Syst. Biol. 74(5):799–817, 2025

https://doi.org/10.1093/sysbio/syaf016
Advance Access Publication March 1, 2025

Phylogenomics of Marine Angelfishes: Diagnosing Sources of Systematic Discordance for an Iconic Reef Fish Family (F: Pomacanthidae)

Lauriane M. Baraf[1,2,*], Julia Y. Hung[1], and Peter F. Cowman[1,2,3]

[1]*College of Science and Engineering, James Cook University, 1 James Cook Dr, Townsville, QLD 4814, Australia*
[2]*Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, 1 James Cook Dr, Townsville, QLD 4814, Australia*
[3]*Biodiversity and Geosciences Program, Queensland Museum Tropics, 70-102 Flinders St, Townsville, QLD 4810, Australia*
[*]*Correspondence to be sent to: College of Science and Engineering, James Cook University, 1 James Cook Dr, Townsville, QLD 4814, Australia; E-mail: lauriane.baraf@my.jcu.edu.au.*

Received 6 May 2024; reviews returned 17 February 2025; accepted 25 February 2025
Associate Editor: Michael Alfaro

Abstract.—Genome-scale data sets are resolving challenging nodes across the tree of life. These data sets, however, characterized by inherent heterogeneity, often push traditional phylogenetic reconstruction methods to their limits. By integrating multiple phylogenomic approaches, we can identify the causes of topological discordance within genomic partitions while accounting for various sources of heterogeneity and systematic errors. Here we conduct the first in-depth analysis of discordance for the reef family Pomacanthidae (marine angelfishes) using target enrichment data of ~1,000 ultraconserved elements from 45 pomacanthid species. Our combined phylogenomic approach resolved the systematics of the family at the base of the pomacanthid tree. Despite this resolution, our analyses also highlight discordance in ancestral nodes associated with the paraphyletic *Centropyge* genus and monotypic *Pygoplites* lineage, and the impact of incomplete lineage sorting in the evolutionary history pomacanthids. Species network searches and model selection supported a reticulated evolutionary history suggesting 3 ancient gene flow events between ghost (or unsampled) lineages at the root of the Pomacanthidae tree and ancestors of *Genicanthus*, *Centropyge*, *Chaetodontoplus*, and *Pomacanthus* lineages. This study advances our understanding of diagnosing topological discordance in genome-scale phylogenies and provide an analytical pathway for limiting systematic errors. In the process of diagnosing discordance, we identify key evolutionary processes involved in the complex evolution of marine angelfishes. While often inconvenient in phylogenetic analyses, patterns of discordance can shed light on underlying biological and evolutionary processes that shape the evolution of biodiversity. [keyword: Phylogenomics; bioinformatics; discordance; reef fish; Pomacanthidae; UCEs.]

Advances in molecular high-throughput sequencing technologies have rapidly expanded the amount of genetic data available by several orders of magnitude, prompting an increased use of genome-scale data sets in phylogenetic studies (Lemmon and Lemmon 2013; Kumar et al. 2019). This has helped to build and improve upon the nodal resolution from previous single or multiple gene approaches of lineages across the Tree of Life (Delsuc et al. 2005; Dunn et al. 2008; Misof et al. 2014; Streicher and Wiens 2017; DiBattista et al. 2018; Chan et al. 2020). However, this increased capacity for phylogenomic inferences has highlighted important limitations of current phylogenetic reconstruction methods. Large genome-scale data are inherently associated with a higher level of heterogeneity across genomic regions that reflects a complex combination of evolutionary and biological processes driving the evolution of lineages (Bravo et al. 2019). This heterogenous signal can lead to conflict among gene tree topologies, whereby each represents its own, potentially distinct, evolutionary artifact. Topological discordance can arise from various biological processes such as incomplete lineage sorting (ILS), horizontal gene transfer, gene duplication or loss, hidden paralogy or hybridization (Nakhleh 2013; Steenwyk et al. 2023). It can also be a result of systematic errors such as flawed chimeric sequences, missing data, sequencing error, improper evolutionary modeling, methodological biases, errors in gene tree estimations, or long branch attraction uncertainty (Westover et al. 2013; Roch et al. 2019; Steenwyk et al. 2023). ILS and introgression are likely the 2 biological factors that have received the most attention from scientists aiming to bridge genomic variation and patterns of phylogenomic discordance (Edwards 2009). As a result, the ongoing development of phylogenomic methods has focused on integrating diverse sources of discordance, with few able to account for ILS, introgression, or their co-occurrence (Liu et al. 2009; Nakhleh 2013; Mirarab et al. 2014; Solís-Lemus and Ané 2016; Edelman et al. 2019; Kubatko and Chifman 2019). To account for variance in molecular data, a species tree must therefore be inferred from multiple single gene trees, which raises concerns on how adequate traditional species tree estimation methods developed for multi-locus approaches are for data sets on a whole or reduced genomic scale.

The widely used ML concatenation method using nonparametric bootstrap proportion to estimate branch support (CA-ML hereafter), has been questioned for oversimplifying the complexity of molecular evolution and biological processes at the root of phylogenetic discordance. One of the core issues being that estimation of node uncertainty is directly linked to the size of the data set as it is calculated based on the variance of resampled sites. While the degree of missing data was a relevant

criterion to measure stochastic error in multiple genes approaches, for large genome-scale data sets containing a higher number of sites relative to the number of taxa, it can result in overinflated bootstrap proportions and misleading phylogenetic inferences (Collins et al. 2005; Kubatko and Degnan 2007). Additionally, CA-ML ignores gene tree heterogeneity by assuming that all genes in the concatenated alignment evolved following a similar evolutionary history. An assumption that is often violated by the various biological processes previously outlined. Consequently, the CA-ML approach has often been combined with coalescent-based methods, which are statistically consistent under the multispecies coalescent (MSC) model and aim to take into account the heterogeneity caused by ILS (Pamilo and Nei 1988; Kubatko and Degnan 2007, 2007; Larget et al. 2010). However, coalescent-based approaches coupled with species tree inferences, known as "summary methods" assume that the pre-estimated gene trees are correct and are therefore, expectedly sensitive to gene tree estimation error (GTEE, Patel et al. 2013; Chou et al. 2015; Molloy and Warnow 2018), which can lead to erroneous species trees (Gatesy and Springer 2013).

Problems with species tree reconstruction stemming from GTEE are increasingly likely to occur with reduced genomic representation data set such as targeted (e.g., ultraconserved element or exon capture) or RADSeq captured data sets, which are composed of short alignments or short loci often including few informative sites. These data often produce many gene trees with low phylogenetic signal (Betancur et al. 2014; Singhal et al. 2017; Karin et al. 2020). To palliate the GTEE issue in coalescent-based methods while conserving evolutionary stochasticity, single-site-based coalescent methods like SVDquartets were developed (Bryant et al. 2012; Chifman and Kubatko 2014). While they resemble concatenated methods in their application, they estimate quartet trees instead of gene trees prior to inferring species trees. In general, simulation studies have shown that CA-ML and SVDquartets methods might infer more accurate species trees than some coalescent-based approaches when levels of GTEE are high (Chou et al. 2015; Molloy and Warnow 2018). In contrast, under low GTEE, coalescent-based methods generally always performed better regardless of the degree of ILS (Edwards et al. 2016; Mirarab et al. 2016; Molloy and Warnow 2018; Patel et al. 2013 but see; Kimball et al. 2013; McCormack et al. 2013). Despite these limitations, currently available methods continue to contribute to our understanding of lineage evolution, provided that they are selected judiciously, taking into account factors such as the accuracy of gene tree estimation (e.g., biased from long branch attraction), the degree of ILS, and the nature of the data (e.g., exons vs. introns, short loci).

Combining complimentary phylogenomic methods proved to be a valuable approach for elucidating problematic areas of the Tree of Life, as it helped disentangle the causes behind discordance in genome-scale data sets (MacGuigan and Near 2019; Meleshko et al. 2021; Morales-Briones et al. 2021; Owen and Miller 2022; Smith et al. 2022). As the most diverse group of vertebrates, fishes have long been of interest to evolutionary scientists aiming to understand the drivers of species diversification, especially among the highly speciose assemblages associated with global coral reefs. Increasingly, reef fish biologists have turned to phylogenomic data sets and methods to explore evolutionary, ecological, and systematic questions (Longo et al. 2017; DiBattista et al. 2018; Santaquiteria et al. 2021; Yi-Kai et al. 2021; Nash et al. 2022; Hughes et al. 2023). However, to the best of our knowledge, only 1 study to date has investigated phylogenomic discordance in a reef fish lineage (Hughes et al. 2023). Yet their immense diversity within the marine ecosystem presents a unique opportunity to explore the intricate interplay between environmental factors, biogeography, and speciation events. Thus, while phylogenomics have primarily been applied to identify areas of discordance within reef fish phylogenies, they can also help to link these discordance patterns to biological and speciation processes to further our understanding of reef fish lineage evolution.

With 90 recognized species (Fricke et al. 2023), marine angelfishes (F: Pomacanthidae) represent a comparatively small family of reef fish species. Their vibrant and unique markings have earned them the status of iconic reef inhabitants worldwide, propelling them to become one of the most exported reef fish family in the marine ornamental fish trade (Wabnitz 2003). Some areas of the Pomacanthidae evolutionary history have been challenging to resolve using a multi-locus approach, whereby the divergence of specific clades remained poorly supported (e.g., *Holacanthus*, *Centropyge*). The combination of long and short branches in previous phylogenies (Gaither et al. 2014; Frédérich et al. 2017; Baraf et al. 2019) for the family suggests successive rapid speciation events with increased chances of ILS and GTEE. Pomacanthids are also known for their propensity to hybridize where at least 48% of species have been reportedly involved in hybridization events (Tea et al. 2020). While ongoing and sustained gene flow between some contemporary pomacanthid species is undoubtedly complicating species delimitation processes, it is not known if past introgression events might also be obscuring the resolution along the backbone of the family's tree.

In this study, we apply multiple phylogenomic methods to a targeted capture data set of ultraconserved elements (UCEs) to explore the causes behind unresolved areas of Pomacanthidae evolutionary history. We examine topological congruence among species trees inferred from CA-ML and summary methods. Finally, to evaluate the potential for ancient introgression, we estimated the proportion of introgressed loci using gene tree triplets and inferred phylogenetic network inferences with maximum pseudo-likelihood (MPL) under the MSC model. This study furthers our understanding and ability to diagnose causes of discordance in genome-scale

phylogenies and provides an analytical pathway for limiting systematic errors. It also helps to identify key evolutionary processes that shaped the complex evolution of marine angelfishes and provides insights into underlying processes shaping reef fish diversification.

MATERIALS AND METHODS

Sample Collection, Library Preparation, and Target Capture of UCEs

The final phylogenomic reconstruction included 45 species (Supplementary Table S3) of the 90 nominal pomacanthids within the family Pomacanthidae. Tissue samples (fin clips or muscle) were preserved in 80–100% ethanol and stored at −20 °C prior to extraction. DNA was extracted using DNEasy Blood and Tissue kits (Qiagen, Hildren, Germany) following the manufacturer's instructions. Each sample was then quantified with a Qubit 2.0 Fluorometer using Qubit dsDNA BR Assay kits and Nanodrop spectrophotometer (Thermo Fisher Scientific Inc.) to measure DNA quality. DNA extracts were shipped to Arbor Bioscience (US) for library preparation and UCE targeted enrichment following MyBaits v.5.0 protocol using the acanthomorph probe set (Alfaro et al. 2018).

UCEs Extraction, Correction, Phasing, and Assembly

Demultiplexed reads were trimmed with the illumiprocessor wrapper program (Faircloth 2013) using the trimmomatic package (Bolger et al. 2014) and default parameters. Trimmed reads were then assembled into contigs using the SPAdes v. 3.10 (Bankevich et al. 2012) in the Phyluce v.1.7.1 program (Faircloth 2016). Following the Phyluce workflows (https://phyluce.readthedocs.io/en/latest/daily-use/daily-use-4-workflows.html), we first mapped assembled contigs and exploded fasta files respectively to the trimmed raw reads, marked duplicates, and calculated contig coverage for all contigs and UCE matched contigs. Resulting BAM files were then used to correct assembled contigs by filtering out low-depth and low-quality base calls. Corrected contigs were then scanned for UCE loci matches against the fish-uce-1k-probes.fasta UCE probe set (Supplementary Fig. S1). UCE contigs were aligned with the Multiple **Alignment** using Fast Fourier Transform (MAFFT) aligner (Katoh et al. 2002) and because most time-calibrated phylogenies estimated a recent emergence of Pomacanthidae (between 30 and 70 Ma), alignments were edge-trimmed using *phyluce_align_seqcap_align*. The final alignment matrix included at least 95% of all samples in each locus (Supplementary Table S1).

To maximize per-locus information, we phased UCE data to account for variation from heterozygous sites, which is lost in contig assemblies. We first mapped the corrected exploded fasta files to the trimmed raw reads and used the resulting BAM files to phase alleles into new assemblies representing 0, 1, and combined 0/1 haplotypes following the workflows outlined in Andermann et al. (2019). Haplotype assemblies were then implemented back into the Phyluce pipeline to produce three 95% complete alignment matrices [0; 1; 0/1].

Phylogenomic Analyses

Alignment matrices for unphased and phased data were processed in IQ-TREE v.2.2.2 (Nguyen et al. 2015; Minh et al. 2020) and used to generate a concatenation-based species tree. We first ran ModelFinder (Kalyaanamoorthy et al. 2017) implemented in IQ-TREE using the option -m MFP + MERGE to find the best-fitted model for our data and allowed for partition merging to reduce the potential for model overfitting. We then inferred a reference maximum-likelihood (ML) species tree by conducting an Shimodaira-Hasegawa approximate likelihood-ratio (SH-aLRT) branch test and measuring branch support with 1,000 ultrafast bootstrap replicates each, using the general time reversible (GTR) + G base substitution model and edge-linked partition model (Guindon et al. 2010; Anisimova et al. 2011; Hoang et al. 2017).

To verify the topology of the reference tree generated by IQ-TREE, we inferred a ML tree in RAxML v.8.2.12. We first subsetted the corrected unphased data set to keep only 1 sample with the highest number of UCE contigs per species before partitioning UCE loci present in the concatenated alignment (95% taxon completeness) following a sliding-window site characteristics method (Tagliacollo and Lanfear 2018). This entropy-based method aims to isolate the conserved core region of UCE markers from their variable flanking regions by dividing each locus into 3 data blocks. The resulting configuration file was used in PartitionFinder2 (Lanfear et al. 2016) to run a rapid bootstrap analysis for best scoring ML tree for 1,000 bootstrap replicates using the best fitting partitioning model selected from the best 10% partition schemes, GTR + G substitution model, and using the fast relaxed clustering search algorithm. To test if the resolution of the CA-ML tree could be improved by identifying and removing influential genes, we followed the IQ-TREE Tutorial on the website workshop's page (at http://www.iqtree.org/workshop/molevol2022). We conducted a partitioned model analysis in IQ-TREE using the same alignments and partitioning scheme than those used to infer the RAxML tree. Best partitioning scheme was then selected by PartitionFinder2 by merging partitions when improving fitness to reduce the potential for over-parameterization. Finally, we concatenated the trees from single and partitioned models before conducting topology and approximately unbiased tests for 1,000 bootstrap replicates using the best partitioning scheme. Subsequent comparisons between topologies inferred from single and partitioned models were conducted to identify genes that might be driving variance in phylogenetic signals. Results were filtered at 3 different levels

to examine variations in concordance factor (CF) values and topologies with the gradual removal of influential partitions linked to increasing log-likelihood difference ($\Delta \log L$) between the 2 topologies.

To account for ILS and compare topologies from concatenation-based and summary methods phylogenies, we first inferred a species tree from a set of gene trees under the MSC model in ASTRAL v.5.7.3 (Zhang et al. 2018). We generated individual gene trees for each UCE locus using partitioned ML analysis in IQ-TREE. To improve species tree estimation we collapsed all nodes with bootstrap values ≤ 30 with newick utilities v.1.6 and removed long branches from gene trees with TreeShrink (Mai and Mirarab 2018; Zhang et al. 2018). We then reran IQ-TREE on the new alignments and collapsed all nodes with bootstrap values ≤ 30 a second time. Resulting trees were inputted in ASTRAL twice, to measure branch support as local posterior probability and with the option -t10 to test for polytomies (Sayyari and Mirarab 2018). Finally, the species tree was rerooted in FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and we ran a concordance factor analysis in IQ-TREE to compare values with those of the reference tree. To identify if resolution could be improved by removing rogue taxa, we subsequently ran the RogueNaRok algorithm (Aberer et al. 2013) on the tree obtained with the Accurate Species TRee ALgorithm (ASTRAL-III).

Concordance Factor Estimations, Anomaly Zone, and Alternative Topologies

To conduct concordance factor estimations, we generated individual gene trees for each UCE locus using partitioned ML analysis and randomly sampled 100 quartets for each internal branch of the reference species tree in IQ-TREE. Gene concordance factor (gCF) and site concordance factor (sCF) were estimated as the fraction of decisive gene trees and decisive alignment sites concordant for each branch, respectively. This will help to quantify and differentiate genuine discordant signals caused by evolutionary processes such as ILS or introgression from those induced stochastically. To investigate the most supported alternative topologies, we use the -df-tree option in IQ-TREE. We then assessed the relationship between gCF and sCF values, and branch lengths (BLs) in coalescent units (number of generations/effective population size) obtained from MSC analyses. Shapiro–Wilk test of normality was first conducted on each variable and Spearman's rank correlation coefficient was used to calculated correlation coefficient (R) and *P*-value. To determine the proportion of anomalous divergences, we calculated BL in coalescent units for each parent-descendant internode using the python script from Linkem et al. (2016) and the ASTRAL-III tree as an input. This approach calculates how many pairs of parent-child internodes fall under the anomaly zone (AZ) boundary as defined by the coalescent theory (Degnan and Rosenberg 2006), wherein the frequency of discordant gene tree is higher than the frequency of non-discordance gene trees.

Test of Introgression and Phylogenetic Network Estimations

Introgression signals were investigated using the Quantifying Introgression via Branch Lengths (QuIBL) python script (Edelman et al. 2019) to estimate proportions of introgressed loci as a potential source of gene tree discordance. The program compares the distribution of internal BLs for each triplet topologies, whereby signatures of ILS and introgression can be translated into a 1-distribution and 2-distribution model, respectively. Bayesian Information Criterion (BIC) scores (Schwarz 1978) are then calculated to evaluate the likelihood of these events. Two sets of gene trees were generated with IQ-TREE using a partitioned ML approach on a 100% complete matrix. Gene tree sets included 15 pomacanthid species from the *Chaetodontoplus*, *Pomacanthus*, *Holacanthus*, and *Pygoplites* genera, and 15 species from the *Centropyge*, *Paracentropyge*, *Genicanthus*, and *Apolemichthys* genera, respectively, both with 1 *Acanthurus* species as an outgroup. Each pomacanthid clade was represented by at least 1 species, which was selected for highest number of contigs or systematic interest. As recommended by authors (Edelman et al. 2019), parameters for likelihood precision threshold, number of steps, and scale factor were set to 0.01, 50, and 0.5, respectively.

We assessed the potential for reticulate or non-tree like evolutionary histories evolution by conducting phylogenetic species network estimations using the MPL approach implemented in PhyloNet v.3.8.2 (Than et al. 2008; Yu and Nakhleh 2015; Cao et al. 2023). To alleviate computational demands, we reduced our data set to 11 pomacanthid species with highest number of informative sites per clade, and 1 outgroup species of the family Acanthuridae. A total of 839 gene trees for each locus were generated using the RAxML.pl script (https://github.com/JuliaPhylo/PhyloNetworks.jl). Five independent network searches were conducted, allowing for 0 to 5 reticulation events, respectively. Each search included 5 runs where BLs and inheritance probabilities were optimized to calculate likelihood for every species network. We computed network likelihoods using the set of RAxML gene trees with the function *CalGTProb* (Yu et al. 2012). Pseudo-likelihood values for best scoring networks were plotted in R and visualize in IcyTree (Vaughan 2017). Additional species network analyses were conducted on another set of 11 randomly selected pomacanthid species to verify the results. Model selection was performed using the Akaike information criterion (Akaike 1973) and BIC (Schwarz 1978), where number of parameters was equal to number of BLs being estimated plus number of hybridization allowed following (Yu et al. 2012). We calculated Akaike information criterion (AIC) weight scores to estimate conditional probabilities of each model following the equation by Wagenmakers and Farrell (2004). Analytical workflow is summarized in Fig. 1.

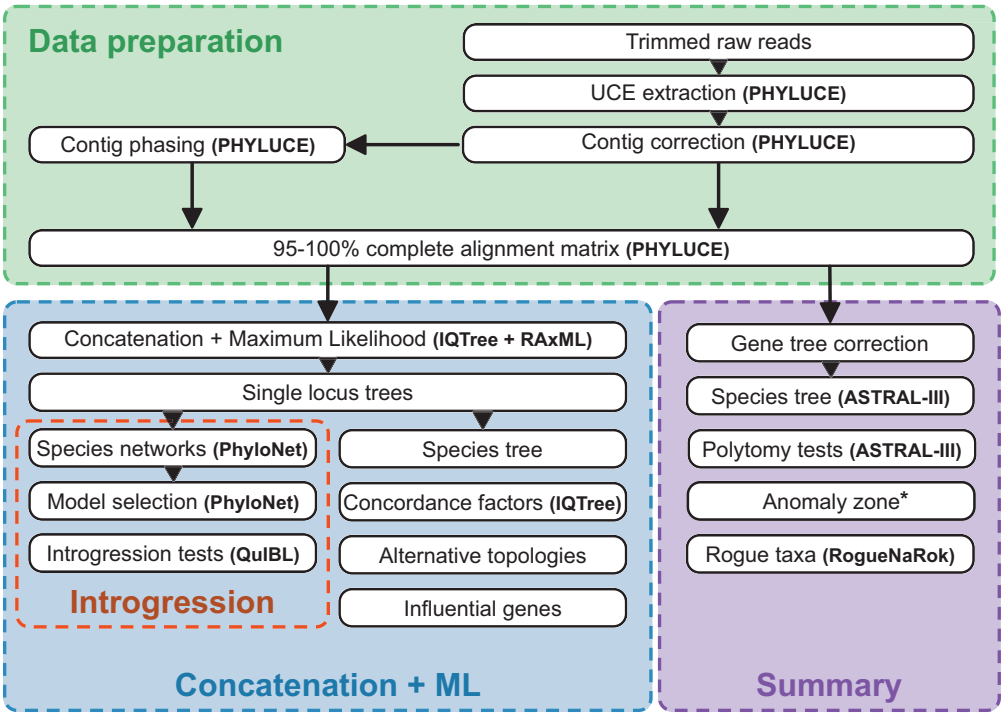

FIGURE 1. Summary of the analytical workflow used in this study. Bold letters in parentheses are bioinformatics softwares and asterisks are previously published studies: * following Linkem et al. (2016).

Divergence Time Estimation

The original data set was reduced to 1 sample per species (45 pomacanthid species) and filtered the concatenated alignment (95% completeness) to only include the first 50 most parsimony-informative sites that all shared the most common substitution model determined by ModelFinder in IQ-TREE. Parameters for Bayesian inference were set in BEAUTi v.2.6.7 (Drummond et al. 2012) as linked partitions across sites and trees, and unlinked molecular clocks. We applied an Hasegawa, Kishino and Yano (HKY) substitution model with empirical base frequencies and discrete gamma model including four rate categories G4 (HKY+F+G4), an optimized relaxed clock and Birth-Death model of speciation prior (Drummond et al. 2006) to each partition. We calibrated the phylogeny to an absolute time by age-constraining 4 nodes using available fossil-based priors representing most recent common ancestors of Nasinae/Acanthuridae (Near et al. 2013), Zanclidae-Nasinae/Acanthuridae (Near et al. 2012), Chaetodontidae (crown, Carnevale 2006) and Chaetodontidae/Pomacanthidae (Micklich et al. 2009) groupings following a log normal prior distribution (Supplementary Table S2). To reduce computational demand, we used the tree obtained with RAxML as a starting tree and imposed monophyly on all clades except for *Holacanthus* and *Pygoplites* lineages, which have sometimes been recovered as sister taxa in previous phylogenies. We conducted 3 independent Markov chain Monte Carlo (MCMC) runs of $400e^6$ generations each, sampling every 5,000th states in BEAST v.2.6.7 (Bouckaert et al. 2019) and assessed convergence in Tracer v.1.7.2 (Rambaut et al. 2018). Tree files were combined and resampled every $1e^6$ states with 20% burn-in in LogCombiner v.2.6.7. Final maximum clade credibility (MCC) tree was generated from 1,203 lineage trees in TreeAnnotator v.2.6.7 with median node height and highest posterior density (HPD) intervals.

RESULTS

After UCE contig correction and alignment trimming, concatenated alignment had a total length of 946,531 bp and contained 167,959 informative sites. Final 95% matrix (concatenated length 946,531 bp) included 880 alignments, representing a 25% increase compared to uncorrected trimmed alignments. Both uncorrected and corrected alignments had comparable amount of missing data (Supplementary Table S1). Average GC-content across UCE loci was relatively low, approximately 39%. This study generated the first UCE-based phylogenetic tree for the family Pomacanthidae, which was highly congruent with previously published phylogenies for the family. While it resolved the diverging order between the 2 clades, *Chaetodontoplus* and *Pomacanthus*, at the base of the tree (Fig. 2), it also helped identifying multiple discordance patterns causing low nodal support across the pomacanthid tree.

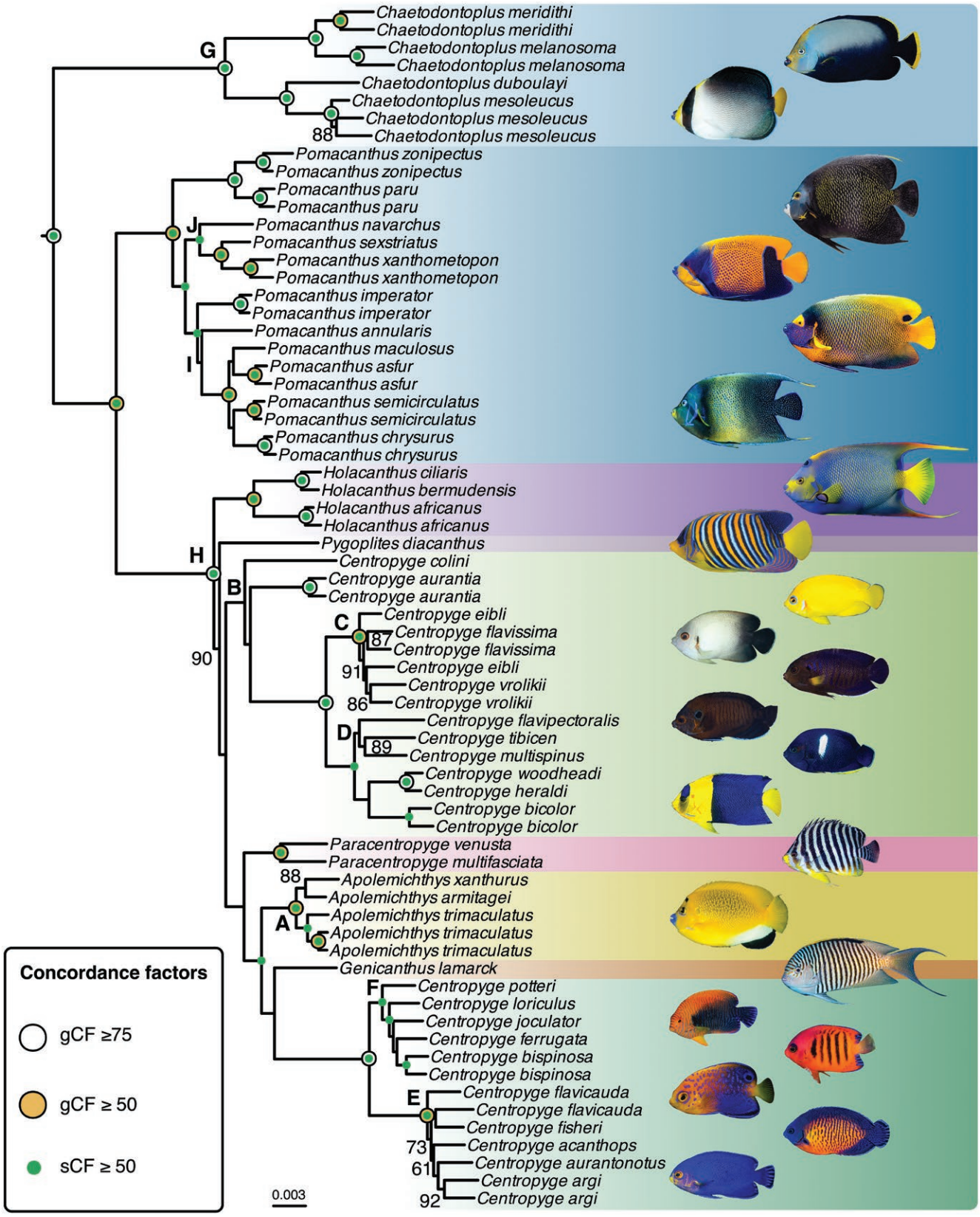

FIGURE 2. Species tree for 45 pomacanthid species generated by IQ-TREE v.2.2.2 from concatenated UCE alignments (95% taxon completeness, corrected contigs). Node numbers represent ultrafast bootstrap support values UFBoot < 95, gene (gCF) ≥ 50 and 75% and site (sCF) ≥ 50 concordance factors. Node letters represent nodes highlighted in Table 2. Photographs by François Libert, Jake Adams, Gregory McFall, and Klaus Stifel (CC BY-NY 2.0).

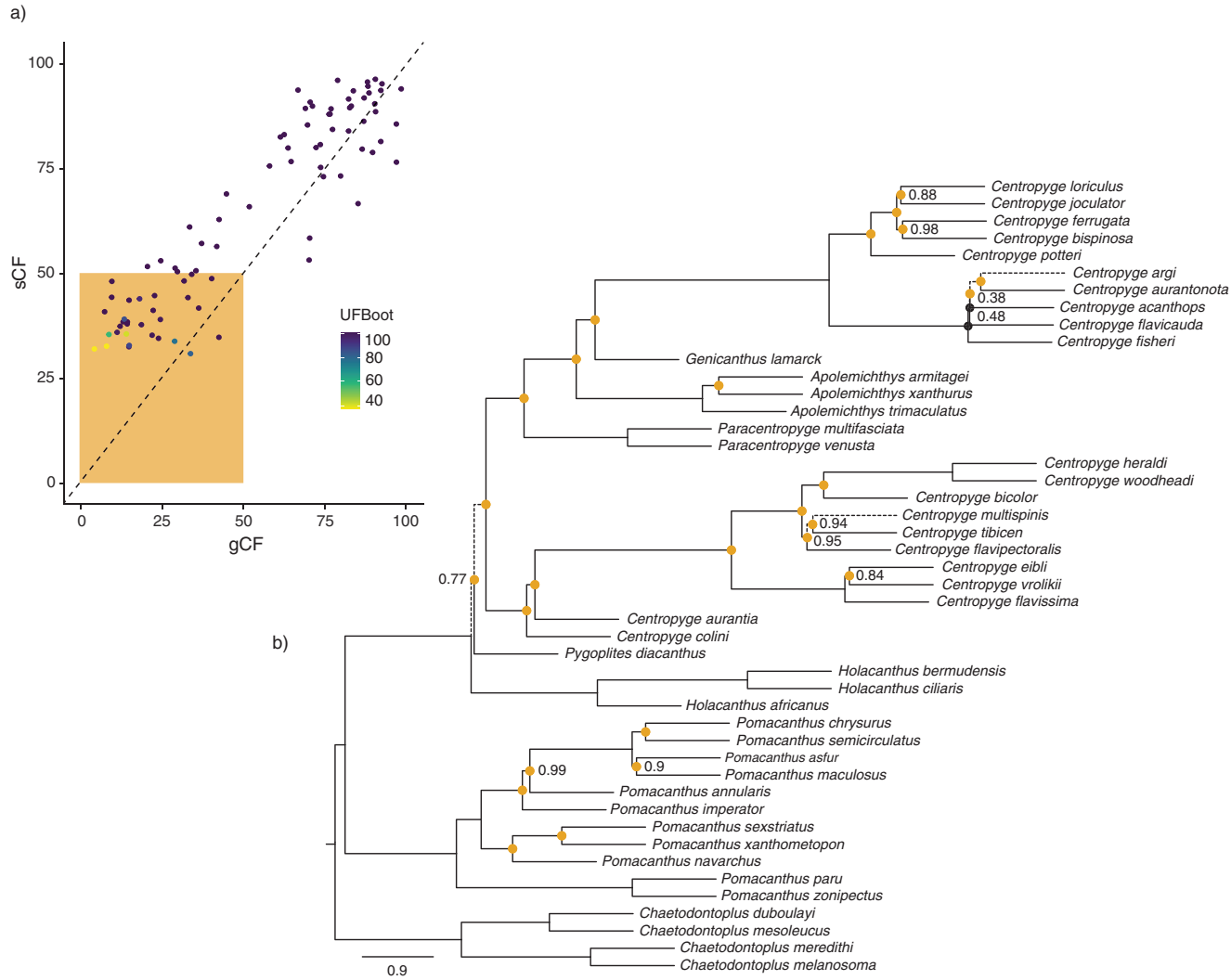

FIGURE 3. a) Relationship between concordance gene (gCF) and site (sCF) concordance factors, and UltraFast bootstrap values estimated with IQ-TREE across pomacanthid phylogenetic reconstruction. Box delimits nodes with combined sCF and gCF values ≤ 50. b) Phylogenetic tree generated by Astral-III tree with 30% threshold. Node labels are local posterior probability values ≤ 1, dashed branches show the 4 anomalous edge pairs, black dots show nodes which failed to reject the polytomy null hypothesis α and should be replaced by a true polytomy ($P \geq 0.05$) and orange dots correspond to nodes within the discordance box (sCF ≤ 50 and/or gCF ≤ 50) in the CA-ML tree.

Lineage Discordance and Topological Comparisons

The defined 50/50 zone for gCF/sCF values, which outlines a cluster of low concordance factor values (Fig. 3a) highlights conflicting phylogenetic signal from single locus trees, and decisive sites. Out of the 68 nodes (excluding outgroups), a total of 35 nodes (51%) were found in the "discordance box," of which 25 nodes (37%) had less than 50% support from both single locus trees and sites. Concordance factor values ranged from 3.06 to 49.45% for gCF and 32.5 to 68.35% for sCF (Fig. 3a and Supplementary Table S4) with both CF values coinciding with lower bootstrap values (Supplementary Fig. S2). BLs ranged from 0.0002 to 0.005 and showed a positive correlation with both gCF and sCF ($P < 0.001$, Supplementary Fig. S3). No taxa were identified as a rogue lineage by RogueNaRok, indicating that the overall resolution of the reference species tree couldn't be improved by removing 1 or multiple taxa. The 2 topologies inferred from single and partitioned models did not have significant differences in their log likelihood ($\Delta\log L = 21.109$, P-value > 0.4, Supplementary Table S5a). However, we identified 32 partitions with $\Delta\log L$ over 10 between the 2 topologies (Supplementary Table S5b), 9 of which had $\Delta\log L \geq 21$ (Supplementary Table S5b), overall including 83 core, 132 left flanking and 111 right flanking UCE regions. The removal of identified influential genetic regions ($N = 32, 18, 9$) following a 3 degree filtration ($10 \geq \Delta\log L \geq 45.8$, $13.8 \geq \Delta\log L \geq 45.8$, $21 \geq \Delta\log L \geq 45.8$) led to similar topologies and generally higher gene tree conflict but equal support from informative sites in comparison to the reference species tree (Supplementary Table S6). Similarly,

topologies inferred from phased UCE data were highly congruent, with discordant nodes not gaining additional support from CF values to the exception of a few *Centropyge* species (e.g., *C. multispinis*, *C. tibicien*, Supplementary Table S7).

The position of *Chaetodontoplus* as the first diverging lineage (gCF = 89, sCF = 80), followed by *Pomacanthus* (gCF = 68, sCF = 58) and *Holacanthus* (gCF = 78, sCF = 72) were strongly supported by both concordance factors (Fig. 2) and summary methods (local posterior probabilities [LPP] = 1, Fig. 3b). We detected 4 edge pairs in the AZ (Fig. 3b), 1 along the backbone of the tree involving *Pygoplites* and *Centropyge* C1, and the 3 others within *Centropyge* C1 and C2. The subsequent divergence of the monotypic genus *Pygoplites* was characterized by one of the shortest branches in our data set (BL = 0.03, in coalescent unit) with some of the lowest support from concordance factors (gCF = 13.1, sCF = 32.5, Fig. 2) and LPP = 0.77 (Fig. 3b). There was no clear phylogenetic signal (sCF ~ sDFs ~ 33%), reflecting scattered informative sites across consequently noisy and uninformative single locus trees. Alternative topologies DF1 and DF2 had *Pygoplites* diverging after or as sister taxa to *Holacanthus*. The first *Centropyge* clade to diverge, *Centropyge* C1, included species from the "*flavissima* complex" (hereafter referred to as C1.1) as a sister group to species of the *Centropyge* (*Centropyge*) subgenus (hereafter referred to as C1.2). Here, *C. colini* diverged first, followed by *C. aurantia*, both being reciprocally monophyletic to the crown *Centropyge* (*Centropyge*) lineage (Fig. 2). *Centropyge* C1 was also characterized by a short BL (BL = 0.14), which received low support from concordance factors (gCF = 12.4, sCF = 37.5). Most single locus trees did not match the topology of the species or DF1/DF2 trees (gene discordance factor due to polyphyly; gDFP = 74.4%) and all 3 topologies received roughly equal support from informative sites (sDFs ~30%). Alternative topologies had *Centropyge* C1 diverging between *Holacanthus* and *Pygoplites* or forming a clade with *Pygoplites* in basal position. The clade, however, was strongly supported by MSC analyses (LPP = 1, Fig. 3b).

Similar discordance patterns were observed for the divergence of *Paracentropyge* (gCF = 29.3, gDFP = 64.6, sCF = 48) and *Apolemichthys* (gCF = 37.7, gDFP = 52.7, sCF = 51.7) genera (Fig. 2 and Supplementary Table S4). Most single locus trees were discordant with the 3 possible clade arrangements for the branch but with most informative sites supporting the branch position of the species tree (Supplementary Table S4). Both clades had some of the longest BLs in the discordant data set, 0.45 and 0.62, respectively. They received strong support from the MSC analyses (LPP = 1, Fig. 3b). Alternative topologies for *Apolemichthys* had the clade diverging as sister to, or prior to, *Paracentropyge*. Alternative topologies for *Paracentropyge* had the clade diverging after *Pygoplites* or in a basal position to *Centropyge* C1.

The divergence of the *Genicanthus* lineage showed genuine discordant phylogenetic signal from single locus trees (gCF = 33.7, gDFP = 37.6, sCF = 41.1). Informative sites equally supported the species tree and third resolutions for this branch (sDF2 = 36.1). The latter comprised of a clade including *Genicanthus* and *Apolemichthys* genera, whereby the single sampled *Genicanthus* species in this study, *G. lamarck*, had a basal position to *Apolemichthys* species. The branching

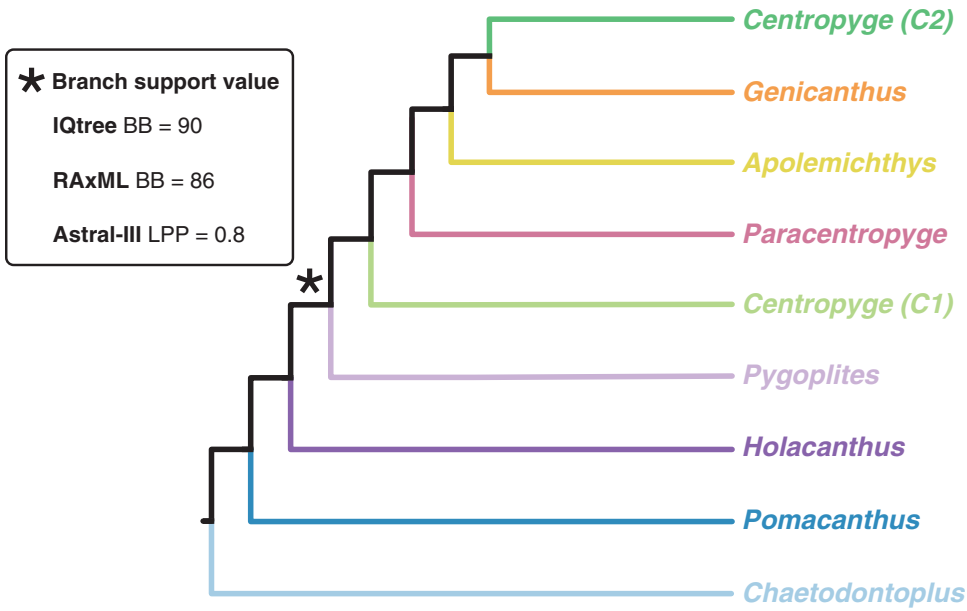

FIGURE 4. Convergence in species tree topologies for the Pomacanthidae family generated by concatenated-based methods using IQ-TREE and RAxML, and summary methods using ASTRAL-III for 9 pomacanthid clades. Asterisk indicates nodes across phylogenetic reconstructions with bootstrap values < 100 and local posterior values < 0.95 along the backbone of the pomacanthid tree. Data presented is for 95% complete alignment matrices and 45 pomacanthid species.

position observed in the reference tree was strongly supported by ASTRAL-III (LPP= 1, Fig. 3b).

Finally, the divergence of the second *Centropyge* clade C2, comprising of the *acanthops* "complex" (*C. acanthops*, *C. resplendens*, *C. argi*, and *C. aurantonotus*) and *C. fisheri* (hereafter referred to as C2.1) sister group to the rest of the *Centropyge* species (C2.2), was strongly supported by concordance factors (gCF = 75, sCF = 89, Fig. 2) and LPP = 1 (Fig. 3b).

Overall, topologies inferred by alignment-based (IQ-TREE and RAxML) and MSC summary methods were highly congruent (Figs. 3b and 4 and Supplementary Fig. S4). Apart from *Pygoplites*, all clades were strongly supported by bootstrap (BB > 95) and LPP = 1 values (Fig. 4).

Species-Level Discordance

All nodes within *Chaetodontoplus* and *Holacanthus* genera were strongly supported by concordance factors (Table 1 and Supplementary Table S4) and MSC analyses (LPP = 1, Fig. 3b). The divergence between *Pomacanthus* species located in the Eastern Pacific and Western Atlantic (*P. zonipectus* and *P. paru*), and those distributed across the Indo-Pacific region was also strongly supported (Fig. 2; Table 1 and Supplementary Table S4). There was varying degree of conflicting signal across single locus trees, which sometimes translated into uninformative locus trees (Table 1). For some species groupings within *Centropyge* C1 and C2, resolution was impacted by anomalous discordance (Fig. 3b and Table 1). Generally, there was high levels of gene tree conflict among *Centropyge* clades (Table 1 and Supplementary Table S4). Only the positions of *C. colini, C. potteri, C. loriculus, and C. bispinosa* within the reference species trees were supported by most informative sites (Fig. 2 and Table 1). In comparison, positions of *C. joculator* and *C. ferrugata* only received moderate support from informative sites (Table 1). There was genuine discordant signal behind the divergence of hybrid *A. armitagei* relative to its parent species *A. trimaculatus* and *A. xanthurus* (Table 1 and Supplementary Table S4).

Test of Introgression and Phylogenetic Network Estimations

One-distribution model (ILS only) was favored over 2-distribution model (non-ILS) across all 910 triplets tested for introgression signals using BLs. For the set of gene trees comprising of *Chaetodontoplus*, *Pomacanthus*, *Holacanthus,* and *Pygoplites* species, mixing proportions for a 2-distribution model (mixprop2) reached a

TABLE 1. Summary table of the discordance in pomacanthid lineages at the species level. Abbreviations and units: gene concordance factor (gCF, %), second resolution of the branch (DF1), third resolution of the branch (DF2), site concordance factor (sCF, %), local posterior probabilities (LPP) and branch lengths (BL, coalescent units). Lineages in bold were well supported at the intrageneric level and letters in parentheses represent labeled nodes in phylogenetic tree (Fig. 2)

Lineage	Species grouping	gCF	sCF	LPP	BL	Diagnosis
Apolemichthys (A)	*A. armitagei* (hybrid *A. xanthurus* × *A. trimaculatus*)	gCF = 31 gDF1 = 34	sCF = 43.4 sDF1 = 45.2	1	0.2	Genuine discordant signal with equal support for both sister pair *A. armitagei*/*A. xanthurus* and *A. armitagei*/*A. trimaculatus*
Centropyge C1 (B)	*C. colini* *C. aurantia*	gCF = 32 gCF = 22	sCF = 49.2 sCF ~ sDF1 ~ sDF2	1	0.2	Conflicting locus trees with most informative sites supporting positions in reference species tree for *C. colini*. Genuine discordant signal and uninformative locus trees for *C. aurantia*
Centropyge C1.1 (C)	*flavissima* "complex"	gCFs < 12	sCF ~ sDF1 ~ sDF2	≥ 0.84	0.04 > BL > 0.1	Noisy and uninformative single locus trees
Centropyge C1.2 (D)	*C. flavipectoralis* *C. multispinis*/*C. tibicien*	15 > gCFs > 20	sCF ~ sDF1 ~ sDF2	> 0.94	0.06	Anomaly zone
Centropyge C2.1 (E)	*acanthops* "complex"	4 > gCFs > 17	32 > sCFs > 45	≥ 0.88	0.004 > BL > 0.1	Anomaly zone
Centropyge C2.2 (F)	*C. potteri* *C. loriculus* *C. bispinosa*	25 > gCFs > 44	sCFs > 52	1	0.5 > BL > 0.8	Conflicting locus trees with most informative sites supporting positions in reference species tree
	C. joculator *C. ferrugata*	gCF = 11.3 gCF = 25.5	sCFs ~ 42	> 0.95	0.08 ≥ BL > 0.1	Noisy locus trees
***Chaetodontoplus* (G)**		gCFs > 69	sCFs > 79	1	2 > BL > 3.8	All nodes well supported
***Holacanthus* (H)**		gCFs > 70	sCFs > 76	1	2 > BL > 3.2	All nodes well supported
Pomacanthus (I)	*P. chrysurus*/*P. semicirculatus*/*P. asfur*/*P. maculosus* *P. annularis*	19 > gCFs > 34	35 > sCFs > 45 sCF ~ sDF1 ~ sDF2	0.9	0.08 > BL > 0.28	Noisy and uninformative locus trees
Pomacanthus (J)	*P. navarchus*/*P. sexstriatus*/*P. xanthometopon* *P. imperator*	19 < gCFs < 42	sCFs > 55.2	1	0.5 > BL > 1.4	Conflicting locus trees with most informative sites supporting positions in reference species tree

maximum of 0.62 (Supplementary Table S8). All triplet topologies with mixprop2 values above 0.50 were supported by less than 60 loci and showed higher BIC scores (ΔBIC > 10) than for a 1-distribution model (Supplementary Table S8). For the set of gene trees comprising of *Centropyge, Paracentropyge, Genicanthus,* and *Apolemichthys* species, mixing proportions (mixprop2) for a 2-distribution model reached a maximum of 0.45 (Supplementary Table S9). Eight triplet topologies had mixprop2 values above 0.40 were supported by most loci ($N > 598$) but showed higher BIC scores (ΔBIC > 10) than for a 1-distribution model (Supplementary Table S9).

Reduced 12-taxon data set for species network searches contained 969 alignments with over 0.95 proportion of taxa. Topology inferred by PhyloNet with zero reticulation was congruent with reference tree except for *Pygoplites* and *Holacanthus*, which formed a sister group (Fig. 5a). Model selection indicated that a reticulated evolutionary history was always a better fit better than bifurcating scheme (Table 2). Best model supported species network with 3 hybridization events (AIC = 29,774, Table 2), which was further supported by AIC weight scores ($w_3 = 1$, $w_{1:2} = 0$, $w_{4:5} = 0$). All ancient introgressive hybridization events involved ghost (or unsampled) lineages at the root of the Pomacanthidae tree (Fig. 5b). First introgressive event involved the MRCA of *Genicanthus* and *Centropyge* clade C2 (Fig. 5b), which were both recipient genomes from ghost lineages ($\gamma = 0.057$). Both *Chaetodontoplus* and *Pomacanthus* were recipient genomes from a common event ($\gamma = 0.57$), with *Pomacanthus* undergoing a separate, subsequent event ($\gamma = 0.75$). Phylogenetic networks obtained with the taxa set of 11 randomly selected pomacanthid species were consistent with those results but supported a ghost introgression event for *Chaetodontoplus* rather than *Pomacanthus* (Supplementary Fig. S5 and Supplementary Table S10).

Divergence Time Estimation

All pomacanthid clades were strongly supported by posterior probability values (PP ≥ 0.99) in the MCC topology generated with BEAST (Fig. 6). Crown age of the family was estimated at 27.38 Ma (95% HPD: 24.9–30.1 Ma) during the Oligocene epoch with most pomacanthid genera originating in the Miocene and Pliocene epochs (Fig. 6). For the most ancestral pomacanthid lineages (e.g., *Chaetodontoplus* and *Pomacanthus*), a timeframe (~38 Ma) for past ghost introgression events might have been possible between the Paleocene and Oligocene epochs, approximately ranging from 64.7 Ma (95% HPD: 61.14–68.45) to the root age of Pomacanthidae. For the MRCA between the *Genicanthus* genus and *Centropyge* clade 2, the window for ghost introgression extended until 5.09 Ma (95% HPD: 7.94–9.58 Ma). Increased lineage diversification occurred from the mid-Miocene (~13 Ma) to the Pleistocene (2.58–0 Ma) epochs (Fig. 6).

DISCUSSION

This study is the first in-depth diagnosis of the discordance in the evolutionary of the family Pomacanthidae using genome-scale data and the second for a reef fish family (Hughes et al. 2023). Out of the 26 nodes that showed high discordance from CF analyses (Figs. 2 and 3), only ten had ultrafast bootstrap values below 95%. This observation reinforces the widely acknowledged consensus that phylogenetic approaches aiming to estimate uncertainty of species relationships by using bootstrap support as a measure of sampling variance can fail to capture intrinsic variation in large data sets (Kumar et al. 2012). Overestimation of bootstrap values is commonly observed in concatenated alignments of large data sets where the resampling of the variance in the data set is bound to converge towards unrealistically high support values because of low sampling variance (Hillis and Bull 1993). This, however, does not undermine the value of bootstrap approximation for assessing branch support. Its measurement of statistical variance, combined with CFs informing on biological variance, makes them complementary approaches for exploring conflicting signals in phylogenomic data sets. Discordant nodes were in majority located in the tree space showing successive or simultaneous radiations over a short period of time (Fig. 2). Despite expected high levels of ILS, final normalized quartet score for the ASTRAL tree, which translate into the proportion of input gene tree quartet trees satisfied by the species tree, was of 84%. While this suggests that ILS only had a low impact on the overall recovery of pomacanthids evolutionary history, when accounting for GTEE violating MSC model assumptions, the quartet score is likely to be lower. As ILS depends on allelic variation and segregation patterns, the number of loci rejecting the MSC model might also increase with additional loci and expanded taxonomic coverage of the family (Jiang et al. 2020). Our analyses did not identify any rogue taxa and the removal of influential partitions that might have been driving some of the observed discordance did not lead to different or more supported topologies (Supplementary Tables S4 and S5). Similarly, haplotype trees inferred from phased UCE data did not show major topological differences nor improved CF values (Supplementary Table S7). There was higher support from informative sites for branching arrangements in the reference tree for some *Centropyge* species (*C. ferrugata, C. multispinis*, and *C. tibicien*). Gene tree conflict at those nodes might therefore be partly explained by heterozygous sites carrying different evolutionary histories, which can be expected in closely related and recently diverged species (Sota and Vogler 2003; Lischer et al. 2014), especially considering their propensity to hybridize.

Phylogenomics Resolve Systematics of Pomacanthidae

In regards of the backbone of the pomacanthid tree, all analyses strongly supported the initial divergence of

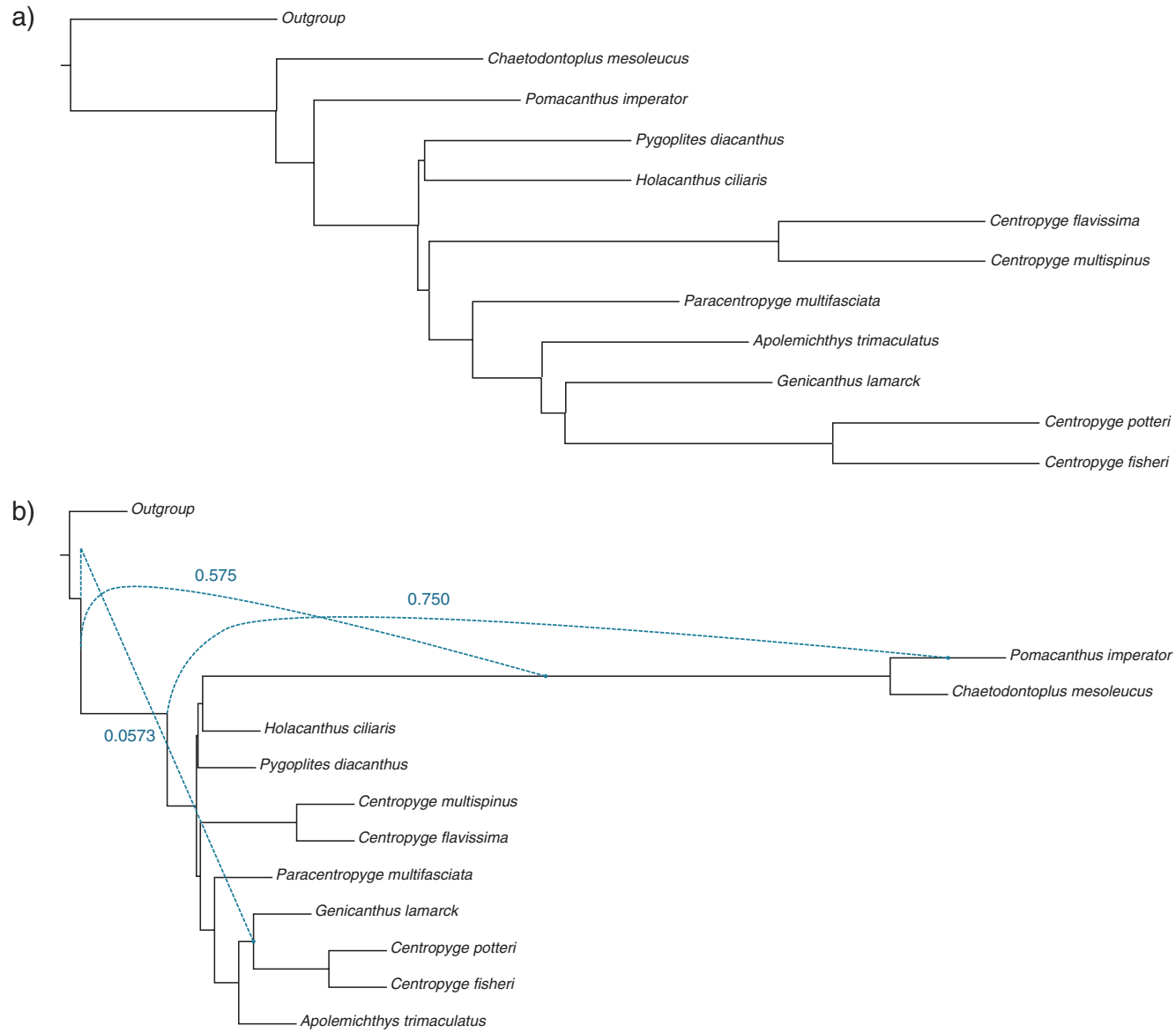

FIGURE 5. Phylogenetic species network with a) zero reticulation and b) 3 reticulation events and highest support from model selection inferred by PhyloNet. Dotted lines indicate reticulation edges and values are inheritance probabilities γ. BLs and inheritance probabilities were optimized to compute likelihood for each species network search.

TABLE 2 . Model selection between bifurcating species tree and 5 species network searches with maximum number of reticulations h_{max} set from 1 to 5. K represents number of parameters and bold line the model that best fitted the data

Topology	h_{max} allowed	# of inferred reticulations	ln(L)	k	# of loci	AIC	ΔAIC	BIC	ΔBIC
IQ-TREE	0	0	−1,710,742.4	20	969	3,421,505	3,391,730	3,421,492	3,391,736.48
Net 1	1	1	−16,346.473	22	969	32,714.95	2,940.31	32,699.82	2,944.3
Net 2	2	2	−15,881.82	24	969	31,787.64	2,013	31,770.52	2,015
Net 3	**3**	**3**	**−14,874.321**	**26**	**969**	**29,774.64**	**NA**	**29,755.52**	**NA**
Net 4	4	4	−16,1851.735	28	969	32,3731.5	293,957	32,3710.3	293,955
Net 5	5	5	−16,1701.106	30	969	32,3710.3	293,936	32,3409.1	293,653

the *Chaetodontoplus* lineage followed by the *Pomacanthus* genus (Fig. 4a–c). The 2 genera were often recovered as sister taxa in previous phylogenies with low support (Bellwood et al. 2004; Frédérich et al. 2017; Baraf et al. 2019). Potential overlap in genetic variation likely could not be fully captured by the limited genetic information of multigenic concatenation approaches, leading to the observed topological conflict between the 2

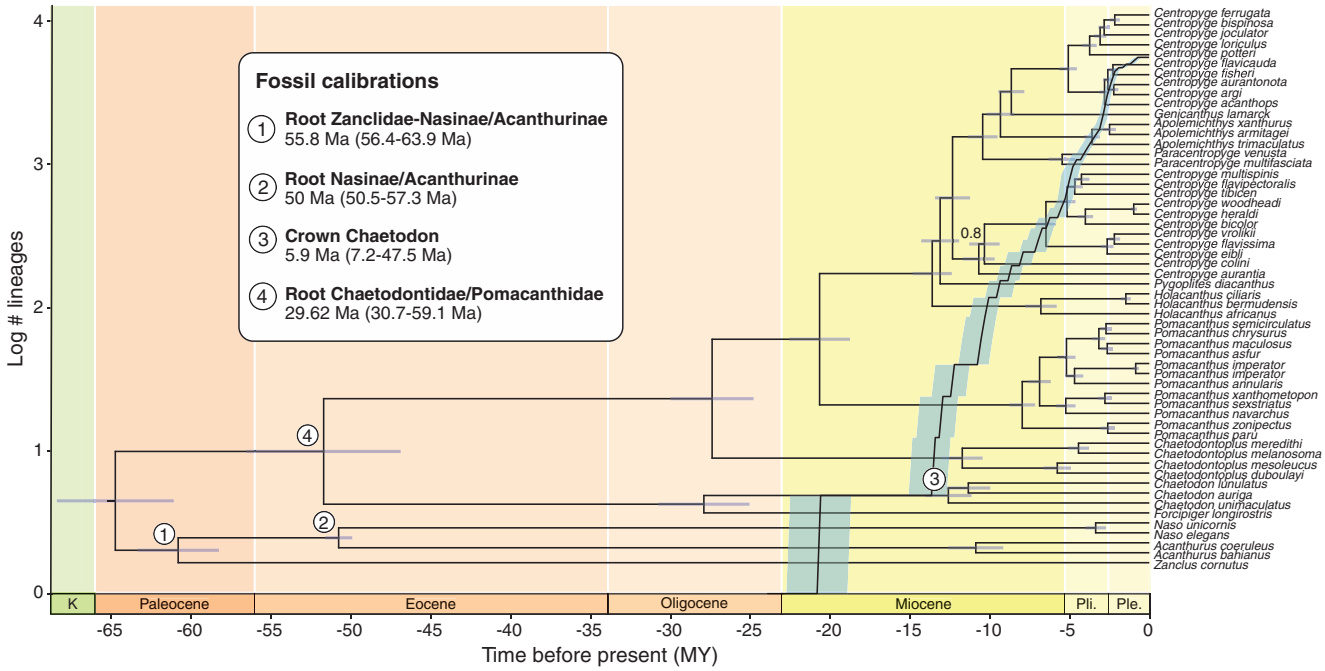

FIGURE 6. Fossil-calibrated chronogram for 45 pomacanthid species. Data presented were compiled from 3 independent MCMC of 400 million generations each conducted in BEAST. Horizontal node bars are 95% HPD intervals and node labels are posterior probability values under 0.99 calculated from MCC. Legend shows temporal constraints as minimal offset date in millions of years (5–95% HPD) on 4 nodes using fossil evidence and lognormal prior distribution. Lineage through-time plot shows pomacanthid lineage diversification with solid black line representing BEAST MCC and shaded area indicating 95% confidence intervals based on 1,000 randomly sampled post-burnin BEAST trees. *x*-axis represents time in million years before present and y-axis represents log-transformed number of pomacanthid lineages.

lineages. Phylogenomic analyses delineated 4 out of the 5 *Centropyge* clades that are usually recovered for this paraphyletic genus. Previous phylogenies had *C. colini* forming an additional fifth clade along with another deep dwelling species *C. narcosis* (not sampled in this study) and a species from the *Apolemichthys* genus, *A. arcuatus* (Baraf et al. 2019; Gaither et al. 2014 but see; Frédérich et al. 2017). Our phylogenetic reconstructions agree with findings of Frédérich et al. (2017) whereby the "fifth" clade (here only represented by *C. colini*) is first to diverge within the clade *Centropyge* C1. Although none of these nodes received strong support. Despite many discussions about species boundaries for *Centropyge* (Bowen et al. 2006; DiBattista et al. 2012; Gaither et al. 2014; Dibattista et al. 2016), present and previous studies show that the paraphyletic genus has well-delineated subgenera where similar species clusters are consistently recovered within the same clades. Therefore, while it may be difficult to untangle the inter-specific relationships within the *Centropyge* genera until full phylogenetic sampling can be achieved (59% taxonomic coverage of currently valid *Centropyge* species), there is a distinct level of phylogenetic delineation between subgenera that highlights the need for a taxonomic revision of this long acknowledged paraphyletic genus (Bellwood et al. 2004; Gaither et al. 2014). Our phylogenomic analyses further support that the subgenera *Xiphypops* (Jordan and Jordan 1922)—type species: *Centropyge fisheri* (Snyder 1904), *Paracentropyge* (Burgess 1991)—type species: *Centropyge multifasciata* (Smith and Radcliffe 1911) and *Centropyge* (Kaup 1860)—type species: *Centropyge tibicien* (Cuvier and Valenciennes 1831) should be resurrected and elevated to genus status.

Stochastic Error and Genuine Discordant Signal

The presence of short internal branches within and among generic lineages rendered the resolution of pomacanthid evolutionary history difficult (Fig. 2). Within the 50/50 discordance box (Fig. 3a), internal BLs ranged from 0.004 to 1.3 in coalescent units, which likely led to stochastic error from limited genetic information and increased gene tree discordance (Degnan and Rosenberg 2006; Philippe et al. 2017; Simion et al. 2020). While short internal branches are known to pose a challenge to resolving the Tree of Life, they also highlight compelling times of rapid, sometimes successive, radiations with increased chances of ILS and speciation with gene flow (Pamilo and Nei 1988; Song et al. 2012; Morales-Briones et al. 2021). Our phylogenomic analyses underlined a zone of uncertainty within the pomacanthid tree, characterized by a succession of short branches (e.g., *Pygoplites*, *Centropyge*, *Paracentropyge*, *Apolemichthys*, and *Genicanthus*), which might reflect a period of rapid diversification in the family evolutionary history.

There was a notable signature of noisy single locus trees whereby the most common resolution for both branches were not the 1 observed in the reference species tree (Supplementary Table S4). This suggested that the divergence of *Apolemichthys* and *Paracentropyge* lineages received low support because of conflicting phylogenetic signal in single locus trees and stochastic error from limited genetic information. Despite our corrected UCE alignments being relatively long (mean length of 1,020 bp), parsimony-informative sites only made up for 18% of all the base pairs (191 sites on average per locus), confirming that our data set contained limited phylogenetic information in comparison to other UCE loci based alignment matrices for other fish lineages (Alfaro et al. 2018; Faircloth et al. 2020). Nonetheless, it provided sufficient site support for their position in the reference species tree.

The topological conflict regarding the divergence of the *Genicanthus* genus resulted from genuine conflicting phylogenetic signal in underlying data (Supplementary Table S4). CFs equally supported *Genicanthus* diverging independently from the backbone of the tree or as sister genus to *Apolemichthys*. Larger data sets containing more or longer loci might help to obtain better resolution of the relationships between these 3 lineages (McCormack et al. 2013). Equally, it might confirm phylogenetically unresolvable conflicts due to low amount of interspecific genetic variation, which can be expected in rapid radiations (Morales-Briones et al. 2021; Scherz et al. 2022).

Alternative topologies revealed conflicting bipartitions separated by 1 or 2 nodes (Supplementary Table S4), pointing to the potential presence of ILS that might reflect large effective ancestral population size (relative to BL) for these 3 genera or successively rapid radiations.

Angelfishes Entering the AZ

For some of the shortest internal branches in the reference species tree, informative sites could not help diagnosing discordance and gene trees were mostly uninformative (Fig. 2 and Supplementary Table S4). Similarly, testing for multifurcating branching schemes did not help to improve the resolution of generic or specific relationships. We found that the challenging resolution of these extremely short branches was caused by extreme ILS, characterizing their presence in the AZ. The AZ can be define as areas within the species tree space where a majority of single gene trees are anomalous gene trees (AGTs) that do not match the topology of the species tree (Degnan and Rosenberg 2006). The boundaries of the AZ can be expanded by stochastic mutational variance (Huang and Knowles 2009). CA-ML methods, which ignore heterogeneity in evolutionary processes across branches (e.g., statistically inconsistent under the MSC model) and infer species trees based on the most common gene tree, are known to struggle in resolving branching arrangements within the AZ (Degnan and Rosenberg 2006; Huang and Knowles 2009; Mendes and Hahn 2018). When inferring species tree using coalescent methods, it is essential to assess for the presence of AGTs as edge pairs within the AZ boundary because they inherently violate the MSC model assumption of no GTEE.

We detected 4 edge/internode pairs in the AZ (Fig. 3b). Most of the anomalous discordance was found within the *Centropyge* subclades C1.2 and C2.1, which both involved extremely short and successive internal branches (Fig. 2). *Centropyge* subclade C2.1, also referred to as the *acanthops* "complex" (Gaither et al. 2014), showed some in-clade species rearrangements between CA-ML and summary methods inferred topologies (Figs. 2 and 3b). These branches had some of the lowest local posterior probability values (Fig. 3b). Considering how closely related species of the C1.2 and *acanthops* "complex" (C2.1) are, the species most likely have high degree of shared genetic variation, potentially from the retention of ancestral polymorphism. This is highly probable considering the multiple reports of captive hybrids, reflecting permeable reproductive barriers between species. For instance, *C. resplendens* (not sampled herein), endemic to the Ascension Island and usually part of the *acanthops* "complex," has been found to produce viable hybrids in captivity with *C. argi* and *C. fisheri* (Baensch and Tamaru 2009; Baensch 2017). Similarly, *C. acanthops* and *C. argi* were successfully captive bred despite unlikely hybrid crossing in the wild due to non-overlapping geographical ranges. Introgressive hybridization after secondary contact, therefore, might also explain the observed discordance for western Atlantic species *C. aurantonotus* and *C. argi*. In contrast, to the best of our knowledge, there have been no official reports of hybrids between the C1.2 trio comprising of *C. flavipectoralis*, *C. tibicien* and *C. multispinis*. Although, this does not exclude the possibility for introgressive hybridization.

The 2 other anomalous internodes involved the branches for *Pygoplites* and *Centropyge* C1 lineages (Fig. 3b). They received the least amount of support from local posterior values (Fig. 3b), where *Pygoplites* had 77% of quartet trees from gene trees present in the species tree. Alternative topologies for this branch had *Pygoplites* diverging prior to or as sister lineage to the *Holacanthus* genus. All 3 best branching arrangements were equally supported by informative sites, but none was the most common resolution for the branch (Supplementary Table S4). Following our discussion on *Apolemichthys* and *Paracentropyge* genera (see above), it is likely that the clustering of these 3 lineages might be a consequence from limited informative sites, which might be partly resolved by increased genomic information. Uninformative gene trees combined with the clear presence of GTEE for *Pygoplites* likely further hindered the resolution of branching patterns for lineages that appear to have rapidly diversified around the same time. Biological factors might also be influencing the position of *Pygoplites* in the pomacanthid tree. Firstly, due to the extant widespread geographical distribution of the monotypic *Pygoplites* genus, 1

can expect a higher genetic diversity with potential for recombination or important gene flow between populations. This would be coherent with the observation of genetic diversity and architecture in mitochondrial markers between individuals from the Pacific Ocean and Red Sea/Indian Ocean forms (Coleman et al. 2016). Replication across *Pygoplites* individuals collected from different geographical locations will be necessary to get a clearer picture of the population diversity and how it relates to other pomacanthid genera at the base of the tree. Previous phylogenetic reconstructions have occasionally recovered *Pygoplites* and *Holacanthus* genera as sister taxa (Bellwood et al. 2004; Frédérich et al. 2017). The resolution of the monotypic genus *Pygoplites* was impaired by anomalous discordance in CA-ML methods and GTEE on coalescent-based methods (Figs. 2 and 3b). In those cases, site-based methods might be a more suitable approach to resolving these types of branching arrangements (Chou et al. 2015; Molloy and Warnow 2018), which might highlight some degree of shared genetic material between the 2 genera and support rapid speciation from a common progenitor.

While branches within the AZ can sometimes be better represented as polytomies (Huang and Knowles 2009), ASTRAL polytomy test rejected a polytomy at these nodes. Overcoming the AZ might be possible with the addition of longer and more phylogenetically informative loci (Linkem et al. 2016). Equally, the proportion of branches in the AZ might expand with increasing taxon sampling as the addition of taxa will shorten internal branches. Simulation studies have shown that AGTs can occur in the presence of gene flow (Solís-Lemus et al. 2016; Long and Kubatko 2018).

Introgression, ILS, or Ghost of Lineage Past?

The Pomacanthidae phylogeny was characterized by a combination of significant gene tree discordance and extremely short branches, which might reflect simultaneous diversification events punctuated by numerous opportunities for introgressive hybridization. It was surprising that for a reef fish family known to extensively hybridize (Tea et al. 2020), no clear signal of introgression across pomacanthids was recovered. Results from the sequence-based approach implemented in QuIBL always supported an evolutionary scenario where ILS was responsible for gene tree discordance rather than caused by introgressive hybridization across pomacanthid species (Supplementary Tables S7 and S8). The area of the pomacanthid tree space dominated by successive short branches, therefore, suggests a series of rapid speciation events, which would have an increased likelihood of ILS occurring (Degnan and Rosenberg 2009; Bravo et al. 2019). While this reflects some degree of shared ancestral polymorphism in diverged lineages following speciation events (Avise and Robinson 2008), it does not exclude that past introgressive hybridization might have taken place across pomacanthid lineages. The number of SNPs extracted from UCE contigs was not sufficient to conduct robust site *D*-statistics. In some instances, statistical support (*P*-value) in ABBA-BABA tests could not be calculated because of too few sites, showing that the application of UCE-based data sets containing low polymorphic sites might be limited when exploring questions about introgression. Other studies showed that SNPs harvested from short and non-coding loci can hold less informativeness and statistical power when used to detect introgression signals (Leaché and Oaks 2017; Węcek et al. 2017; Barlow et al. 2018; Everson et al. 2019; Alda et al. 2021; Rancilhac et al. 2021). This issue is likely to be exacerbated in closely related or rapidly diversifying lineages, as well as in loci that are not under strong selection or have lower mutation rates. To determine whether these results reflect the "true" evolutionary history of pomacanthids or merely highlights the limitations of our data set, complementary analyses using a more informative data set will be necessary. A greater number of SNPs might be obtainable from less conservative genomic reduction approaches like restriction-site associated DNA (RAD) sequencing data or from whole genomes, potentially helping to refine future detection of introgression patterns within the family (Eaton and Ree 2013; Massatti et al. 2016; Near et al. 2018).

Model selection showed than any reticulated evolutionary history was a better fit than bifurcating tree model (Table 2). Best ranking species network had 3 hybridization events, any additional reticulation (h_{max} = 4–5) did not improve the log likelihood of the network (Table 2). All ancient introgression events involved ghosts (or unsampled) lineages at the root of the pomacanthid tree. It recovered *Pomacanthus* and the most recent common ancestor (MRCA) of *Chaetodontoplus/Pomacanthus* as recipient genomes with inheritance values above 50%, and to a lesser extent for the MRCA of *Genicanthus/Centropyge* C2 (γ = 5.7%, Fig. 5b). This suggests that *Pomacanthus* and *Chaetodontoplus* lineages inherited most of their genetic material from unsampled parental populations, which might explain some of the conflict observed in the branching arrangement of the 2 genera in previous phylogenies (Bellwood et al. 2004; Gaither et al. 2014; Frédérich et al. 2017; Baraf et al. 2019). Our phylogenetic network analyses always recovered the 2 genera as sister lineages in species networks (excl. Net 4). Furthermore, the low inheritance probability for the ancient introgression event on the MRCA of *Genicanthus/Centropyge* C2 suggests that the genuine discordant signal behind the position of *Genicanthus* (Fig. 2) most likely resulted from a combination of ILS and, to a lesser extent, introgressive hybridization. Timing of those events can be estimated between the MRCA of Acanthuridae and Pomacanthidae, approximately from 64.7 Ma (95% HPD: 61.14–68.45 Ma) and the crown age of Pomacanthidae around 27.38 Ma (95% HPD: 24.9–30.1 Ma), and the MRCA between *Genicanthus* and *Centropyge C2* around 8.66 Ma (95% HPD: 7.35–9.58 Ma). However, the use of different outgroups in species network inferences or discovery of additional fossil evidence for node calibration in Bayesian reconstructions

might lead to different introgression scenarios and time estimates.

Assessing the presence of introgression from midgroup or ghost lineages has gained significant interest. One reason being that they can confound the recovery of introgression signals in ABBA-BABA tests, leading to differing gene flow scenarios (Durand et al. 2011; Hibbins and Hahn 2022; Tricou et al. 2022). In such cases, *D*-statistic methods are unable to differentiate between introgression signals emitted by a sister ghost lineage from those of a sampled lineage, potentially resulting in the misidentification of donor genome for the same recipient genome (Eaton and Ree 2013; Eaton et al. 2015; Tricou et al. 2022). Other factors such as variation in population size or ancestral population structure can lead to false positives in *D*-statistic (Durand et al. 2011; Martin et al. 2015). These methods might increasingly struggle with progressively complex reticulated patterns for example, multiple introgression within 1 quartet (Elworth et al. 2018; Everson et al. 2019).

CONCLUSION

This study provides the first in-depth assessment of topological and gene tree discordance for Pomacanthidae. Despite limited phylogenetic information in our data set, a multi-phylogenomic approach help to resolve most of the family systematics and identify multiple sources of conflict across the pomacanthid tree. Some of the topological discordance caused by conflicting phylogenetic signal in single locus trees could be clarified by proportion of informative sites (e.g., *Apolemichthys* and *Paracentropyge*). Anomalous discordance was detected for *Pygoplites* and *Centropyge* C1, further highlighting the limitations of CA-ML methods that infer species tree from pre estimated, sometimes uninformative, gene trees. More moderate levels of ILS, associated with short branches and stochastic errors, likely caused most of the gene tree discordance across the pomacanthid tree. This was supported by introgression tests, which always favoured ILS over introgressive hybridization for the deeper short branches. Those observations suggest a time of rapid and successive speciation events in the evolutionary history of the family, whereby the short timeframe between each diverging clade did not allow for ancestral genetic variation to fully segregate into distinct lineages. This might also be indicative of a large effective ancestral population size. To distinguish between different scenarios of ILS, subsequent population genetics studies focusing on demographic history or examination of patterns of genetic variation across genomes will be required.

Resolving the systematics of closely related species is inherently challenging. In this study, it was likely accentuated by the short loci and limited informative characters in our data set. This issue is particularly pronounced when attempting to resolve divergence events characterized by uninformative gene trees and high GTEE (e.g., *Pygoplites diacanthus*), which would likely benefit from a site-based approach using a less reduced-representation library or whole genomes. The limitations of our data set were further underlined by the insufficient number of single-nucleotide polymorphisms (SNPs) retrieved from UCE contigs, which hindered our ability to conduct robust site-based phylogenetic reconstructions and *D*-statistics. Nevertheless, the present workflow, including site-based methods, is likely to assist future diagnoses of discordance in genomic data sets generated from less conservative genome reduction approaches or for more divergent taxa. Species networks searches, however, strongly supported a reticulated evolutionary history for the family, with the best model suggesting 3 reticulations involving ghost lineages at the root of the pomacanthid tree. Ancient introgressive hybridization for *Chaetodontoplus* and *Pomacanthus* might explain the topological conflict observed between the 2 genera in previous phylogenies. Moreover, it suggested that the genuine discordance signal in underlying data for *Genicanthus* was driven by past introgression.

Hence, the remaining incongruence in pomacanthid evolutionary history might not be resolvable from a phylogenetic standpoint, even with genome-scale data and especially for deeper nodes. Pomacanthid species potentially hold low genetic diversity between 1 another, a plausible narrative when considering that some vicariously isolated species can produce interspecific hybrids in captivity (e.g., *Centropyge*). There have even been recent reports of speculated cross-genera hybridization between *Pygoplites* and *Apolemichthys* (pers. comm. M. Wittenrich from PomaLabs). Future studies investigating introgression among pomacanthids should therefore consider population structure to be able to distinguish between evolutionary processes like introgression from shared genetic polymorphism from ancestral population.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.r4xgxd2n4.

ACKNOWLEDGMENTS

We thank Prof. David Bellwood, Prof. Brian Bowen, Dr. Michelle Gaither, Aspen Oudshoorn for providing or assisting with tissue sample collection. We also thank Prof. Cecile Ané and Hanaka Mera for insightful discussions. We are grateful to editors and 2 anonymous reviewers for their helpful comments and suggestions on the present manuscript.

FUNDING

This work was funded by the ARC DECRA Fellowships (DE170100516) and ARC Centre of

Excellence Programme (CE140100020) to P.F.C. and David Yellowlees Excellence in Research Award to L.M.B.

DATA AVAILABILITY

Raw reads for target enrichment of UCEs are available on the NCBI under BioProject PRJNA1101094. Alignments and tree files can be accessed on the Dryad repository https://datadryad.org/dataset/doi:10.5061/dryad.r4xgxd2n4. Scripts can be found on L.M.B Github https://github.com/ Lavarchus.

REFERENCES

Aberer A.J., Krompass D., Stamatakis A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Syst. Biol. 62:162–166. doi:10.1093/sysbio/sys078

Akaike H. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60:255–265. doi:10.1093/biomet/60.2.255

Alda F., Ludt W.B., Elías D.J., McMahan C.D., Chakrabarty P. 2021. Comparing ultraconserved elements and exons for phylogenomic analyses of middle American cichlids: when data agree to disagree. Genome Biol. Evol. 13:evab161. doi:10.1093/gbe/evab161

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. Nat. Ecol. Evol. 2:688–696. doi:10.1038/s41559-018-0494-6

Andermann T., Fernandes A.M., Olsson U., Töpel M., Pfeil B., Oxelman B., Aleixo A., Faircloth B.C., Antonelli A. 2019. Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. Syst. Biol. 68:32–46. doi:10.1093/sysbio/syy039

Anisimova M., Gil M., Dufayard J.-F., Dessimoz C., Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–699. doi:10.1093/sysbio/syr041

Avise J.C., Robinson T.J. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. Syst. Biol. 57:503–507. doi:10.1080/10635150802164587

Baensch F. 2017. Dwarf angelfish, in: marine ornamental species aquaculture. Hoboken, NJ: John Wiley & Sons, Ltd. p. 279–298. doi:10.1002/9781119169147.ch16

Baensch F.U., Tamaru C.S. 2009. Captive hybridization of two geographically isolated pygmy angelfish species, *Centropyge fisheri* and *Centropyge resplendens*. J. Fish Biol. 75:2571–2584. doi:10.1111/j.1095-8649.2009.02452.x

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477. doi:10.1089/CMB.2012.0021

Baraf L.M., Pratchett M.S., Cowman P.F. 2019. Ancestral biogeography and ecology of marine angelfishes (F: Pomacanthidae). Mol. Phylogenet. Evol. 140:106596. doi:10.1016/j.ympev.2019.106596

Barlow A., Cahill J.A., Hartmann S., Theunert C., Xenikoudakis G., Fortes G.G., Paijmans J.L.A., Rabeder G., Frischauf C., Grandal-d'Anglade A., García-Vázquez A., Murtskhvaladze M., Saarma U., Anijalg P., Skrbinšek T., Bertorelle G., Gasparian B., Bar-Oz G., Pinhasi R., Slatkin M., Dalén L., Shapiro B., Hofreiter M. 2018. Partial genomic survival of cave bears in living brown bears. Nat. Ecol. Evol. 2:1563–1570. doi:10.1038/s41559-018-0654-8

Bellwood D.R., Van Herwerden L., Konow N. 2004. Evolution and biogeography of marine angelfishes (Pisces: Pomacanthidae). Mol. Phylogenet. Evol. 33:140–155. doi:10.1016/j.ympev.2004.04.015

Betancur-R, R., Naylor G.J.P., Ortí G. 2014. Conserved genes, sampling error, and phylogenomic inference. Syst. Biol. 63:257–262. doi:10.1093/sysbio/syt073

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., Maio N.D., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15:e1006650. doi:10.1371/journal.pcbi.1006650

Bowen B.W., Muss A., Rocha L.A., Grant W.S. 2006. Shallow mtDNA coalescence in Atlantic pygmy angelfishes (genus Centropyge) indicates a recent invasion from the Indian Ocean. J. Hered. 97:1–12. doi:10.1093/jhered/esj006

Bravo G.A., Antonelli A., Bacon C.D., Bartoszek K., Blom M.P.K., Huynh S., Jones G., Knowles L.L., Lamichhaney S., Marcussen T., Morlon H., Nakhleh L.K., Oxelman B., Pfeil B., Schliep A., Wahlberg N., Werneck F.P., Wiedenhoeft J., Willows-Munro S., Edwards S.V. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7:e6399. doi:10.7717/peerj.6399

Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932. doi:10.1093/molbev/mss086

Burgess W. E. 1991. Two new genera of angelfishes, family Pomacanthidae. Trop. Fish Hobbyist. 39(7):68–70.

Cao Z., Liu X., Ogilvie H., Yan Z., Nakhleh L. 2023. CHAPTER 5 Practical Aspects of Phylogenetic Network Analysis Using PhyloNet. In L. Kubatko & L. Knowles (Ed.), Species Tree Inference: A Guide to Methods and Applications. Princeton University Press. p. 89–199. doi: 10.1515/9780691245157-00

Carnevale G. 2006. Morphology and biology of the Miocene butterflyfish *Chaetodon ficheuri* (Teleostei: Chaetodontidae). Zool. J. Linn. Soc. 146:251–267.

Chan K.O., Hutter C.R., Wood Jr P.L., Grismer L.L., Brown R.N. 2020. Larger, unfiltered data sets are more effective at resolving phylogenetic conflict: introns, exons, and UCEs resolve ambiguities in Golden-backed frogs (Anura: Ranidae; genus Hylarana). Mol. Phylogenet. Evol. 151:106899.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324. doi:10.1093/bioinformatics/btu530

Chou J., Gupta A., Yaduvanshi S., Davidson R., Nute M., Mirarab S., Warnow T. 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics 16:S2. doi:10.1186/1471-2164-16-S10-S2

Coleman R., Eble J.A., DiBattista J.D. 2016. Regal phylogeography: range-wide survey of the marine angelfish *Pygoplites diacanthus* reveals evolutionary partitions between the Red Sea, Indian Ocean, and Pacific Ocean. Mol. Phylogenet. Evol. 100:243–253. doi:10.1016/j.ympev.2016.04.005

Collins T.M., Fedrigo O., Naylor G.J.P. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol. 54:493–500. doi:10.1080/10635150590947339

Cuvier, G., Valenciennes, A. 1831. Histoire naturelle des poissons. Tome septième. Livre septième. Des Squamipennes. Livre huitième. Des poissons à pharyngiens labyrinthiformes. 7:441-531

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68–768. doi:10.1371/journal.pgen.0020068

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340. doi:10.1016/j.tree.2009.01.009

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

Dibattista J.D., Gaither M.R., Hobbs J.-P.A., Rocha L.A., Bowen B.W. 2016. Angelfishes, paper tigers, and the devilish taxonomy of the *Centropyge flavissima* complex. J. Hered. 107:647–653. doi:10.1093/jhered/esw062

DiBattista J.D., Alfaro M.E., Sorenson L., Choat J.H., Hobbs J.P.A., Sinclair-Taylor T.H., Rocha L.A., Chang J., Luiz O.J., Cowman P.F., Friedman M., Berumen M.L. 2018. Ice ages and butterflyfishes: phylogenomics elucidates the ecological and evolutionary history of reef fishes in an endemism hotspot. Ecol. Evol. 8:10989–11008. doi:10.1002/ece3.4566

DiBattista J.D., Waldrop E., Bowen B.W., Rocha L.A. 2012. Twisted sister species of pygmy angelfishes: discordance between taxonomy, coloration, and phylogenetics. Coral Reefs 31:839–851. doi:10.1007/s00338-012-0907-y

Drummond A.J., Ho S.Y., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

Drummond A.J., Suchard M.A., Xie D., Andrew R. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Dunn C.W., Hejnol A., Matus D.Q., Pang K., Browne W.E., Smith S.A., Seaver E., Rouse G.W., Obst M., Edgecombe G.D., Sørensen M.V., Haddock S.H.D., Schmidt-Rhaesa A., Okusu A., Kristensen R.M., Wheeler W.C., Martindale M.Q., Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749. doi:10.1038/nature06614

Durand E.Y., Patterson N., Reich D., Slatkin M. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252. doi:10.1093/molbev/msr048

Eaton D.A.R., Hipp A.L., González-Rodríguez A., Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution 69:2587–2601. doi:10.1111/evo.12758

Eaton D.A.R., Ree R.H. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Syst. Biol. 62:689–706. doi:10.1093/sysbio/syt032

Edelman N.B., Frandsen P.B., Miyagi M., Clavijo B., Davey J., Dikow R.B., García-Accinelli G., Van Belleghem S.M., Patterson N., Neafsey D.E., Challis R., Kumar S., P Moreira G.R., Salazar C., Chouteau M., Counterman B.A., Papa R., Blaxter M., Reed R.D., Dasmahapatra K.K., Kronforst M., Joron M., Jiggins C.D., Owen McMillan W., Di Palma F., Blumberg A.J., Wakeley J., Jaffe D., Mallet J. 2019. Genomic architecture and introgression shape a butterfly radiation. Science 366:594–599.

Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19. doi:10.1111/j.1558-5646.2008.00549.x

Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462. doi:10.1016/j.ympev.2015.10.027

Elworth R.A.L., Allen C., Benedict T., Dulworth P., Nakhleh L., 2018. DGEN: a test statistic for detection of general introgression scenarios. Presented at the In 18th International Workshop on Algorithms in Bioinformatics (WABI 2018)., Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Everson K.M., McLaughlin J.F., Cato I.A., Evans M.M., Gastaldi A.R., Mills K.K., Shink K.G., Wilbur S.M., Winker K. 2019. Speciation, gene flow, and seasonal migration in Catharus thrushes (Aves:Turdidae). Mol. Phylogenet. Evol. 139:106564. doi:10.1016/j.ympev.2019.106564

Faircloth B.C. 2013. llumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. doi:10.6079/J9ILL

Faircloth B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788.

Faircloth B.C., Alda F., Hoekzema K., Burns M.D., Oliveira C., Albert J.S., Melo B.F., Ochoa L.E., Roxo F.F., Chakrabarty P., Sidlauskas B.L., Alfaro M.E. 2020. A target enrichment bait set for studying relationships among ostariophysan fishes. Copeia 108:47–60. doi:10.1643/CG-18-139

Frédérich B., Santini F., Konow N., Schnitzler J., Lecchini D., Alfaro M.E. 2017. Body shape convergence driven by small size optimum in marine angelfishes. Biol. Lett. 13:20170154. doi:10.1098/rsbl.2017.0154

Fricke R., Eschmeyer W., van der Laan R. 2023. Eschmeyer's catalog of fishes: genera, species, references. (http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp). Electronic version accessed 10 Dec 2023

Gaither M.R., Schultz J.K., Bellwood D.R., Pyle R.L., DiBattista J.D., Rocha L.A., Bowen B.W. 2014. Evolution of pygmy angelfishes: Recent divergences, introgression, and the usefulness of color in taxonomy. Mol. Phylogenet. Evol. 74:38–47. doi:10.1016/j.ympev.2014.01.017

Gatesy J., Springer M.S. 2013. Concatenation versus coalescence versus "concatalescence.". Proc. Natl. Acad. Sci. 110:E1179–E1179. doi:10.1073/pnas.1221121110

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Hibbins M.S., Hahn M.W. 2022. Phylogenomic approaches to detecting and characterizing introgression. Genetics 220:iyab220. doi:10.1093/genetics/iyab173

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis, biology. Sys. Bio. 42:182–192. doi:10.1093/sysbio/42.2.182

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Le S.V. 2017. UFBoot2: improving the ultrafast bootstrap approximation. molecular biology and evolution. Mol. Biol. Evol. 35:518–522.

Huang H., Knowles L.L. 2009. What is the danger of the anomaly zone for empirical phylogenetics? Syst. Biol. 58:527–536. doi:10.1093/sysbio/syp047

Hughes L.C., Nash C.M., White W.T., Westneat M.W. 2023. Concordance and discordance in the phylogenomics of the wrasses and parrotfishes (Teleostei: Labridae). Syst. Biol. 72:530–543. doi:10.1093/sysbio/syac072

Jiang X., Edwards S.V., Liu L. 2020. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. Syst. Biol. 69:795–812. doi:10.1093/sysbio/syaa008

Jordan, D. S., Jordan, E. K. 1922. Descriptions of two new species of fishes, from Honolulu, Hawaii. Proc. U. S. Natl. Mus. 42(1915):525–527.

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.

Karin B.R., Gamble T., Jackman T.R., Vidal N. 2020. Optimizing phylogenomics with rapidly evolving long exons: comparison with anchored hybrid enrichment and ultraconserved elements. Mol. Biol. Evol. 37:904–922. doi:10.1093/molbev/msz263

Katoh K., Misawa K., Kuma K., Miyata T.; Research, T.M.-N. acids, 2002, undefined. 2002. . MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kaup, J. J. 1860. Ueber die Chaetodontidae. Arch. Naturegeschichte. 26(1):133–156.

Kimball R.T., Wang N., Heimer-McGinn V., Ferguson C., Braun E.L. 2013. Identifying localized biases in large data sets: a case study using the avian tree of life. Mol. Phylogenet. Evol. 69:1021–1032. doi:10.1016/j.ympev.2013.05.029

Kubatko L.S., Chifman J. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BMC Evol. Biol. 19:112. doi:10.1186/s12862-019-1439-7

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24. doi:10.1080/10635150601146041

Kumar K.R., Cowley M.J., Davis R.L. 2019. Next-generation sequencing and emerging technologies. Semin. Thromb. Hemost. 45:661–673. doi:10.1055/s-0039-1688446

Kumar S., Filipski A.J., Battistuzzi F.U., Kosakovsky Pond S.L., Tamura K. 2012. Statistics and truth in phylogenomics. Mol. Biol. Evol. 29:457–472. doi:10.1093/molbev/msr202

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:msw260–msw773.

Larget B.R., Kotha S.K., Dewey C.N., Ané C. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910–2911. doi:10.1093/bioinformatics/btq539

Leaché A.D., Oaks J.R. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 48:69–84. doi:10.1146/annurev-ecolsys-110316-022645

Lemmon E.M., Lemmon A.R. 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44:99–121. doi:10.1146/annurev-ecolsys-110512-135822
Linkem C.W., Minin V.N., Leaché A.D. 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). Syst. Biol. 65:465–477. doi:10.1093/SYSBIO/SYW001
Lischer H.E.L., Excoffier L., Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of microtus voles. Mol. Biol. Evol. 31:817–831. doi:10.1093/molbev/mst271
Liu L., Yu L., Pearl D.K., Edwards S.V. 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58:468–477. doi:10.1093/sysbio/syp031
Long C., Kubatko L. 2018. The effect of gene flow on coalescent-based species-tree inference. Syst. Biol. 67:770–785. doi:10.1093/sysbio/syy020
Longo S.J., Faircloth B.C., Meyer A., Westneat M.W., Alfaro M.E., Wainwright P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Mol. Phylogenet. Evol. 113:33–48. doi:10.1016/j.ympev.2017.05.002
MacGuigan D.J., Near T.J. 2019. Phylogenomic signatures of ancient introgression in a rogue lineage of darters (Teleostei: Percidae). Syst. Biol. 68:329–346. doi:10.1093/SYSBIO/SYY074
Mai U., Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics 19:272. doi:10.1186/s12864-018-4620-2
Martin S.H., Davey J.W., Jiggins C.D. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257. doi:10.1093/MOLBEV/MSU269
Massatti R., Reznicek A.A., Knowles L.L. 2016. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: a case study in Carex sect. Racemosae. Am. J. Bot. 103:337–347. doi:10.3732/ajb.1500315
McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8:e54848. doi:10.1371/journal.pone.0054848
Meleshko O., Martin M.D., Korneliussen T.S., Schröck C., Lamkowski P., Schmutz J., Healey A., Piatkowski B.T., Shaw A.J., Weston D.J., Flatberg K.I., Szövényi P., Hassel K., Stenøien H.K. 2021. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. Mol. Biol. Evol. 38:2750–2766. doi:10.1093/molbev/msab063
Mendes F.K., Hahn M.W. 2018. Why concatenation fails near the anomaly zone. Syst. Biol. 67:158–169. doi:10.1093/sysbio/syx063
Micklich N.R., Tyler J.C., Johnson G.D., Świdnicka E., Bannikov A.F. 2009. First fossil records of the tholichthys larval stage of butterfly fishes (Perciformes, Chaetodontidae), from the Oligocene of Europe. Paläontologische Zeitschrift 83:479–497. doi:10.1007/s12542-009-0031-7
Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534. doi:10.1093/molbev/msaa015
Mirarab S., Bayzid M.S., Boussau B., Warnow T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346:1250463. doi:10.1126/science.1250463
Mirarab S., Bayzid M.S., Warnow T. 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65:366–380. doi:10.1093/sysbio/syu063
Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Zhou X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.
Molloy E.K., Warnow T. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst. Biol. 67:285–303. doi:10.1093/sysbio/syx077
Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A., Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in amaranthaceae s.l. Syst. Biol. 70:219–235. doi:10.1093/sysbio/syaa066
Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. Trends Ecol. Evol. 28:719–728. doi:10.1016/j.tree.2013.09.004
Nash C.M., Lungstrom L.L., Hughes L.C., Westneat M.W. 2022. Phylogenomics and body shape morphometrics reveal recent diversification in the goatfishes (Syngnatharia: Mullidae). Mol. Phylogenet. Evol. 177:107616. doi:10.1016/j.ympev.2022.107616
Near T.J., Dornburg A., Eytan R.I., Keck B.P., Smith W.L., Kuhn K.L., Moore J.A., Price S.A., Burbrink F.T., Friedman M., Wainwright P.C. 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci. U.S.A. 110:12738–12743. doi:10.5061/dryad.d3mb4
Near T.J., Eytan R.I., Dornburg A., Kuhn K.L., Moore J.A., Davis M.P., Wainwright P.C., Friedman M., Smith W.L. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. U.S.A. 109:13698–13703. doi:10.1073/pnas.1206625109
Near T.J., MacGuigan D.J., Parker E., Struthers C.D., Jones C.D., Dornburg A. 2018. Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. Mol. Phylogenet. Evol. 129:268–279. doi:10.1016/j.ympev.2018.09.001
Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQTREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32:268–274.
Owen C.L., Miller G.L. 2022. Phylogenomics of the Aphididae: deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. Syst. Entomol. 47:470–486. doi:10.1111/syen.12542
Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.
Patel S., Kimball R., Braun E. 2013. Error in phylogenetic estimation for bushes in the tree of life. J. Phylogenet. Evol. Biol. 1:1–10. doi:10.4172/2329-9002.1000110
Philippe H., Vienne D.M.D., Ranwez V., Roure B., Baurain D., Delsuc F. 2017. Pitfalls in supermatrix phylogenomics. Eur. J. Taxon. 283:1–25. doi:10.5852/ejt.2017.283
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.
Rancilhac L., Irisarri I., Angelini C., Arntzen J.W., Babik W., Bossuyt F., Künzel S., Lüddecke T., Pasmans F., Sanchez E., Weisrock D., Veith M., Wielstra B., Steinfartz S., Hofreiter M., Philippe H., Vences M. 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. Mol. Phylogenet. Evol. 155:106967. doi:10.1016/j.ympev.2020.106967
Roch S., Nute M., Warnow T. 2019. Long-branch attraction in species tree estimation: inconsistency of partitioned likelihood and topology-based summary methods. Syst. Biol. 68:281–297. doi:10.1093/sysbio/syy061
Santaquiteria A., Siqueira A.C., Duarte-Ribeiro E., Carnevale G., White W.T., Pogonoski J.J., Ricardo B.R. 2021. Phylogenomics and historical biogeography of seahorses, dragonets, goatfishes, and allies (Teleostei: Syngnatharia): assessing factors driving uncertainty in biogeographic inferences. Syst. Biol. 70:1145–1162.
Sayyari E., Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes 9:132. doi:10.3390/genes9030132
Scherz M.D., Masonick P., Meyer A., Hulsey C.D. 2022. Between a rock and a hard polytomy: phylogenomics of the rock-dwelling mbuna cichlids of Lake Malawi. Syst. Biol. 71:741–757. doi:10.1093/sysbio/syac006
Schwarz G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464.
Simion P., Delsuc F., Philippe H., 2020. To what extent current limits of phylogenomics can be overcome? In: Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era. ed. No commercial publisher.
Singhal S., Grundler M., Colli G., Rabosky D.L. 2017. Squamate Conserved Loci (SqCL): A unified set of conserved loci for

phylogenomics and population genetics of squamate reptiles. Mol. Ecol. Resour. 17:e12–e24. doi:10.1111/1755-0998.12681
Smith B.T., Merwin J., Provost K.L., Thom G., Brumfield R.T., Ferreira M., Mauck W.M., Moyle R.G., Wright T.F., Joseph L. 2022. Phylogenomic analysis of the parrots of the world distinguishes artifactual from biological sources of gene tree discordance. Syst. Biol. 72:228–241. doi:10.1093/sysbio/syac055
Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 12:e1005896. doi:10.1371/journal.pgen.1005896
Smith, H. M., Radcliffe, L. 1911. Descriptions of three new fishes of the family Chaetodontidae from the Philippine Islands. Proc. U. S. Natl. Mus. 40(1822):319–326.
Snyder, J. O. 1904. A catalogue of the shore fishes collected by the steamer Albatross about the Hawaiian Islands in 1902. Bull. U. S. Fish Comm. 22(1902):513–538.
Solís-Lemus C., Yang M., Ané C. 2016. Inconsistency of species tree methods under gene flow. Syst. Biol. 65:843–851. doi:10.1093/sysbio/syw030
Song S., Liu L., Edwards S.V., Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad. Sci. U.S.A. 109:14942–14947. doi:10.1073/pnas.1211733109
Sota T., Vogler A.P. 2003. Reconstructing species phylogeny of the carabid beetles *Ohomopterus* using multiple nuclear DNA sequences: heterogeneous information content and the performance of simultaneous analyses. Mol. Phylogenet. Evol. 26:139–154. doi:10.1016/S1055-7903(02)00311-1
Steenwyk J.L., Li Y., Zhou X., Shen X.X., Rokas A. 2023. Incongruence in the phylogenomics era. Nat. Rev. Genet. 24:834–850. doi:10.1038/s41576-023-00620-x
Streicher J.W., Wiens J.J. 2017. Phylogenomic analyses of more than 4,000 nuclear loci resolve the origin of snakes among lizard families. Biol. Lett. 13:20170393. doi:10.1098/rsbl.2017.0393
Tagliacollo V.A., Lanfear R. 2018. Estimating improved partitioning schemes for ultraconserved elements. Mol. Biol. Evol. 35:1798–1811. doi:10.1093/molbev/msy069
Tea Y.-K., Hobbs J.P.A., Vitelli F., DiBattista J.D., Ho S.Y.M., Lo N. 2020. Angels in disguise: sympatric hybridization in the marine angelfishes is widespread and occurs between deeply divergent lineages. Proc. R. Soc. B. 287:20201459. doi:10.1098/rspb.2020.1459
Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinf. 9:322. doi:10.1186/1471-2105-9-322
Tricou T., Tannier E., De Vienne D.M. 2022. Ghost lineages highly influence the interpretation of introgression tests. Syst. Biol. 71:1147–1158. doi:10.1093/sysbio/syac011
Vaughan T.G. 2017. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. Bioinformatics 33:2392–2394. doi:10.1093/bioinformatics/btx155
Wabnitz C. 2003. From ocean to aquarium: the global trade in marine ornamental species. Cambridge, UK: UNEP/Earthprint.
Wagenmakers E.J., Farrell S. 2004. AIC model selection using Akaike weights. Psychon. Bull. Rev. 11:192–196. doi:10.3758/BF03206482
Węcek K., Hartmann S., Paijmans J.L.A., Taron U., Xenikoudakis G., Cahill J.A., Heintzman P.D., Shapiro B., Baryshnikov G., Bunevich A.N., Crees J.J., Dobosz R., Manaserian N., Okarma H., Tokarska M., Turvey S.T., Wójcik J.M., Żyła W., Szymura J.M., Hofreiter M., Barlow A. 2017. Complex admixture preceded and followed the extinction of wisent in the wild. Mol. Biol. Evol. 34:598–612. doi:10.1093/molbev/msw254
Westover K.M., Rusinko J.P., Hoin J., Neal M. 2013. Rogue taxa phenomenon: a biological companion to simulation analysis. Mol. Phylogenet. Evol. 69:1–3. doi:10.1016/j.ympev.2013.05.010
Yi-Kai T., Xin X., Joseph D.D., Nathan L., Peter F.C., Simon Y.H. 2021. Phylogenomic analysis of concatenated ultraconserved elements reveals the recent evolutionary radiation of the fairy wrasses (Teleostei: Labridae: Cirrhilabrus). Syst. Biol. 71:1–12. doi:10.1093/sysbio/syab012
Yu Y., Degnan J.H., Nakhleh L. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. PLoS Genet. 8:e1002660. doi:10.1371/journal.pgen.1002660
Yu Y., Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genom. 16:S10. doi:10.1186/1471-2164-16-S10-S10
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19:153. doi:10.1186/s12859-018-2129-y