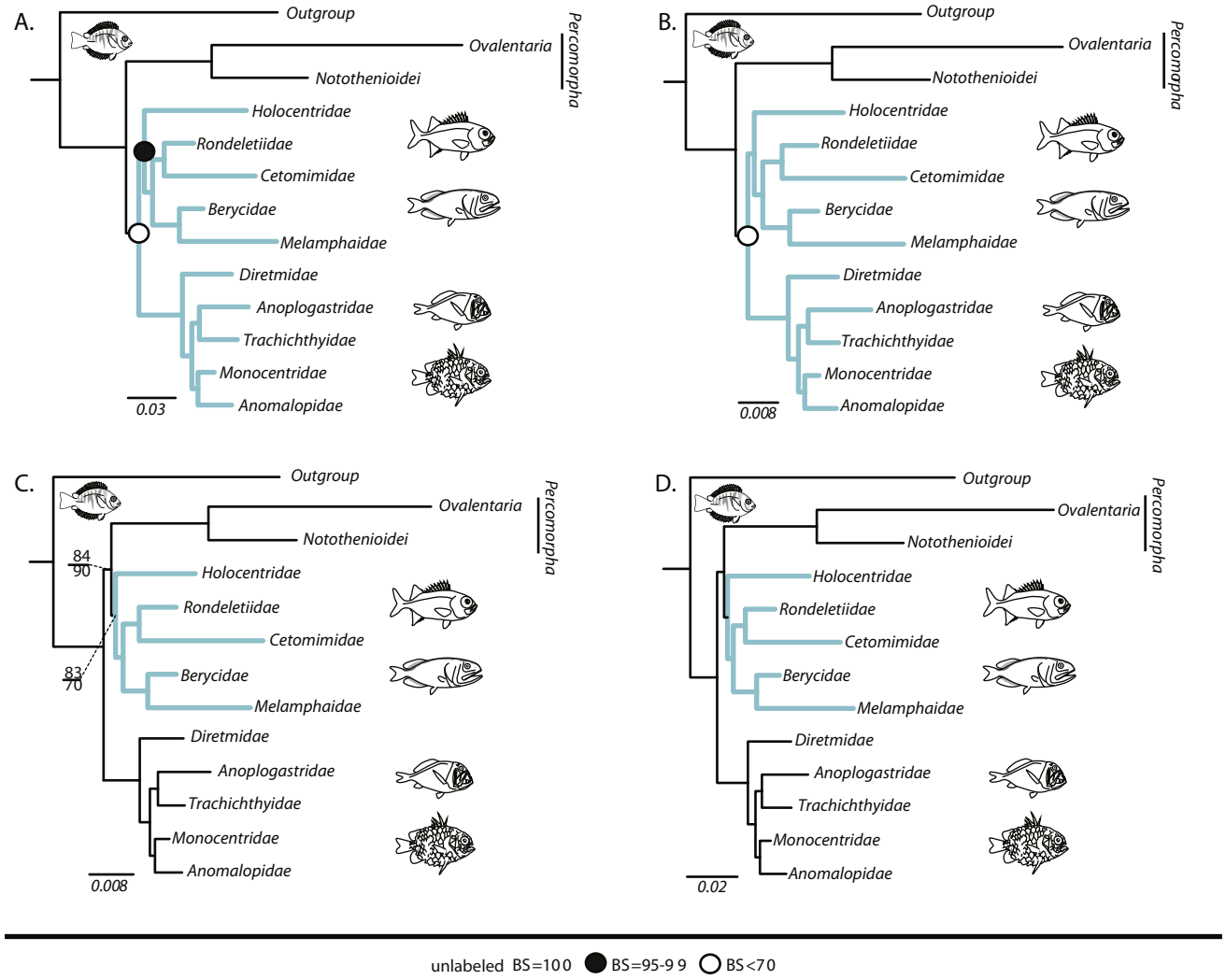


**Fig. 3.** Phylogenetic relationships based on maximum likelihood analyses of the anchored hybrid enrichment nucleotide data. A. Maximum likelihood inferred phylogeny based on the concatenated nucleotide alignment of 132 genes. B. Maximum likelihood inferred phylogeny based on the concatenated nucleotide alignment of 132 genes with all third codon positions removed. C. Maximum likelihood inferred phylogeny based on the concatenated nucleotide alignment of codon positions based on the locus-median change point of QIHP. D. Maximum likelihood inferred phylogeny based on the concatenated nucleotide alignment with third positions RY-coded. No circles at nodes represent bootstrap support values (BS) of 100, black circles represent BSS between 95 and 99, and open circles represent BS of 70–90. Node values above each node in C indicate median BSS values based on analyses using site rates estimated on the guide topology of Near et al. (2013), while values below each node indicate median BSS values from analyses conditioned on site rites from a guide tree depicting Holocentridae as sister to Percomorpha. Highlighted branches indicate the inferred sister clade to Percomorpha. All branch lengths are in substitution units.

the resolution of the sister lineage to Percomorpha (BSS < 70; Fig. 3B).

When conditioning analyses on partitions of below the locus median change point of QIHP, the topology deviates from the previous analyses of the concatenated matrix. There is moderate to strong support for a clade containing Holocentridae, Rondeletiidae, Cetomimidae, Berycidae, and Melamphaidae (Fig. 3C; 84 < BSS < 90) as the sister lineage to Percomorpha across all replicates regardless of guidetree chosen (Fig. 3C; 70 < BSS < 83). Relationships within this clade remain identical to those inferred in the previous two analyses. This topology also strongly supports the previously inferred relationships among Diretmidae, Anoplogastridae, Trachichthyidae, Monocentridae, and Anomalopidae (Fig. 3C; BSS = 100). Quantification of base frequencies from the filtered versus retained datasets demonstrates that filtration based on QIHP largely omitted partitions with elevated GC content (Fig. 5). Although QIHP calculations do not assume a set topology, starting topologies have the potential to alter site rate estimates. However, comparisons of site rates generated under different starting topologies revealed a strong correlation, suggesting resulting QIHP calculations to be robust to different starting guide trees (Fig. S1), as is evident from our visualization of locus-median change points (Fig. S2). When accounting for GC bias potentially misleading inference by RY coding of third codon positions (Phillips and Penny, 2003; Phillips et al., 2004), (Fig. 3D), results are identical to those conditioned on QIHP values below the locus median change point (Fig. 3D).

### 3.3. Amino acid and species-tree analyses

Bayesian analyses of the concatenated amino acid dataset under a CAT-GTR + Γ4 model results in a well-supported phylogeny that is identical to that inferred from the filtered nucleotide dataset with partitions conditioned on the degree of homoplasy (Fig. 6). These results were also identical to the topology inferred using a multi-coalescent model and all individual loci (Fig. S2A) though

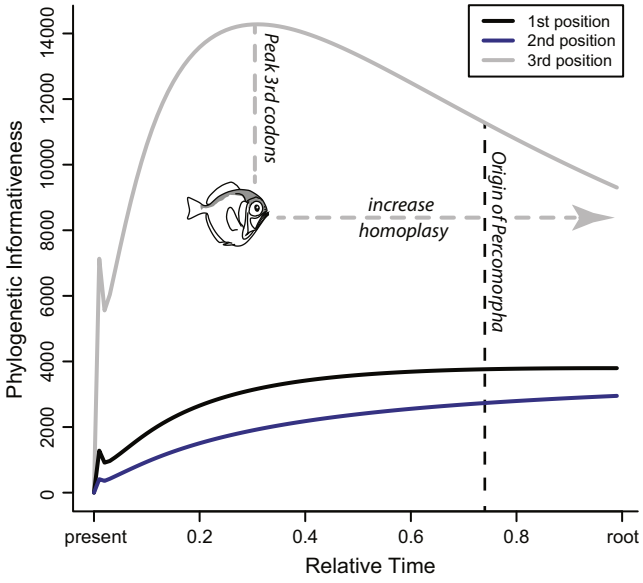


**Fig. 4.** Visualizations of the phylogenetic informativeness profiles for each codon position in the anchored hybrid enrichment dataset.

support was low. Analyses conditioned on loci that were RY coded yielded strong support for an identical topology to that conditioned on all loci, albeit with stronger support (Fig. S2B). Together these results support a clade comprised of Holocentridae, Rondeletiidae, Cetomimidae, Berycidae, and Melamphaidae as the sister lineage to Percomorpha.

## 4. Discussion

### 4.1. Bias and phylogenomic inference

Convergence in nucleotide state is predicted to mislead inference based on the stochastic interaction of time, internode length, and substitution rate underlying the resulting site patterns. Our results empirically validate these theoretical expectations (Townsend et al., 2012), and underscore the observation that homoplasy is not always apparent and will not necessarily lead to topological discordance (Roje, 2010). Many results in our study, such as the phylogenetic resolution of Berycidae, were unaffected by which data set was used in the analysis. However, as exemplified by our efforts to resolve the sister lineage to Percomorpha, patterns of convergence such as those generated by compositional bias are highly problematic for topological inference for some taxa.

Phylogenomic studies have observed that portions of the genome with a bias toward one or more of the four nucleotides, such as the GC bias in the third codon position of mammals, can mislead phylogenetic inference (Foster and Hickey, 1999; Mooers and Holmes, 2000; Romiguier et al., 2013; Cox et al., 2014). Screening our AHE dataset sampled from acanthopterygian teleosts revealed an elevated frequency of GC at the third codon position that ranged between 65% and 80% across all genes in a given species. The GC content in the first and second codon positions ranged between 42% and 54% (Table 2). This result mirrors a high GC3 in the markers commonly used for fish phylogenetic studies (Table 3) over the past decade (Dornburg et al., 2008, 2015; Alfaro et al., 2009; Near et al., 2012b; Wainwright et al., 2012; Betancur-R et al., 2013a;

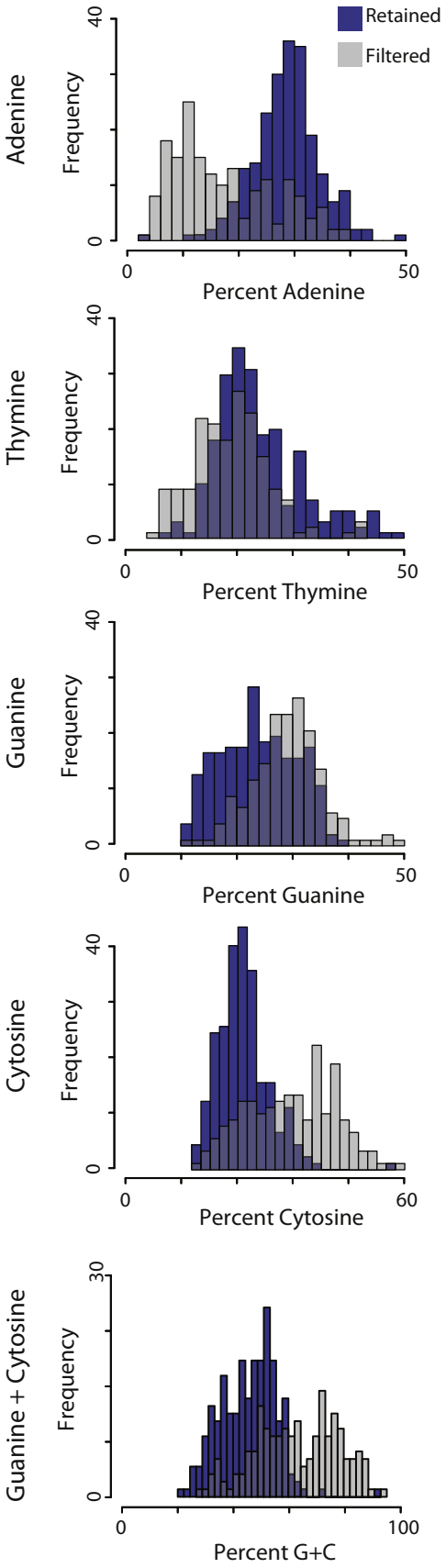


**Fig. 5.** Nucleotide frequencies of the codon partitions in one of the retained (blue) versus filtered (grey) data sets. Frequencies within each panel correspond to the frequency within each partition of either adenine, thymine, guanine, cytosine, or guanine + cytosine. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

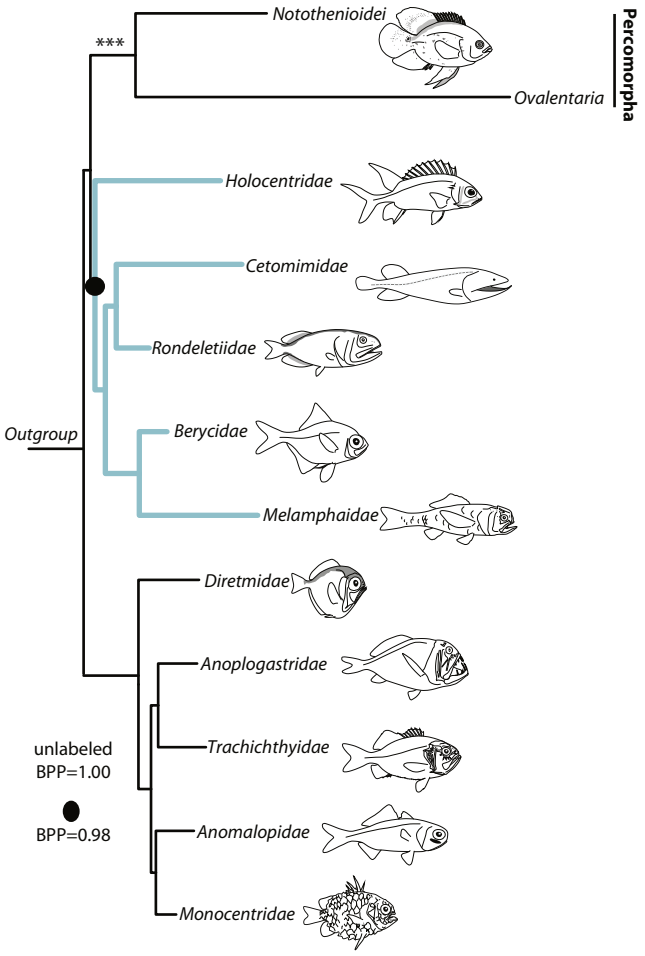


**Fig. 6.** Maximum clade credibility tree of the posterior distribution of phylogenies based on amino acid analysis of next-generation sequence data using PhyloBayes. No circles at nodes represent Bayesian Posterior Probability (BPP) estimates of 1.00 and the black circle represents a BPP of 0.98. All branch lengths are in substitution units and the highlighted branches correspond with the inferred sister lineage to Percomorpha. *** symbols indicate a branch length scaling of 25% of all subtending branches to accommodate graphical constraints.

Dornburg et al., 2015; Santini and Carnevale, 2015). Although the mechanisms generating codon biases are heterogeneous (Galtier et al., 2001; Wan et al., 2004; Plotkin and Kudla, 2010), codon biases are a ubiquitous genomic phenomenon that has been observed in taxa as disparate as mammals (Gustafsson et al., 2004; Lartillot, 2013; Romiguier et al., 2013), percomorph teleosts (Betancur-R et al., 2013b), *Drosophila* (Powell and Moriyama, 1997), and viruses (Shackelton et al., 2006; Wong et al., 2010). Although codon biases are often found in highly conserved and highly expressed genes (Sharp and Li, 1987; de Miranda et al., 2000; Urrutia and Hurst, 2001), codon biases have also been found in fast-evolving genes (Blouin et al., 1998), a finding supported in this case by our phylogenetic profiling of both loci and codon positions captured by AHE (Fig. 2).

**Table 3**
GC content of nuclear genes commonly used in phylogenetic investigations of acanthomorph fishes, based on the alignment from Near et al. (2013)

| Gene | GC% | GC% 1st Codon | GC% 2nd Codon | GC% 3rd Codon |
|---|---|---|---|---|
| *glyt* | 52 | 54 | 37.4 | 64.7 |
| *Plag* | 57.6 | 54.6 | 44.1 | 74.2 |
| *ptr* | 52.3 | 41.9 | 33.5 | 71.4 |
| *sreb* | 57.3 | 52.7 | 43.9 | 75.3 |
| *RAG1* | 53.3 | 54.5 | 42 | 63.3 |
| *ZIC1* | 59.1 | 57.6 | 47 | 72.6 |
| *ENC1* | 53.7 | 56.5 | 38.7 | 65.9 |
| *Myh6* | 53.3 | 52.7 | 35.8 | 68.4 |
| *SH3PX3* | 55.2 | 54.1 | 35.5 | 76 |
| *tbr* | 60.7 | 53.7 | 60.8 | 68 |
| Average | 55.7 | 53.2 | 41.8 | 70 |

Our examination of third codon positions in the AHE dataset underscores that simply removing all third positions *a priori* will not fully mitigate the impact of saturation in large DNA sequence datasets. Other factors such as the redundancy of the genetic code for arginine and leucine that allows synonymous GC changes in both first and third codon positions (Palidwor et al., 2010) may also bias phylogenetic analyses. Correspondingly, our filtration approach resulted in the omission of subsets of both first and third codon position partitions. As GC-rich genomic regions are found across all major vertebrate groups (Belle et al., 2002), data scrutiny and filtration will remain an increasingly relevant aspect of phylogenomic analyses.

### 4.2. Taxon sampling, filtration, transformation, and consistent inference

Increased taxon sampling can alleviate noise and bias by correctly polarizing substitutions across an inferred tree (Hillis, 1998; Heath et al., 2008). However, this benefit of additional taxon sampling is greater when the taxa sampled diverge near the node of interest (Townsend and Lopez-Giraldez, 2010). Therefore, in collecting this dataset we attempted to sample every major lineage in the candidate pool of lineages that could be sister Percomorpha. This strategy allowed us to maximize coverage of deep nodes in the phylogeny near the divergence of Percomorpha. Based on theory, additional sampling would be subject to diminishing returns relative to increased character sampling when attempting to resolve deep divergences in a tree characterized by deep nodes followed by long subtending branches (Townsend and Lopez-Giraldez, 2010 c.f., Fig. 4). Given previous phylogenetic studies, the distribution of internodes is likely distributed towards the recent and away from the deeper portions of the tree within the crown of each of the lineages sampled in this study (Miya et al., 2003; Near et al., 2012b, 2013; Broughton et al., 2013; Dornburg et al., 2014a). It is possible that the addition of lineages not included in this study such as Barbourisiidae, Gibberichthyidae, and Hispidoberycidae could serve to break up long branches deep in the tree. However, exactly where the threshold between optimal taxon and character sampling lies in a partly resolved tree is difficult to determine at this point. While our analysis demonstrates that approaches aimed to target homoplasy such as RY coding and data filtration offer a potentially inexpensive alternative to additional taxon sampling, further development of theoretical advances that quantify the relative benefit of taxon versus gene sampling would be highly desirable for optimal experimental design.

As many portions of the Tree of Life are characterized by short internodes and long subtending branches (Venditti et al., 2010), the development of bioinformatic tools that can filter out site types from phylogenomic data that contribute high observed or unobserved homoplasy will continue to increase in importance. The need for a rational discriminatory approach that goes beyond crude heuristics of codon position is highlighted in our analyses by the observation that removal of all third codon positions from the AHE dataset resulted in no improvement in topological resolution (Fig. 4B). This lack of dissipation of homoplasy in response to removal of a significant portion of the contributing sites may be a counter-intuitive consequence of the large size of the GC-biased dataset. While adding more data can increase the number

of sites contributing to an accurate topology, adding more data will also increase the number of homoplasious sites (Philippe et al., 2011; Townsend et al., 2012; Lemmon and Lemmon, 2013). In contrast to the arbitrary removal of all third codon positions, the targeted removal of specific codon positions at specific genes, which were identified as problematic, resulted in the well-supported phylogenetic resolution of a clade containing Holocentridae, Rondeletiidae, Cetomimidae, Berycidae, and Melamphaidae as sister lineage of Percomorpha (Fig. 4C). An identical phylogenetic topology was observed in the analysis of the concatenated amino acid sequences (Fig. 4) and in a likelihood analysis that recoded third positions into purines (R) and pyrimidines (Y; Fig. 4D). Both of these approaches are expected to mitigate the impact of compositional bias in phylogenomic inference, thereby lending support to results obtained using our dataset filtration approach (Betancur-R et al., 2013b; Cox et al., 2014; Li et al., 2014).

Reducing compositional heterogeneity through the use of recoding procedures such as RY coding (Phillips and Penny, 2003; Phillips et al., 2004) has successfully reconciled topological incongruence in lineages as diverse as flatfishes and early land plants (Betancur-R et al., 2013b; Li et al., 2014). This reconciliation is encouraging and corresponds with simulations that suggest the potential of RY coding to greatly increase phylogenetic accuracy in the presence of compositional heterogeneity (Ishikawa et al., 2012). Nevertheless, RY coding may not be sufficient to suppress the influence of homoplasy on phylogenetic inferences in more divergent sequences or under conditions such as a bias in the ratio of transitions to transversions (Ishikawa et al., 2012). RY coding is also unlikely to mitigate homoplasious sites arising from processes that do not produce a compositional bias. In the case of our dataset, fast rates and compositional bias appear to be tightly linked; both dataset filtration and RY coding remove the potential for characters to mislead based on their evolution at a rate high enough that it leads to phylogenetic noise. Assuming the probability of resolution can be accurately quantified, dataset filtration could in principle mitigate the problem of homoplasy across a large range of datasets (Naylor and Brown, 1997), and is not limited to specific cases of base-pair heterogeneity. Filtration offers a tool that can be integrated with models of compositional heterogeneity to better achieve accurate phylogenetic resolution and congruence between competing hypotheses. An accurate filtration approach also provides the means to integrate datasets from multiple genomic regions that may vary in signal and noise content such as on non-coding ultraconserved elements (UCEs) (Crawford et al., 2012a; Faircloth et al., 2012; McCormack et al., 2012) or RADSeq datasets. UCEs are generally more AT rich (Duret and Arndt, 2008; Romiguier et al., 2013); however, informativeness criteria for fragment lengths at different timescales for this type of phylogenomic data has not been investigated. Comparing topological results between data filtration methods and data acquisition methods represents an exciting and potentially fruitful avenue of research that, when combined with other approaches such as assessments of model performance (Brown, 2014b, 2014a), will aid in our stabilization of a Genomic Tree of Life.

### 4.3. Sister to Percomorpha? A phylogenomic perspective

The staggering morphological diversity of percomorph teleosts combined with the scant fossil record of their early diverging lineages (Patterson, 1993), has complicated the identification of the sister lineage of Percomorpha. A range of phylogenetic hypotheses based on analyses of morphology (Stiassny and Moore, 1992; Johnson and Patterson, 1993; Moore, 1993b, 1993a; Baldwin and Johnson, 1995) and molecular data (Miya et al., 2005; Betancur-R et al., 2014; Near et al., 2013) have been proposed. Our resolution of the sister lineage to Percomorpha is congruent with early mitogenomic work (Miya et al., 2003; but see Miya et al., 2005) and is strongly supported both in the phylogenomic analyses of the DNA sequences filtered using PI based metrics, RY coding, and in the concatenated amino acid alignments (Figs. 3C & D and 5). It is important to note that our analyses do not span the entire spectrum of phylogenomic methods and there are other factors that can drive topological incongruence (Brown, 2014b, 2014a). However, our analyses highlight uncertainty in our understanding of phenotypic and ecological diversification across the fish lineages sampled. For example, Cetomimidae, Melamphaidae, and Rondeletiidae all inhabit deep-sea ecosystems (Paxton et al., 2001; Merrett and Moore, 2005; Bartow, 2011). The resolution of Holocentridae as the sister lineage of a clade that includes these species and Berycidae suggests either independent transitions to deep-sea habitats, or a shift to shallower (epipelagic < 200 m) habitats in Holocentridae and some species of Berycidae. Transitions between deep-sea and marine habitats in the photic zone are not unusual across the species in this study. For example, Berycidae are primarily slope and shelf dwelling fishes that occupy a range of depths commonly spanning 50–700 m (Dürr and González, 2002), with juveniles often occurring in shallow near-shore waters and even estuaries (Mundy, 1990; Smith, 2000). Transitions between deep-sea and shallow marine habitats are not limited to this clade, and shifts in depth ranges spanning hundreds, if not thousands of meters are observed across numerous lineages in this study. Increased taxon sampling in future studies of the phylogenetics and timing of diversification in this part of the teleost Tree of Life offers an exciting perspective from which to investigate the temporal correlations between habitat usage and geologic processes such as Cenozoic deep-sea anoxic events (Thomas, 2007), in generating and maintaining biodiversity in deep-sea fish communities.

## 5. Conclusion

Here we have shown that combining multiple approaches to dataset scrutiny can mitigate the impact of homoplasy in a phylogenomic dataset. We demonstrated that despite being conserved enough to identify across all gnathostomes, the loci captured by AHE exhibit high levels of nucleotide convergence characteristic of loci with a GC bias at the third codon position (GC3) at this taxonomic scale. Accommodating this bias in the phylogenetic analyses of the AHE dataset results in a phylogeny where a clade comprised of Holocentridae (squirrelfishes), Berycidae (alfonsinos), Melamphaidae (bigscale fishes), Cetomimidae (flabby whalefishes), and Rondeletiidae (redmouth whalefishes) is resolved as the sister lineage of Percomorpha. Our study highlights the difficulties in reconciling incongruent topologies between phylogenomic analyses and published multi-locus datasets, and further suggests that GC3 is a feature of many phylogenetic "legacy markers." While the continual accumulation of sequence data on public databases, reduction in the cost of sequencing, and coordinated initiatives to resolve the phylogeny of major lineages have resulted in a 21st century phylogenetic renaissance (Wheat and Wahlberg, 2013; dos Reis et al., 2014; Graham Reynolds et al., 2014; Rochette et al., 2014), the promise of a fully assembled Tree of Life is an achievement that will require increasingly complex investigations of nodes that continue to elude resolution.

## Acknowledgements

This work was part of a symposium on the evolution of percomorph fishes at the 9th Indo Pacific Fish Conference in Okinawa, Japan during June of 2013. We thank Masaki Miya for his organizational help and motivation of this work and participants of the con-

ference for their valuable feedback. Brian Park and Sarah Federman provided valuable assistance with analyses. Mark Westneat at the Field Museum of Natural History, Andrew Bentley at the University of Kansas Natural History Collection, and Katherine Maslenikov at the Burke Museum provided tissues for DNA sequencing. Gregory Watkins-Colwell provided assistance with specimen and tissue collections of fishes at the Yale Peabody Museum of Natural History. Steve Weston provided guidance on parallel processing in R and with openMPI. Sandra Emme provided assistance with DNA sequencing. This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center and by National Science Foundation awards to TJN (DEB-0716155 and ANT-0839007), EML (DEB 1120516), EML and ARL (IIP 1313554), and to AD and TJN (DEB-1011328). All data and source code generated for this study are available on Zenodo (DOI: 105281/zenodo.321684); PhyInformR is hosted on CRAN and github: carolinafishes.github.io.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.02.017.

## References

Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G., Harmon, L.J., 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. USA 106, 13410–13414.

Archie, J.W., 1989. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Biol. 38, 253–269.

Baldwin, C.C., Johnson, G.D., 1995. A larva of the Atlantic flashlight fish, *Kryptophanaron alfredi* (Beryciformes: Anomalopidae), with a comparison of beryciform and stephanoberyciform larvae. Bull. Mar. Sci. 56, 1–24.

Bartow, K.A., 2011. Taxonomy and Ecology of the Deep-Pelagic Fish Family Melamphaidae, With Emphasis on Interactions With a Mid-Ocean Ridge System. Florida Atlantic University.

Belle, Elise M.S., Smith, Nick, Eyre-Walker, Adam, 2002. Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. J. Mol. Evol. 55 (3), 356–363.

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., 2013a. The tree of life and a new classification of bony fishes. PLoS Curr. 5.

Betancur-R, R., Li, C., Munroe, T.A., Ballesteros, J.A., Ortí, G., 2013b. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol. 62, 763–785.

Betancur-R, R., Wiley, E.O., Miya, M., Lecointre, G., Bailly, N., Ortí, G., 2014. New and revised classification of bony fishes. Version 2. <http://www.deepfin.org/Classification_v2.htm%3E>. Electronic version accessed 24 Jun 2014.

Blouin, M.S., Yowell, C.A., Courtney, C.H., Dame, J.B., 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. Mol. Biol. Evol. 15, 1719–1727.

Brenner, S., Elgar, G., Sanford, R., Macrae, A., Venkatesh, B., Aparicio, S., 1993. Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. Nature 366, 265–268.

Broughton, R.E., Betancur-R, R., Li, C., Arratia, G., Ortí, G., 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. 5.

Brown, J.M., 2014a. Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. Syst. Biol. 63, 334–348.

Brown, J.M., 2014b. Predictive approaches to assessing the fit of evolutionary models. Syst. Biol. 63, 289–292.

Carpenter, S., Kitchell, J., Hodgson, J., Cochran, P., Elser, J., Elser, M., Lodge, D., Kretchmer, D., He, X., Von Ende, C., 1987. Regulation of lake primary productivity by food web structure. Ecology 68, 1863–1876.

Chanet, B., Guintard, C., Betti, E., Gallut, C., Dettaï, A., Lecointre, G., 2013. Evidence for a close phylogenetic relationship between the teleost orders Tetraodontiformes and Lophiiformes based on an analysis of soft anatomy. Cybium 37, 179–198.

Colgan, D.J., Zhang, C.G., Paxton, J.R., 2000. Phylogenetic investigations of the Stephanoberyciformes and Beryciformes, particularly whalefishes (Euteleostei: Cetomimidae), based on partial 12S rDNA and 16S rDNA sequences. Mol. Phylogenet. Evol. 17, 15–25.

Cox, C.J., Li, B., Foster, P.G., Embley, T.M., Civáň, P., 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63, 272–279.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Glenn, T.C., 2012a. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8, 783–786.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K., Glenn, T. C., 2012b. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8, 783–786.

Crawford, N.G., Parham, J.F., Sellas, A.B., Faircloth, B.C., Glenn, T.C., Papenfuss, T.J., Henderson, J.B., Hansen, M.H., Simison, W.B., 2015. A phylogenomic analysis of turtles. Mol. Phylogenet. Evol. 83, 250–257.

Crawley, S.S., Hilu, K.W., 2012. Caryophyllales: evaluating phylogenetic signal in trnK intron versus matK. J. System. Evol. 50, 387–410.

Dávalos, L.M., Perkins, S.L., 2008. Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. Genomics 91, 433–442.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.

de Miranda, A.B., Alvarez-Valin, F., Jabbari, K., Degrave, W.M., Bernardi, G., 2000. Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in *Mycobacterium tuberculosis* and *Mycobacterium leprae*. J. Mol. Evol. 50, 45–55.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.

Dornburg, A., Friedman, M., Near, T.J., 2015. Phylogenetic analysis of molecular and morphological data highlights uncertainty in the relationships of fossil and living species of Elopomorpha (Actinopterygii: Teleostei). Mol. Phylogenet. Evol. 89, 205–218.

Dornburg, A., Moore, J.A., Beaulieu, J.M., Eytan, R.I., Near, T.J., 2014a. The impact of shifts in marine biodiversity hotspots on patterns of range evolution: evidence from the Holocentridae (squirrelfishes and soldierfishes). Evolution 69, 146–161.

Dornburg, A., Santini, F., Alfaro, M.E., 2008. The influence of model averaging on clade posteriors: an example using the triggerfishes (Family Balistidae). Syst. Biol. 57, 905–919.

Dornburg, A., Townsend, J.P., Friedman, M., Near, T.J., 2014b. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. BMC Evol. Biol. 14, 169.

Dornburg, A., Fisk, J.N., Tamagnon, J., Townsend, J.P., 2016. PhyinformR: Phylogenetic experimantal design and Phylogenomic data exploration in R. BMC Evol. Biol. 16, 262.

dos Reis, M., Donoghue, P.C., Yang, Z., 2014. Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. Biol. Lett. 10, 20131003.

Doyle, V.P., Young, R.E., Naylor, G.J., Brown, J.M., 2015. Can we identify genes with increased phylogenetic reliability? Syst. Biol. 64, 824–837.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2010. Geneious v7.0.6. Available from: <http://www.geneious.com/>.

Duret, L., Arndt, P.F., 2008. The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet. 4, e1000071.

Dürr, J., González, J., 2002. Feeding habits of *Beryx splendens* and *Beryx decadactylus* (Berycidae) off the Canary Islands. Fish. Res. 54, 363–374.

Eschmeyer, W.N., Fricke, R. (Eds.), 2015. Catalog of Fishes electronic version (9 Sept 2015), California Academy of Sciences, San Francisco. <http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>.

Eytan, R.I., Evans, B.R., Dornburg, A., Lemmon, A.R., Lemmon, E.M., Wainwright, P.C., Near, T.J., 2015. Are 100 enough? Inferring acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. BMC Evol. Biol. 15, 113.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.

Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS ONE 8, e65923.

Forey, P.L., 1973. A revision of the elopiform fishes, fossil and recent. Bull. Br. Mus. (Natural History) Geol. 10 (Supplement), 1–222.

Foster, P.G., Hickey, D.A., 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48, 284–290.

Galtier, N., Daubin, V., 2008. Dealing with incongruence in phylogenomic analyses. Philos. Trans. Roy. Soc. B: Biol. Sci. 363, 4023–4029.

Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159, 907–911.

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. Phylogenet. Evol. 92, 140–146.

Goodman, M., 1967. Deciphering primate phylogeny from macromolecular specificities. Am. J. Phys. Anthropol. 26, 255–275.

Graham Reynolds, R., Niemiller, M.L., Revell, L.J., 2014. Toward a tree-of-life for the boas and pythons: multilocus species-level phylogeny with unprecedented taxon sampling. Mol. Phylogenet. Evol. 71, 201–213.

Gustafsson, C., Govindarajan, S., Minshull, J., 2004. Codon bias and heterologous protein expression. Trends Biotechnol. 22, 346–353.

Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. System. Evol. 46, 239–257.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.
Hilu, K.W., Black, C.M., Oza, D., 2014. Impact of gene molecular evolution on phylogenetic reconstruction: a case study in the rosids (Superorder Rosanae, Angiosperms). PLoS ONE 9, e99725.
Hundt, P.J., Iglésias, S.P., Hoey, A.S., Simons, A.M., 2014. A multilocus molecular phylogeny of combtooth blennies (Percomorpha: Blennioidei: Blenniidae): multiple invasions of intertidal habitats. Mol. Phylogenet. Evol. 70, 47–56.
Ishikawa, S.A., Inagaki, Y., Hashimoto, T., 2012. RY-coding and non-homogeneous models can ameliorate the maximum-likelihood inferences from nucleotide sequence data with parallel compositional heterogeneity. Evol. Bioinform. Online 8, 357.
Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y., Faircloth, B.C., Nabholz, B., Howard, J.T., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320–1331.
Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.
Johnson, G.D., Patterson, C., 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. Bull. Mar. Sci. 52, 554–626.
Killick, Rebecca, Eckley, Idris, 2014. Changepoint: an R package for changepoint analysis. J. Stat. Softw. 58 (3), 1–19.
Klopfstein, S., Kropf, C., Quicke, D.L., 2010. An evaluation of phylogenetic informativeness profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). Syst. Biol. 59, 226–241.
La Mesa, M., Eastman, J.T., Vacchi, M., 2004. The role of notothenioid fish in the food web of the Ross Sea shelf waters: a review. Polar Biol. 27, 321–338.
Lartillot, N., 2013. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. Mol. Biol. Evol. 30, 489–502.
Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109.
Lartillot, N., Rodrigue, N., Stubbs, D., Richer, J., 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. 62, 611–615.
Lautredou, A.-C., Motomura, H., Gallut, C., Ozouf-Costaz, C., Cruaud, C., Lecointre, G., Dettai, A., 2013. New nuclear markers and exploration of the relationships among Serraniformes (Acanthomorpha, Teleostei): the importance of working at multiple scales. Mol. Phylogenet. Evol. 67, 140–155.
Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.
Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121.
Li, B., Lopes, J.S., Foster, P.G., Embley, T.M., Cox, C.J., 2014. Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. Mol. Biol. Evol. 31, 1697–1709.
Lin, J., Chen, G., Gu, L., Shen, Y., Zheng, M., Zheng, W., Hu, X., Zhang, X., Qiu, Y., Liu, X., 2013. Phylogenetic affinity of tree shrews to Glires is attributed to fast evolution rate. Mol. Phylogenet. Evol. 71, 193–200.
Lopez-Giraldez, F., Townsend, J.P., 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol. Biol. 11.
Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24, 133–141.
McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22, 746–754.
Merrett, N., Moore, J., 2005. A new genus and species of deep demersal fish (Teleostei: Stephanoberycidae) from the tropical eastern North Atlantic. J. Fish Biol. 67, 1699–1710.
Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc 2010 (6), pdb-prot5448.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548.
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., Takahashi, M., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 26, 121–138.
Mooers, A.O., Holmes, E.C., 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. 15, 365–369.
Moore, J.A., 1993a. The Phylogeny and Evolution of the Trachichthyiformes (Teleostei: Percomorpha), with Comments on the intrarelationships of the Acanthomorpha Ph.D. Dissertation. Yale University, New Haven, CT.
Moore, J.A., 1993b. Phylogeny of the Trachichthyiformes (Teleostei: Percomorpha). Bull. Mar. Sci. 52, 114–136.
Morgan, C.C., Foster, P.G., Webb, A.E., Pisani, D., McInerney, J.O., O'Connell, M.J., 2013. Heterogeneous models place the root of the placental mammal phylogeny. Mol. Biol. Evol. 30, 2145–2156.
Mundy, B.C., 1990. Development of larvae and juveniles of the alfonsins, Beryx splendens and B. decadactylus (Berycidae, Beryciformes). Bull. Mar. Sci. 46, 257–273.
Naylor, G.J.P., Brown, W.M., 1997. Structural biology and phylogenetic estimation. Nature 388, 527–528.
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. 110, 12738–12743.
Near, T.J., Dornburg, A., Harrington, R.C., Oliveira, C., Pietsch, T.W., Thacker, C.E., Satoh, T.P., Katayama, E., Wainwright, P.C., Eastman, J.T., 2015. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. BMC Evol. Biol. 15, 109.
Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T., Zane, L., Fernandez, D.A., Jones, C.D., 2012a. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc. Natl. Acad. Sci. USA 109, 3434–3439.
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., 2012b. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. USA 109, 13698–13703.
Nelson, J.S., Grande, T.C., Wilson, M.V.H., 2016. Fishes of the World. John Wiley, Hoboken.
Palidwor, G.A., Perkins, T.J., Xia, X., 2010. A general model of codon bias due to GC mutational bias. PLoS ONE 5, e13431.
Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., Rossiter, S.J., 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502, 228–231.
Patterson, C., 1993. An overview of the early fossil record of acanthomorphs. Bull. Mar. Sci. 52, 29–59.
Paxton, J.R., Johnson, G.D., Trnski, T., 2001. Larvae and Juveniles of the Deepsea "whalefishes", Barbourisia and Rondeletia (Pisces: Beryciformes), with Comments on Relationships. Rec. Aust. Mus. 53, 407–425.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Worheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9.
Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.
Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171–185.
Pisani, D., 2004. Identifying and removing fast-evolving sites using compatibility analysis: an example from the Arthropoda. Syst. Biol. 53, 978–989.
Plotkin, J.B., Kudla, G., 2010. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12, 32–42.
Pond, S.L.K., Muse, S.V., 2005. HyPhy: Hypothesis Testing Using Phylogenies. Statistical Methods in Molecular Evolution. Springer, pp. 125–181.
Powell, J.R., Moriyama, E.N., 1997. Evolution of codon usage bias in Drosophila. Proc. Natl. Acad. Sci. 94, 7784–7790.
Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Moriarty Lemmon, E., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next generation DNA sequencing. Nature 526, 569–573.
Rhyne, A.L., Tlusty, M.F., Schofield, P.J., Kaufman, L., Morris Jr, J.A., Bruckner, A.W., 2012. Revealing the appetite of the marine aquarium fish trade: the volume and biodiversity of fish imported into the United States. PLoS ONE 7, e35808.
Rochette, N.C., Brochier-Armanet, C., Gouy, M., 2014. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol. Biol. Evol., mst272
Roje, D.M., 2010. Incorporating molecular phylogenetics with larval morphology while mitigating the effects of substitution saturation on phylogeny estimation: a new hypothesis of relationships for the flatfish family Pleuronectidae (Percomorpha: Pleuronectiformes). Mol. Phylogenet. Evol. 56, 586–600.
Rokas, A., Chatzimanolis, S., 2008. From Gene-Scale to Genome-Scale Phylogenetics: The Data Flood in, but the Challenges Remain. Phylogenomics. Springer, pp. 1–12.
Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., Douzery, E.J., 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol. Biol. Evol. 30, 2134–2144.
Rosenberg, A.A., Bolster, W.J., Alexander, K.E., Leavenworth, W.B., Cooper, A.B., McKenzie, M.G., 2005. The history of ocean resources: modeling cod biomass using historical records. Front. Ecol. Environ. 3, 78–84.
Rota-Stabelli, O., Lartillot, N., Philippe, H., Pisani, D., 2013. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. Syst. Biol. 62, 121–133.
Rutschmann, S., Matschiner, M., Damerau, M., Muschick, M., Lehmann, M.F., Hanel, R., Salzburger, W., 2011. Parallel ecological diversification in Antarctic notothenioid fishes as evidence for adaptive radiation. Mol. Ecol. 20, 4707–4721.
Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331.
Sanciangco, M.D., Carpenter, K.E., Betancur-R, R., 2015. Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). Mol. Phylogenet. Evol. 94, 565–576.
Santini, F., Carnevale, G., 2015. First multilocus and densely sampled timetree of trevallies, pompanos and allies (Carangoidei, Percomorpha) suggests a Cretaceous origin and Eocene radiation of a major clade of piscivores. Mol. Phylogenet. Evol. 83, 33–39.

Santini, F., Harmon, L.J., Carnevale, G., Alfaro, M.E., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Biol. 9, 164.
Santini, F., Sorenson, L., Marcroft, T., Dornburg, A., Alfaro, M.E., 2013. A multilocus molecular phylogeny of boxfishes (Aracanidae, Ostraciidae; Tetraodontiformes). Mol. Phylogenet. Evol. 66, 153–160.
Schuster, S.C., 2007. Next-generation sequencing transforms today's biology. Nature 200.
Seehausen, O., 2006. African cichlid fish: a model system in adaptive radiation research. Proc. Roy. Soc. B: Biol. Sci. 273, 1987–1998.
Seo, T.-K., 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25, 960–971.
Sethi, S.A., Branch, T.A., Watson, R., 2010. Global fishery development patterns are driven by profit but not trophic level. Proc. Natl. Acad. Sci. 107, 12163–12167.
Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol. 62, 551–563.
Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428, 717–723.
Sharp, P., Li, W.-H., 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4, 222–230.
Smith, K.A., 2000. Active and passive dispersal of Centroberyx affinis (Berycidae) and Gonorynchus greyi (Gonorynchidae) larvae on the Sydney shelf. Mar. Freshw. Res. 51, 229–234.
Smith, W.L., Wheeler, W.C., 2006. Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. J. Hered. 97, 206–217.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21), 2688–2690.
Stiassny, M.L.J., Moore, J.A., 1992. A review of the pelvic girdle of acanthomorph fishes, with comments on hypotheses of acanthomorph intrarelationships. Zool. J. Linn. Soc. 104, 209–242.
Su, Z., Wang, Z., Lopez-Giraldez, F., Townsend, J.P., 2014. The impact of incorporating molecular evolutionary model into predictions of phylogenetic signal and noise. Phylogenet. Phylogenom. System. 2, 11.
Thacker, C.E., Satoh, T.P., Katayama, E., Harrington, R.C., Eytan, R.I., Near, T.J., 2015. Molecular phylogeny of percomorpha resolves trichonotus as the sister lineage to gobioidei (Teleostei: Gobiiformes) and confirms the polyphyly of Trachinoidei. Mol. Phylogenet. Evol. 93, 172–179.
Thomas, E., 2007. Cenozoic mass extinctions in the deep sea: what perturbs the largest habitat on Earth? Geol. Soc. Am. 424, 1.
Townsend, J.P., 2007. Profiling phylogenetic informativeness. Syst. Biol. 56, 222–231.
Townsend, J.P., Leuenberger, C., 2011. Taxon sampling and the optimal rates of evolution for phylogenetic inference. Syst. Biol. 60, 358–365.
Townsend, J.P., Lopez-Giraldez, F., 2010. Optimal selection of gene and ingroup taxon sampling for resolving phylogenetic relationships. Syst. Biol. 59, 446–457.
Townsend, J.P., Su, Z., Tekle, Y.I., 2012. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. Syst. Biol. 61, 835–849.
Urrutia, A.O., Hurst, L.D., 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics 159, 1191–1199.
Venditti, C., Meade, A., Pagel, M., 2010. Phylogenies reveal new interpretation of speciation and the Red Queen. Nature 463, 349–352.
Wainwright, P.C., Smith, W.L., Price, S.A., Tang, K.L., Sparks, J.S., Ferry, L.A., Kuhn, K. L., Eytan, R.I., Near, T.J., 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in Labroidei and beyond. Syst. Biol. 61, 1001–1027.
Wan, X.-F., Xu, D., Kleinhofs, A., Zhou, J., 2004. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evol. Biol. 4, 19.
Wheat, C.W., Wahlberg, N., 2013. Phylogenomic insights into the Cambrian explosion, the colonization of land and the evolution of flight in Arthropoda. Syst. Biol. 62, 93–109.
Winemiller, K.O., 1990. Spatial and temporal variation in tropical fish trophic networks. Ecol. Monogr. 60, 331–367.
Wong, E.H.M., Smith, D.K., Rabadan, R., Peiris, M., Poon, L.L.M., 2010. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. BMC Evol. Biol. 10, 253.