

Addressing Gene Tree Discordance and Non-Stationarity to Resolve a Multi-Locus Phylogeny of the Flatfishes (Teleostei: Pleuronectiformes)

RICARDO BETANCUR-R.¹, CHENHONG LI², THOMAS A. MUNROE³, JESUS A. BALLESTEROS¹, AND GUILLERMO ORTÍ^{1,*}

¹Department of Biological Sciences, The George Washington University, 2023 G St. NW, Washington, D.C. 20052, USA; ²College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China; and ³National Systematics Laboratory NMFS/NOAA, Post Office Box 37012, Smithsonian Institution NHB, WC 60, MRC-153, Washington, D.C. 20013-7012, USA

*Correspondence to be sent to: Department of Biological Sciences, The George Washington University, 2023 G St. NW, Washington, D.C. 20052, USA; E-mail: gorti@gwu.edu.

Received 11 December 2012; reviews returned 1 February 2013; accepted 28 May 2013
Associate Editor: Mark Holder

Abstract.—Non-homogeneous processes and, in particular, base compositional non-stationarity have long been recognized as a critical source of systematic error. But only a small fraction of current molecular systematic studies methodically examine and effectively account for the potentially confounding effect of non-stationarity. The problem is especially overlooked in multi-locus or phylogenomic scale analyses, in part because no efficient tools exist to accommodate base composition heterogeneity in large data sets. We present a detailed analysis of a data set with 20 genes and 214 taxa to study the phylogeny of flatfishes (Pleuronectiformes) and their position among percomorphs. Most genes vary significantly in base composition among taxa and fail to resolve flatfish monophyly and other emblematic groups, suggesting that non-stationarity may be causing systematic error. We show a strong association between base compositional bias and topological discordance among individual gene partitions and their inferred trees. Phylogenetic methods applying non-homogeneous models to accommodate non-stationarity have relatively minor effect to reduce gene tree discordance, suggesting that available computer programs applying these methods do not scale up efficiently to the data set of modest size analysed in this study. By comparing phylogenetic trees obtained with species tree (STAR) and concatenation approaches, we show that gene tree discordance in our data set is most likely due to base compositional biases than to incomplete lineage sorting. Multi-locus analyses suggest that the combined phylogenetic signal from all loci in a concatenated data set overcomes systematic biases induced by non-stationarity at each partition. Finally, relationships among flatfishes and their relatives are discussed in the light of these results. We find support for the monophyly of flatfishes and confirm findings from previous molecular phylogenetic studies suggesting their close affinity with several carangimorph groups (i.e., jack and allies, barracuda, archerfish, billfish and swordfish, threadfin, moonfish, beach salmon, and snook and barramundi). [Base compositional bias; Carangimorphariae; compositional attraction; concatenation; gene tree-species tree; multi-locus data set; non-homogeneous models; non-stationarity; Pleuronectiformes; systematic error.]

The advent of genomics and large data sets has spawned enthusiasm and higher confidence in phylogenetic estimation. However, larger and more complex data sets also present larger and more complex challenges for phylogenetic analysis. Complexities involve accommodating diverse biological processes with potential to generate incongruent historical signal, such as incomplete lineage sorting (ILS), horizontal gene transfer, and hidden paralogy (Maddison 1997). Although much effort has been devoted recently to account for ILS using species-tree methods (Maddison 1997; Edwards et al. 2007; Kubatko et al. 2009; Liu et al. 2009a, 2009b), most studies implementing such approaches assume that this biological phenomenon is the main cause of gene tree discordance. However, large multi-locus data sets also may feature dramatic systematic biases for phylogenetic inference as a consequence of rate- and pattern-heterogeneity among data partitions (Pagel and Meade 2004).

Conventional likelihood-based approaches increasingly use partitioned analyses to specify independent DNA substitution models for accommodating among-site rate or pattern heterogeneity (Li et al. 2008; Lanfear et al. 2012); less frequently, mixture-model analyses also have been applied in this context (Pagel and Meade 2004; Roure and Philippe 2011;

Evans and Sullivan 2012). Although these methods account for heterogeneity among sites or partitions in the alignment, they are based on homogeneous models of DNA substitution, assuming uniform processes across taxa.

The importance of non-homogeneity across the evolutionary process has been recognized for some time, in particular the effects of rate variation among lineages or heterotachy (López et al. 2002; Lockhart et al. 2006; Lartillot et al. 2007; Evans et al. 2010; Roure and Philippe 2011) and base composition heterogeneity among taxa or non-stationarity (Weisburg et al. 1989; Lockhart et al. 1992; Conant and Lewis 2001; Jermini et al. 2004; Phillips et al. 2004; Phillips et al. 2006; Sheffield et al. 2009). Departures from the homogeneity condition can thus result in severe systematic error (e.g., Collins et al. 2005; Rodríguez-Ezpeleta et al. 2007) by incorrectly grouping distantly related taxa that converge in fast rates of substitution (i.e., long-branch attraction) or extreme nucleotide composition (i.e., compositional attraction).

Most phylogenetic programs used for phylogenetic reconstruction operate under the assumption of stationarity in base composition, but seldom is this assumption assessed or its violation accounted for. Many DNA markers used for higher-level phylogenetic studies, however, appear to suffer from non-stationarity,

with prominent examples including a diverse array of taxa, such as bats (Van Den Bussche 1998; Teeling et al. 2000), mammal orders (Gibson et al. 2005), birds (Nabholz et al. 2011), fishes (Li and Orti 2007; Orti and Li 2009), hexapods (Delsuc et al. 2003), beetles (Sheffield et al. 2009; Song et al. 2010), yeast (Collins et al. 2005), and basal eukaryote lineages (Rodriguez-Ezpeleta et al. 2007). Some early proposals to correct for nucleotide compositional heterogeneity included computing neighbor-joining trees based on LogDet distances (Lake 1994; Lockhart et al. 1994) or removing, recoding, or translating problematic partitions (Lake 1994; Phillips and Penny 2003; Phillips et al. 2004; Gibson et al. 2005; Evans et al. 2010), but these strategies are not always effective and may induce loss of phylogenetic signal. More recently, complex non-stationary models that take into account base compositional heterogeneity have been developed (Galtier and Gouy 1998; Foster 2004; Blanquart and Lartillot 2006; Boussau and Gouy 2006; Gowri-Shankar and Rattray 2007; Dutheil and Boussau 2008), however, these non-homogeneous (nh) models are computationally demanding and few empirical studies have assessed their efficacy (e.g., Sheffield et al. 2009).

Most studies addressing violations of the assumption of compositional homogeneity investigate systematic error with respect to specific taxa that appear misplaced in the resulting topology (e.g., Li and Orti 2007; Sheffield et al. 2009). However, no empirical papers have assessed systematically the degree of compositional heterogeneity among different genetic markers in a multi-locus data set and the extent to which it may adversely affect phylogenetic estimation. Large data sets, such as the one analysed here, may have one or more non-stationary partitions, but the direction and magnitude of their bias may differ in such a way that their combined effect is neutralized with respect to the prevailing phylogenetic signal. Although systematic biases introduced by individual non-stationary partitions may be averaged out or greatly reduced by high signal-to-noise ratios achieved by combining multiple genetic partitions (but see Jeffroy et al. 2006; Rodriguez-Ezpeleta et al. 2007), loci with stronger compositional heterogeneity may result in severe phylogenetic biases (Castresana 2007). Rather than assuming this to be true, this possibility needs to be explored for each particular case.

In this study, we investigate the effects of non-stationary partitions in phylogenetic estimation of single-gene and multi-locus data sets generated for taxa from the flatfish order Pleuronectiformes. Flatfishes, including the popular flounders, halibuts, turbot, and soles, are well-known for having both eyes on the same side of the head, representing the most extreme deviation from symmetry among vertebrates (Munroe 2005). Pleuronectiform monophyly has been supported on the basis of three morphological synapomorphies (Ahlstrom et al. 1984; Chapleau 1993; Munroe 2005) and the group is traditionally divided into two suborders: Psettodoidei, with one genus (*Psettodes*,

including three species), and Pleuronectoidei, including the remaining taxa (ca. 800 species in 120 genera). Fish species with morphology intermediate between a flatfish and a symmetric body plan are rare and known only from the fossil record (i.e., †*Amphistium* and †*Heteronectes*), suggesting that flatfish asymmetry had a single evolutionary origin (Friedman 2008; Friedman 2012). Molecular phylogenetic studies using a handful of markers, however, have so far been inconclusive about pleuronectiform monophyly. The few cases that in fact supported a flatfish clade, either include a comprehensive sampling inside the order but poor (Berendzen and Dimmick 2002) or even non-existent (Azevedo et al. 2008) outgroup taxa or, conversely, have a diverse coverage of percomorph lineages but flatfish sampling is restricted to a few pleuronectoids (Kawahara et al. 2008; Little et al. 2010). In contrast, other large-scale molecular studies comprising broader taxonomic representation have failed to recover the monophyly of this prominent group (Smith and Wheeler 2006; Li et al. 2009; Li et al. 2011). These findings are intriguing considering that some ichthyologists in the early 1900s challenged the idea of single origin of fish asymmetry by suggesting that *Psettodes* is an asymmetrical percoid that could potentially be placed in the family Serranidae (Kyle 1923; Chabanaud 1949).

Aside from the debate concerning pleuronectiform monophyly, their phylogenetic affinities with other percomorphs were also, until recently, enigmatic. Flatfishes have historically been separated from the remaining percomorph lineages but traditional anatomical studies have failed to identify their sister group due to an apparent lack of shared-derived characteristics (Regan 1912; Norman 1934; Chapleau 1993; Johnson and Patterson 1993; Hensley 1997). Recent molecular phylogenies have shed new light, suggesting a close affinity between flatfishes and other seemingly disparate percomorph taxa, such as barracudas, archerfishes, snooks, moonfishes, swordfishes and billfishes, threadfins, and carangoids (Chen et al. 2003; Miya et al. 2003; Smith and Wheeler 2006; Li et al. 2009; Little et al. 2010; Li et al. 2011; Betancur-R et al. 2013). This group of taxa was initially referred to as “clade L” sensu (Chen et al., 2003) or Carangimorpha by (Li et al., 2009), and more recently formally named as Carangimorphariae in a revised classification of bony fishes (Betancur-R et al. 2013), which we follow here. The new classification ranks Carangimorphariae (hereafter carangimorphs), together with Anabantomorphariae and Ovalentariae, as subseries of Carangimorpharia, one of the six major series within the subdivision Percomorphaceae (percomorphs).

In this article, we present a thorough investigation of flatfish phylogeny by combining rich genetic coverage (20 loci; ca. 20 kbp) with dense taxonomic breadth (>200 taxa), including all putative flatfish and carangimorph lineages plus a diverse percomorph outgroup. Initial analyses of individual genes resulted in highly incongruent gene trees, the majority of which failed to recover monophyly of Pleuronectiformes

and/or its suborders, carangimorphs, as well as other emblematic groups. To address this lack of congruence, we first investigate the presence of base compositional heterogeneity. We report strong deviations from stationarity in multiple partitions of our data set, assess the degree to which non-stationarity impacts gene tree estimation, and evaluate currently available nh models to reduce this effect. We also explore the extent to which ILS may cause the observed phylogenetic discordance. Finally, we assess whether biases induced by non-stationarity partitions are compensated for in the analysis of the combined multi-locus data set. We show that despite substantial incongruence at the gene level, phylogenetic analyses based on multi-locus sequence data provide significant resolution of flatfish monophyly and interrelationships, ultimately supporting the hypothesis that flatfish asymmetry has a single evolutionary origin.

METHODS AND DATA

Taxonomic Sampling

Flatfish diversity is represented in our study by a sample of 79 specimens from 75 nominal species (out of *ca.* 800 spp. described for the order), distributed among 59 genera (out of *ca.* 120), 13 families (out of 15), and the two recognized suborders (Psettidoidei and Pleuronectoidei). The number of specimens sampled for each of the 13 families was as follows: Achiridae (5), Achiropsettidae (4), Bothidae (10), Citharidae (3), Cynoglossidae (4), Paralichthyidae (11), Pleuronectidae (17), Poecilopsettidae (2), Psettodidae (4), Rhombosoleidae (4), Samaridae (3), Scophthalmidae (3), and Soleidae (9). Two family-level lineages (Paralichthodidae and the *incertae sedis* genus *Tephrinectes*) were not included in the analyses due to unavailability of material.

To test hypotheses of flatfish monophyly and interrelationships, we also examined a broad arrange of percomorph taxa (134 species in 91 families), laying emphasis on carangimorph lineages suggested by recent molecular studies (Smith and Wheeler 2006; Li et al. 2009; Little et al. 2010; Li et al. 2011; Betancur-R et al. 2013). Sequences from two non-percomorph taxa (*Zeus faber* and *Polymixia japonica*) were used as outgroups. Most taxa examined have voucher specimens deposited at the ichthyology collection, University of Kansas, or other repositories. A complete list of material examined (214 taxa) and catalog numbers is given in Supplementary Table 1. All supplementary files listed here, including figures, tables, and aligned data sets are available from the Dryad data repository at <http://datadryad.org>, doi:10.5061/dryad.dj31p. Tree and alignment files are also available from TreeBase at <http://purl.org/phylo/treebase/phylogenies/study/TB2:S14273>. See also <http://datadryad.org/review?wfid=11972&token=65b90899-3dd5-492e-8ddf-b0d1156b98d4> and <http://purl.org/phylo/treebase/phylogenies/>

study/TB2:S14239?x-access-code=94f63fbf67e7dfdaf60a7322804f83a4&format=html.

Molecular Data and Sequence Alignment

We generated sequences for 20 genetic markers, including one mitochondrial (mtDNA) gene (16S ribosomal DNA) and 19 single-copy nuclear exons (nDNA). Table 1 provides detailed information on markers examined and abbreviations used throughout the text. The 19 nDNA markers used include the popular RAG1 fragment (exon 3) (López et al. 2004) and four markers (IRBP, MLL, RH, and RNF213) designed and/or optimized by other research groups (Chen et al. 2003; Dettai and Lecointre 2005; Dettai and Lecointre 2008; Li et al. 2009). The other 14 loci were either previously published by the Euteleost Tree of Life consortium (Li et al. 2007; Li et al. 2008; Li et al. 2010; Li et al. 2011; Betancur-R et al. 2013; Broughton et al. 2013): KIAA1239, MYH6, PLAGL2, RIPK4, SH3PX3, SIDKEY, SLC10A3, UBE3A, UBE3A-like, ZNF503, and ZNF536; or they were newly developed for this study (EHHADH, GCS1, and GPR61) using the same approach. DNA extraction and amplification protocols via nested PCR followed previous studies (Li et al. 2007; Li et al. 2008; Li et al. 2010; Li et al. 2011). A summary of 103 primer sequences and optimized PCR conditions used for the 20 markers is presented in Supplementary

TABLE 1. Markers examined

Locus	Type	Abbreviation
16S rRNA	Mitochondrial ribosomal	16S
Gene for peroxisomal enoyl-CoA hydratase/L-3-hydroxyacyl-CoA dehydrogenase ^a	Nuclear exon	EHHADH
Hypothetical protein ^a	Nuclear exon	GCS1
G protein-coupled receptor 61 ^a	Nuclear exon	GPR61
Interphotoreceptor retinoid-binding protein	Nuclear exon	IRBP
LOC562320	Nuclear exon	KIAA1239
Mixed-lineage leukemia-like protein	Nuclear exon	MLL
Cardiac muscle myosin heavy chain 6	Nuclear exon	MYH6
Pleiomorphic adenoma gene-like 2	Nuclear exon	PLAGL2
Recombination activating gene 1	Nuclear exon	RAG1
Rhodopsin	Nuclear exon	RH
Receptor-interacting serine-threonine kinase 4	Nuclear exon	RIPK4
Isolate NJ1 ring finger protein 213	Nuclear exon	RNF213
Similar to SH3 and PX domain containing three genes	Nuclear exon	SH3PX3
si:dkey-174m14.3	Nuclear exon	SIDKEY
Hypothetical protein	Nuclear exon	SLC10A3
Ubiquitin protein ligase E3A	Nuclear exon	UBE3A
Ubiquitin protein ligase E3A-like	Nuclear exon	UBE3A-like
Zinc finger protein 503	Nuclear exon	ZNF503
Zinc finger protein 536	Nuclear exon	ZNF536

^aNew markers developed in this study.

Table 2. The PCR amplicons obtained were submitted for purification and sequencing in both directions to High Throughput Sequencing Solutions (HTSeq.org), University of Washington, Seattle, Washington.

Although the great majority of sequences (83%) were generated in this study, a fraction was obtained from GenBank or other genomic databases (e.g., Ensemble and the Cichlid Genome Consortium). In order to minimize missing data, a few of the sequences retrieved from public databases (5.6% of the total sequences) were used to compile multi-locus data for composite taxa. These comprise genus-level sequences for carangimorphs (including flatfishes) and genus-level or family-level sequences for other taxa. Alternative species examined and non-conspecific entries in the data matrix are specified in Supplementary Table 1.

Contigs were assembled from forward and reverse sequences using CodonCode Aligner v3.5.4 (CodonCode Corporation) or Sequencher v4 (Gene Codes Corporation). The 16S sequences were aligned in SATE v1.4 (Liu et al. 2009b) using 50 iterations and aligned blocks including a high proportion of indels were removed from the final data set. Exon markers were aligned individually either manually or using the Translation Align algorithm in Geneious v5.4 (Biomatters Ltd.). All exon alignments were inspected to verify the absence of stop codons or disrupted reading frames. Because nested PCR is highly prone to cross-contamination, we vetted the data during several sequencing stages by visually inspecting individual gene trees generated under maximum likelihood (ML) using the PhyML (Guindon and Gascuel 2003) module in Geneious.

Diagnosing Non-Stationarity

As a first approximation, we assessed stationarity using the χ^2 test implemented in PAUP* v4.0b10 (Swofford 2002), separately for 16S and for each codon position of individual exons (total 58 partitions), after excluding missing taxa for each partition. Due to issues of type II error with the χ^2 test (Foster 2004), we also used the disparity index (I_D) of Kumar and Gadagkar 2001, implemented in MEGA5 (Tamura et al. 2011) as an alternative metric to detect non-stationarity. The I_D was originally developed for identifying non-stationary genes in a pairwise framework by summarizing observed differences in composition for any given pair of sequences (e.g., human and mouse). However, detecting deviant partitions becomes difficult for large phylogenetic data sets with hundreds of taxa. Therefore, as a second approximation, we estimated mean I_D values for each of the 58 partitions (average of all pairwise values among taxa) to assess their degree of compositional heterogeneity, and compared these values against the results obtained with the χ^2 test. Finally, we also attempted to conduct Foster's test as implemented

in P4 v0.89 (Foster 2004) which uses posterior predictive simulations to assess the fit of homogeneous versus heterogeneous models to each data partition. Unlike χ^2 and I_D , this test accounts for evolutionary non-independence; however, its implementation was computationally inefficient even for a reduced data set. Partitions diagnosed with significant non-stationarity were earmarked for differential treatment in subsequent phylogenetic analyses.

Analyses Using Homogeneous Models

Given the size of our data set (214 taxa and 20 genes), phylogenetic analyses using nh models on the complete matrix were infeasible, but these were conducted for a subset of the data (see below). For the complete data set we implemented recoding strategies, such as RY-coding or AGY-coding, to mask the adverse effects of base compositional heterogeneity (Phillips et al. 2004; Gibson et al. 2005). As an additional (and more radical) approach to mask the nucleotide bias, we translated the exon sequences into amino acids and performed ML searches on the resulting protein data set combined with 16S DNA (hereafter DNA-AA). Analyses of individual gene alignments and of the concatenated data set were conducted under ML using Treefinder v1.0-2008 (Jobb et al. 2004) and RAxML v7.2.8 (Stamatakis 2006). Each run was replicated 50 times to check consistency of results. Three strategies were implemented for coding non-stationary partitions diagnosed in the previous step: nucleotide (raw data with no coding, as a control), AGY, and RY. We further checked whether recoded partitions resulted in homogeneous base frequencies across taxa using the χ^2 statistic. To analyse recoded data, we used the GTR3 and GTR2 models implemented in Treefinder v1.0-2008 (Jobb et al. 2004) for AGY- and RY-coded matrices (Gibson et al. 2005), respectively. To the best of our knowledge, Treefinder is the only efficient ML algorithm that implements these models. Nonetheless, for comparison, we also estimated trees using RAxML under a (homogeneous) GTR model for the recoded partitions, with the caveat that these may be subject to over-parameterization.

ML analyses on the DNA data sets were partitioned by gene type (mtDNA vs. nDNA) and by codon position for nuclear exons as follows: one partition for 16S and three partitions for all exon genes, resulting in four partitions for the concatenated data set. For estimation of individual gene trees, the partitioning strategy used three partitions (by codon position) for each individual exon and one partition for 16S. The best-fit model of sequence evolution for each partition that was not diagnosed with non-stationarity was assessed using the "Propose Model" feature with the Akaike Information Criterion (AIC) in TreeFinder (TreeFinder searches only). Partitions diagnosed with non-stationarity were recoded and analysed under a GTR2 or GTR3 model, as explained above. The DNA-AA data set was analysed in RAxML under the GTR model for the DNA partition and the JTTF

for the amino acid partition, as selected with ProtTest (Abascal et al. 2005). Nodal support was assessed in RAXML using the rapid bootstrapping algorithm with automatic estimation of replicates.

Testing the Effects of Non-Stationary Partitions on Phylogenetic Estimation

After diagnosing base compositional heterogeneity in our data set, we explore its potential impact on the phylogenetic results from four different angles. Our exploration is based on analyses of individual genes and multi-locus data. Because this is an empirical study and the true topology is unknown, we chose to focus on three criteria: 1) topological congruence in recovering the monophyly of five emblematic taxa; 2) topological disparity among resulting trees; and 3) support values as a proxy for phylogenetic signal. The five indicator, taxonomic groups are the flatfish order Pleuronectiformes and suborder Pleuronectoidei, as well as three supraordinal clades: Carangimorphariae, Anabantomorphariae, and Carangimorphariae + Anabantomorphariae (see Discussion section and Miya et al. 2005; Smith and Wheeler 2006; Kawahara et al. 2008; Li et al. 2011; Betancur-R et al. 2013). Topological disparity was measured with the PH85 distance using the R package Ape (Paradis et al. 2004) or the Robinson-Foulds (RF) distance implemented in Hashrf v1.0 (Sul and Williams 2008). Whenever necessary, differences in taxon sampling across genes were corrected by estimating the topological distance rate (topological distance rate = topological distance / (number of taxa – 3)) (Sul and Williams 2008). Support values were averaged for each tree across internal branches (mean support = Σ support values / (number of taxa – 3)) and whisker plots were used to compare the resulting mean values as well as the RF distance matrices. Because some of the methods we explore are computationally intensive (e.g., programs implementing nh models) or require full sequence coverage across taxa (i.e., species-tree methods), many analyses were conducted on a subset of 37 species (representing major lineages) and 18 genes for which complete sequences were available (MLL and SH3PX3 were excluded). On the basis of these comparisons we addressed the following questions (not all were evaluated using the three criteria):

1) *How does non-stationarity affect gene tree estimation?*—To quantitatively assess the impact of inferring gene trees under homogeneous models when the assumption of stationarity is violated by the data, we compared the mean I_D value for each partition (as a proxy of non-stationarity) against the topological distance between that gene tree and a reference tree. This comparison uses as a fixed reference the topology of the species tree obtained by analysis of the concatenated data (hereafter referred to as the concatenation tree or topology, shown in Fig. 3), chosen under the assumption that it is more likely to reflect maximum signal-to-noise ratio

by combining more data and averaging out the noise of non-stationarity at individual gene partitions (see prediction 4 below). If base compositional heterogeneity causes a systematic bias for each individual gene data set, we expect a positive correlation between the average value of I_D and the topological distance between the gene tree and the concatenation tree for gene partitions diagnosed with significant non-stationarity. For these partitions, higher values of I_D imply stronger base compositional bias that would cause larger deviations with respect to the averaged concatenation topology. In contrast, stationary partitions should result in non-significant correlations. Trees for each gene were estimated under ML in RAXML using partitioned analyses with homogenous models, as explained above. To control for other confounding factors that may cause topological discordance, for example random mutational variance (noise or homoplasy) due to fragment length (approximated by the amount of phylogenetic signal in each gene partition), we used a generalized linear model (glm function in R; R Development Core Team 2011). With this approach we test the response of topological distance between the concatenation tree and each gene tree (dependent variable) with respect to three independent variables: mean I_D , sequence length, and number of variable sites for each fragment.

Another, more qualitative method for assessing and visualizing the effects of non-stationarity is to trace GC content across the phylogeny for partitions diagnosed with compositional heterogeneity. Using a non-stationary model, we optimized GC content for five heterogeneous markers on their inferred gene tree topology and traced this parameter by coloring internal branches proportional to their inferred GC content. The GC content was optimized using a non-stationary model with four equilibrium frequencies implemented in nhPhyML (Boussau and Gouy 2006) onto the RAXML-derived topologies (based on an homogeneous GTR model). With this approach, it can be seen whether putatively unrelated taxa with extreme values of GC are attracted or grouped together as a result of convergent nucleotide composition (Lockhart et al. 1994; Conant and Lewis 2001).

2) *Does the implementation of nh models reduce gene tree incongruence?*—If the observed topological incongruence among gene trees is mainly due to base compositional biases, we expect that estimation of gene trees with nh models should result in reduced discordance with respect to trees based on models that assume stationarity. This expectation is most relevant for partitions diagnosed with significant GC biases. A failure to meet this expectation could mean that current methods implementing nh models are inefficient (or too sensitive to details of the model) or that other factors may be interacting in unpredictable ways to maintain high levels of incongruence among gene trees.

We implemented nh models under Bayesian and likelihood frameworks using P4 (Foster 2004), nhPyloBayes v0.2.3 (Blanquart and Lartillot 2006), and nhPhyML. The initial plan was to run these applications on the complete, concatenated data set. Given that these programs are computationally inefficient for large data sets, we restricted these analyses to the reduced data subset only. Even then, preliminary analyses on the concatenated subset suggested they would require a running time of 6–12 months to achieve convergence; thus, nh analyses only were conducted on individual genes. The P4 approach to compositional heterogeneity requires the user to specify *a priori* a number of base composition vectors, consistent with the distribution of base frequencies. The values on the vectors and their allocation on the tree are free parameters estimated by Markov Chain Monte Carlo (MCMC) simulation. The program nhPyloBayes uses a doubly stochastic system, first creating breakpoints along the tree. After each breakpoint, the sequences evolve toward new equilibrium frequencies, *a priori* independent from those ancestral to the breakpoint. Unlike P4, the total number of breakpoints is a free variable of the model and is automatically estimated during the MCMC. The nhPhyML program implements the nh model of Galtier and Gouy (1998) in a likelihood framework, using the nearest neighbor interchange (NNI) algorithm of PhyML to explore tree space. The algorithm optimizes equilibrium frequencies for either each branch of the tree or, like P4, using a predefined number of composition vectors. The model is parameter-rich and the program has limitations for exploration of the tree space, requiring the user to supply an initial tree topology.

The P4 analyses were partitioned by codon position and the best-fit model for each partition was estimated with the AIC in MrModeltest 2.3 (Nylander 2004); homogeneous models were specified for partitions without significant compositional heterogeneity, whereas deviant partitions were modeled under non-stationarity using three composition vectors. The P4 and nhPyloBayes MCMC chains were each run using a computer cluster for 2–4 months until P4 sampled 30 million generations, collecting trees every 1000 generations, and until nhPyloBayes obtained 4000–12,000 samples using four independent chains. To determine if the chains were long enough, P4 log files were assessed using the effective sample size (ESS) statistic, where $ESS > 200$ suggest convergence. We also repeated the P4 searches using five composition vectors but all the runs failed to converge ($ESS < 20$) and thus the results are not reported. For nhPyloBayes, the “bpcomp” function compares the set of trees among pairs of sampled chains, calculating the pairwise differences of estimated posterior probability of tree-bipartition, where maximum differences among comparisons of < 0.1 or between 0.1 and 0.3 are good or acceptable indications of chain convergence, respectively (Blanquart and Lartillot 2006). For nhPhyML, four equilibrium frequencies were specified. Topologies

obtained under homogeneous models for each gene in both Bayesian and ML frameworks were used as unfixed starting trees. Bayesian analyses were conducted in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003), with settings similar to those used for P4; ML trees were obtained with RAxML, as explained above. Because nhPhyML has limitations exploring tree space, for each gene we chose the nhPhyML tree with the highest likelihood score obtained from either the RAxML or the MrBayes starting trees. Unlike P4, nhPyloBayes and nhPhyML do not allow implementation of partitioned analyses.

3—Is the discordance between gene trees the result of systematic error or coalescent variance?—It is possible that the observed incongruence among gene trees is (at least in part) due to discordant genealogical histories stemming from ILS (Edwards et al. 2007; Kubatko et al. 2009; Liu et al. 2009a, 2009b; Huang et al. 2010). The actual effects of gene tree discordance on phylogenetic results (the species tree) will depend in part on the factors causing this discordance (ILS or GC bias, or more generally coalescent variance or mutational variance; Huang et al. 2010) and on the methods that attempt (or not) to correct for systematic errors caused by one or more factors. To explore this possibility and its putative consequences, we implemented coalescent-based species-tree methods. Species-tree analyses were infeasible on the complete data set given the random nature of missing species in our matrix and were thus restricted to the data subset. We used the tree-based program STAR implemented in the R package Phybase (Liu et al. 2009a, 2009b). Input trees included those estimated under different models (homogeneous and non-homogeneous), coding strategies (nucleotide, AGY, and RY), and methods (RAxML, TreeFider, MrBayes, P4, nhPyloBayes, and nhPhyML).

If gene tree disparity is the product of widespread ILS, concatenation results could be misled or at least give different results compared to methods that account for coalescent variance. The actual difference in species trees estimated by both methods is hard to predict due to complex interactions among factors affecting mutational variance (or systematic error, as suspected in this case), ILS, and the actual phylogenetic history (topology, branch lengths) of the target species (Huang et al. 2010). On the basis of comparisons among analytical approaches that account for either non-stationarity or ILS, or both, we explore 1) whether the discrepancy between STAR and concatenation trees is reduced by applying methods that account for non-stationarity; and 2) whether the discrepancy between species trees estimated under homogeneous and nh models is reduced by applying methods that account for ILS. Although these comparisons will not directly address the putative cause of observed gene tree discordance, they may shed some light on the effect that this disparity and the methods that attempt to account for it have on the resulting phylogeny.

4) Do concatenated multi-locus data sets counteract detrimental effects of non-stationarity among discordant genes?—As previously suggested (Baker and DeSalle 1997; Rokas et al. 2003; Philippe et al. 2011), it may be reasonable to assume that additive phylogenetic signal contained in genome-wide multi-locus data sets may neutralize any systematic bias induced by non-stationarity in each gene partition by increasing signal-to-noise ratio. This may be the case when convergence in base composition among taxa differs from gene to gene. To test this, we assessed the species trees and the concatenated trees using the three criteria outlined above. In addition, we optimized GC content in the concatenation tree to evaluate the prevalence and distribution of branches with extreme GC values. Because we had gene and concatenation trees based on both the full and the reduced data sets, we also tested the effect of taxonomic sampling in determining phylogenetic signal and topological congruence (e.g., Zwickl and Hillis 2002; Rokas and Carroll 2005; Heath et al. 2008).

Finally, we also compared the likelihood of the concatenation topology against each gene tree topology under the most general nh model implemented in nhPhyML, which optimizes a separate equilibrium frequency for each branch in the tree. We used fixed topologies to optimize this general model and also ran approximately unbiased (AU) tests as implemented in CONSEL v.1.2 (Shimodaira and Hasegawa 2001) to compare the concatenation topology against the tree topology obtained for each gene marker using RAxML. Assuming the concatenation topology as a null hypothesis, we compared likelihood scores between this hypothesis and the RAxML topology obtained for each gene. If the concatenation topology is in fact correcting the noise generated by GC biases, we expect gene tree topologies that are forced to match the concatenated tree to have higher likelihood scores under nh models than topologies obtained using RAxML.

RESULTS

Data set Attributes

This study generated 2847 sequences for 20 markers and 214 taxa, and the final data set comprised 19,461 aligned sites. The average presence of data across the alignment was 80.2%, including the 584 additional sequences retrieved from public databases; the number of variable sites in the concatenated alignment was 11,501, and for each locus it ranged from 329–1038 (mean = 575). A summary of data set statistics for each marker and overall, including alignment length, sequence variation, pairwise distances, distribution of present/missing sequences and taxa, and GenBank accession numbers (JQ937381–JQ940227) is given in Supplementary Table 3 (see also Supplementary Table 1).

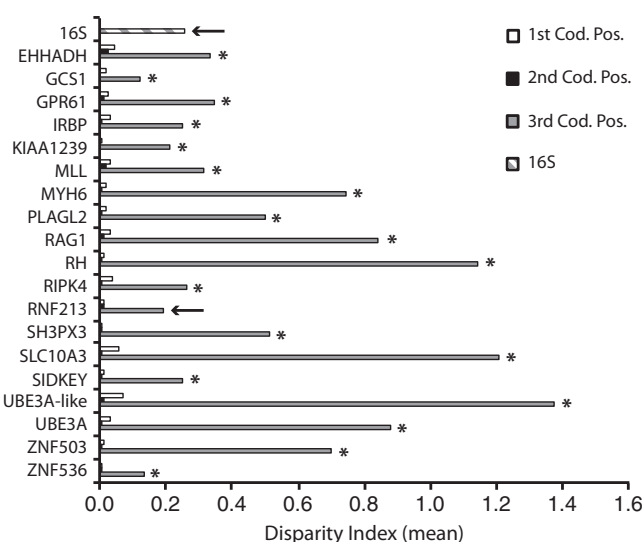


FIGURE 1. Mean disparity index (I_D) values for 58 partitions (by gene and codon position). The asterisk denotes 18 partitions with significant compositional heterogeneity according to the χ^2 test in PAUP* ($P < 0.001$). Arrows indicate two non-significant partitions with, otherwise, relatively high I_D values.

Base Compositional Heterogeneity

The results of the χ^2 test indicate that third codon positions of the 19 nuclear markers have significant levels of compositional heterogeneity ($P < 0.001$). Likewise, mean I_D values are substantially higher for third codon positions (0.123–1.370) and 16S (255) relative to first (0.006–0.071) and second codon positions (0.000–0.029; Fig. 1, Table 2). Although the χ^2 test resulted in non-significant heterogeneity for 16S and third codon positions of locus RNF213, the mean I_D values for these partitions (0.195 and 0.255, respectively) are higher than those in other partitions (e.g., GCS1 = 0.123, ZNF526 = 0.134) with significant χ^2 P -values. We thus treat 16S and all third codon positions of nuclear exons (i.e., 20 out of 58 possible partitions) as non-stationary partitions for downstream analyses (Fig. 1, Table 2).

ML Analyses on Individual Genes and the Concatenated Data set

The ML reconstructions resulted in highly discordant and poorly supported gene trees (mean bootstrap values across nodes and genes = 52.3%), with 18 (out of 20) failing to support flatfish monophyly (Fig. 2; Table 3). Likewise, monophyly of other indicator groups recognized by either morphology (e.g., flatfish suborders; Chapleau 1993; Munroe 2005) or recent molecular studies (e.g., carangimorphs; Li et al. 2009; Li et al. 2011) was not recovered in over half of the gene trees. Despite high degree of discordance at the gene level, the concatenation trees obtained using the three different coding strategies for the

TABLE 2. Mean disparity index (I_D) values for each marker and for each codon position

MARKER	ALL	1st codon pos.	2nd codon pos.	3rd codon pos.
16S	0.255	—	—	—
EHHADH	0.181	0.043	0.029	0.336*
GCS1	0.044	0.017	0.000	0.123*
GPR61	0.162	0.024	0.013	0.344*
IRBP	0.108	0.034	0.007	0.251*
KIAA	0.069	0.007	0.000	0.215*
MLL	0.151	0.03	0.02	0.316*
MYH6	0.259	0.018	0.005	0.745*
Plagl2	0.19	0.02	0.002	0.503*
RAG1	0.346	0.031	0.014	0.838*
RH	0.363	0.014	0.008	1.139*
RIPK4	0.1	0.039	0.005	0.264*
RNF213	0.069	0.015	0.011	0.195
SH3PX3	0.154	0.006	0.001	0.516*
SLC10A3	0.572	0.059	0.007	1.209*
SIDKEY	0.068	0.01	0.002	0.252*
UBE3A-like	0.72	0.071	0.014	1.37*
UBE3A	0.356	0.03	0.003	0.881*
ZN503	0.253	0.014	0.003	0.7*
ZNF536	0.043	0.008	0.003	0.134*

Significant compositional heterogeneity according to the χ^2 test in PAUP ($P < 0.001$).

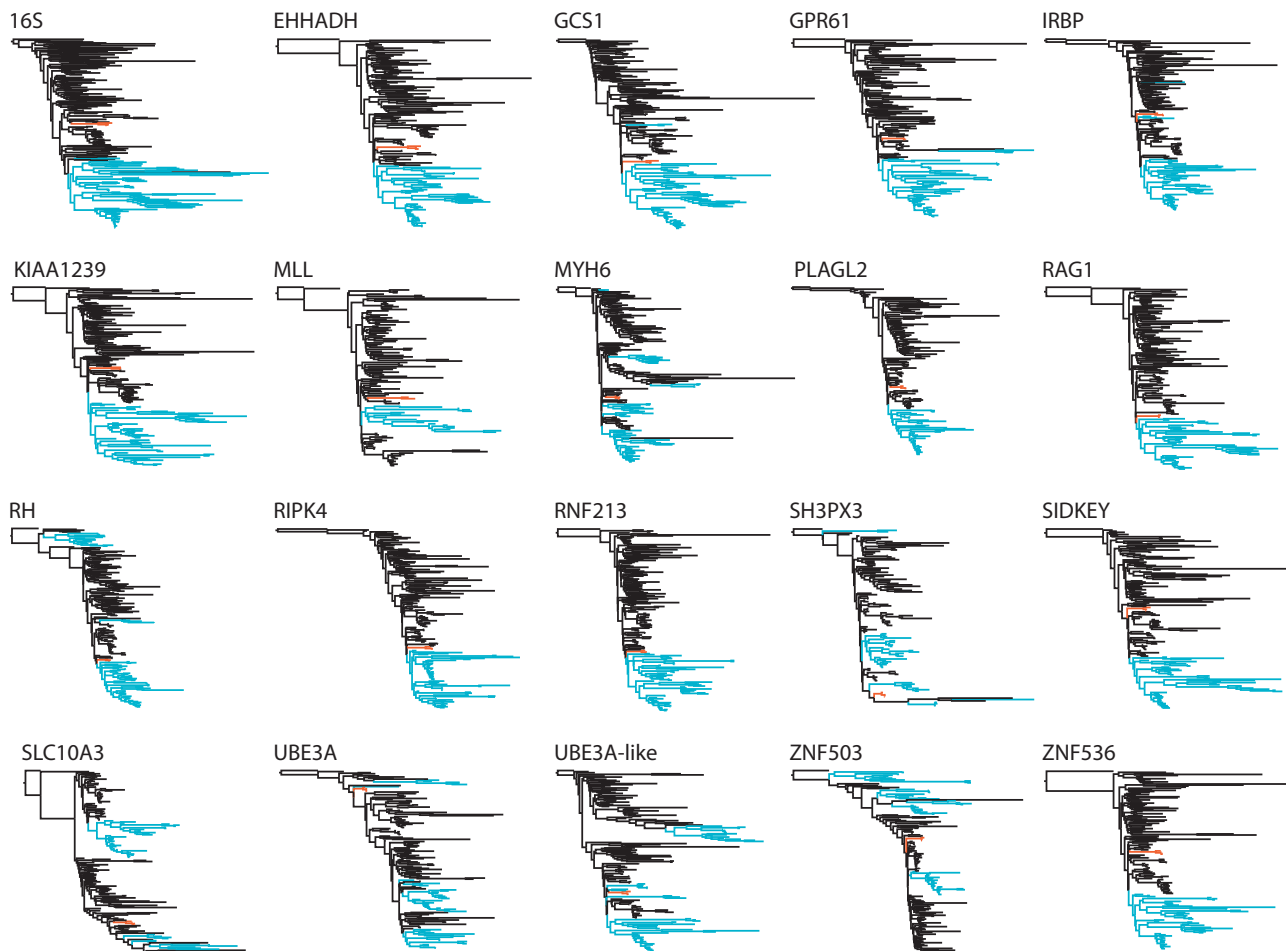


FIGURE 2. ML trees estimated with RAxML for each of the 20 markers examined. Color branches represent flatfish lineages (cyan, Pleuronectoidei; orange, Psettodei).

TABLE 3. Number of phylogenetic trees (out of total trees obtained in each case) that support the monophyly of five representative suprafamilial or supraordinal taxa for different analytical approaches outlined in the text

	Pleuronectoidei	Pleuronectiformes	Carangimorphariae	Anabantomorphariae	Anabantomorphariae + Carangimorphariae
GENE TREES					
All data (RAxML)	7/20	2/20	9/20	9/20	8/20
Subset (RAxML)	12/18	4/18	13/18	11/18	10/18
Subset (MrBayes)	14/18	2/18	14/18	13/18	12/18
Subset (P4)	14/18	0/18	13/18	14/18	10/18
Subset (nhPhyloBayes)	11/18	1/18	10/18	12/18	8/18
Subset (nhPhyML)	14/18	3/18	14/18	12/18	12/18
ALL DATA (concatenation—214 taxa, 20 genes)					
RAxML (nucleotide)	1	1	1	1	1
RAxML (AGY)	1	0	1	1	1
RAxML (RY)	1	1	1	1	1
TreeFinder (nucleotide)	1	1	1	1	1
TreeFinder (AGY)	1	1	1	1	1
TreeFinder (RY)	1	0	1	1	1
SUBSET (multi-locus—37 taxa, 18 genes)					
STAR species tree—RAxML (nucleotide)	1	0	1	1	1
STAR species tree—MrBayes (nucleotide)	1	0	1	1	1
STAR species tree—TreeFinder (nucleotide)	1	0	1	1	1
STAR species tree—TreeFinder (AGY)	1	0	1	1	1
STAR species tree—TreeFinder (RY)	1	0	1	1	1
STAR species tree—P4	1	0	1	1	1
STAR species tree—nhPhyloBase	1	0	1	1	1
STAR species tree—nhPhyML	1	0	1	1	1
Concatenation—RAxML (nucleotide)	1	0	1	1	1
Concatenation—MrBayes (nucleotide)	1	0	1	1	1
Concatenation—TreeFinder (nucleotide)	1	0	1	1	1
Concatenation—TreeFinder (AGY)	1	0	1	1	1
Concatenation—TreeFinder (RY)	1	0	1	1	1

DNA data sets (nucleotide, AGY, and RY) were highly congruent and well supported (mean bootstrap values across nodes and coding strategies = 82.6%), with most topologies recovering the monophyly of flatfishes and indicator groups (Table 3). The topology obtained with the DNA-AA data set also was largely congruent with trees estimated with untranslated DNA sequences only (Supplementary Fig. 1); however, the translation to amino acids resulted in considerable loss of phylogenetic signal (mean bootstrap values = 77.3%) and did not resolve flatfishes as a monophyletic group. Therefore, we exclude DNA-AA analyses from further consideration.

The χ^2 test resulted in non-significant P -values after AGY- or RY-recoding of 20 partitions with compositional heterogeneity. In addition, mean support values across nodes were similar for the AGY (RAxML bootstrap = 83.4) and the nucleotide (RAxML bootstrap = 83.3%) trees, which were in turn slightly higher relative to the RY trees (RAxML bootstrap = 81.2%). These results suggest that the milder method of data recoding (AGY) provides the best trade-off in terms of GC bias correction and loss of phylogenetic signal (Gibson et al. 2005). We thus report the tree obtained with TreeFinder, estimated under the GTR3 model for the AGY-recoded partitions, as our favored phylogenetic hypothesis (Fig. 3). The major implications of this result from an ichthyological perspective are presented in the Discussion section.

The Effects of Non-Stationary Partitions on Phylogenetic Estimation

1) *How does non-stationarity affect gene tree estimation?*—As predicted, scatterplots of topological distance between each gene tree and the concatenation tree (estimated with homogeneous models) against mean I_D (Fig. 4) show a strong positive correlation for non-stationary partitions (16S + third codon positions; $R^2=0.343$, $P=0.007$), whereas homogeneous partitions resulted in non-significant correlations (first codon positions: $R^2=0.033$, $P=0.45$; second codon positions: $R^2=0.008$, $P=0.71$). As noted above, bootstrap support values for all gene trees are low (mean 52.3% across nodes and gene trees), suggesting that the phylogenetic signal contained in each data set is weak and may contribute significantly to mutational variance. However, using a GLM to account for other confounding variables, such as sequence length and variable sites, still resulted in a significant fit between topological disparity and mean I_D for non-stationary partitions, albeit the significance is slightly weaker ($P=0.024$). In addition, the two other independent variables did not have a significant response (sequence length: $P=0.65$; variable sites: $P=0.82$).

Trees from five genes with high levels of compositional heterogeneity (MYH6, RH, SLC10A3, UBE3A, and UBE3A-like; mean $I_D=0.75-1.37$; see Fig. 1, Table 2) exhibit polyphyletic flatfishes and, most importantly,

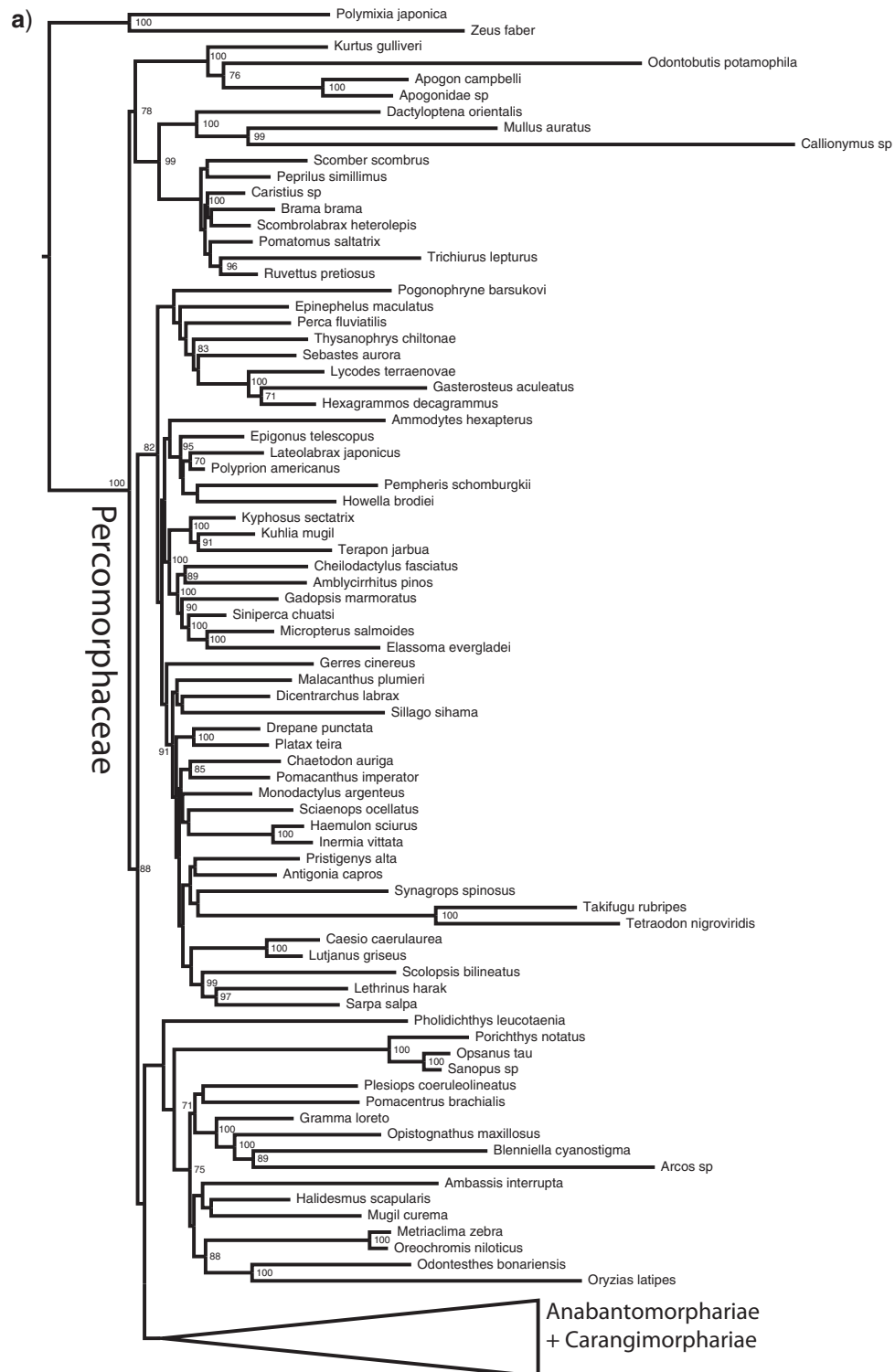


FIGURE 3. Favored phylogenetic hypotheses for the complete data set (214 taxa and 20 genes) obtained with TreeFinder using the GRT3 model for AGY-recorded partitions (see also Fig. 8). Numbers on internal branches indicate support values ($\geq 75\%$), estimated with the rapid bootstrapping algorithm in RAxML. a) complete tree; b) expanded relationships for the Anabantomorphariae + Carangimorphariae clade; c) expanded relationships for Pleuronectiformes.

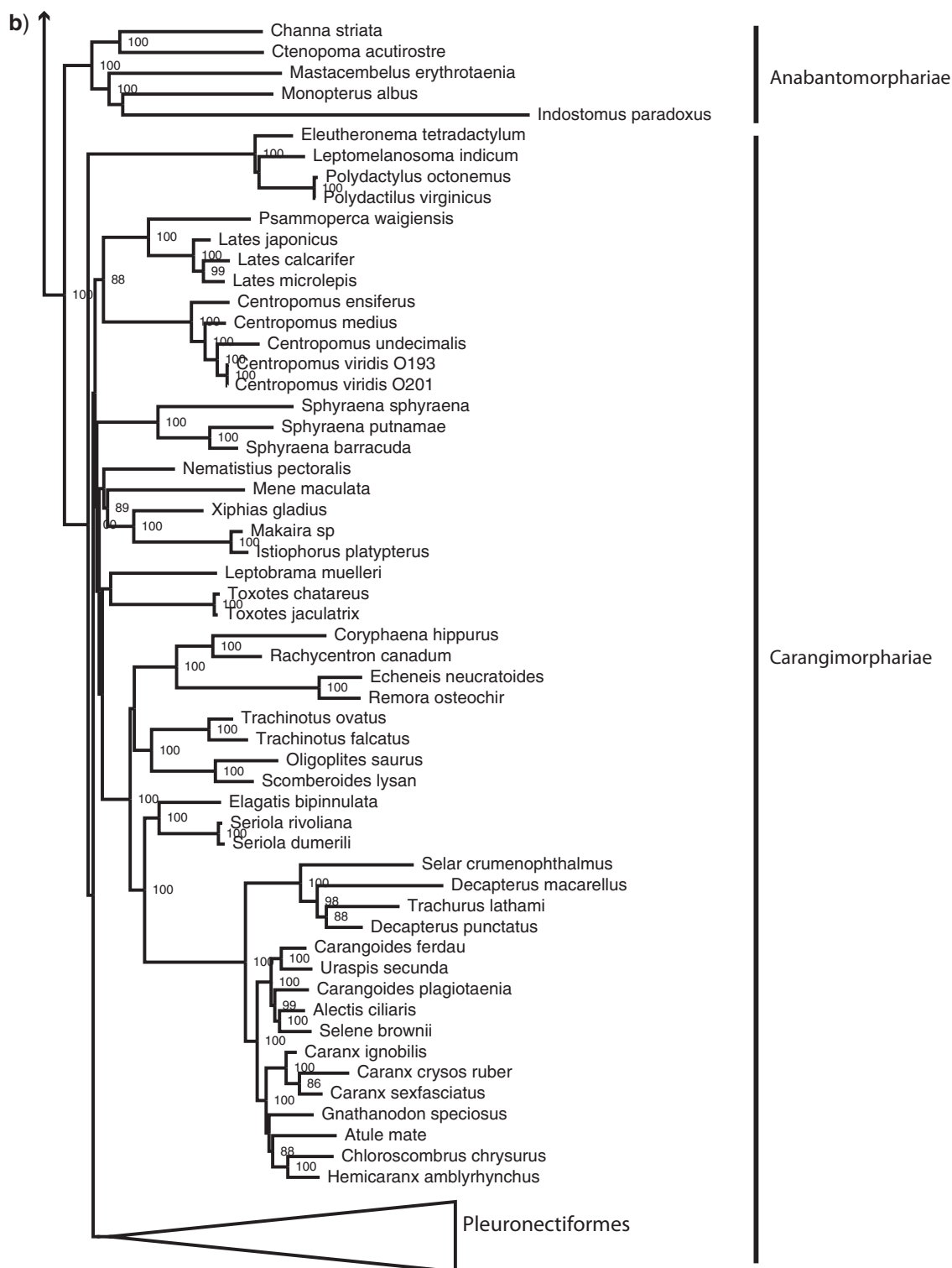


FIGURE 3. Continued

these trees include clades with extreme GC content that are composed of both flatfish (mostly bothids and rhombosoleids) and non-flatfish taxa (Fig. 5). Taken together, these results suggest that violation of the

assumption of stationarity by our data is leading to dangerous levels of systematic error in gene tree estimation. However, if the direction of bias or the distribution of convergence in base composition among

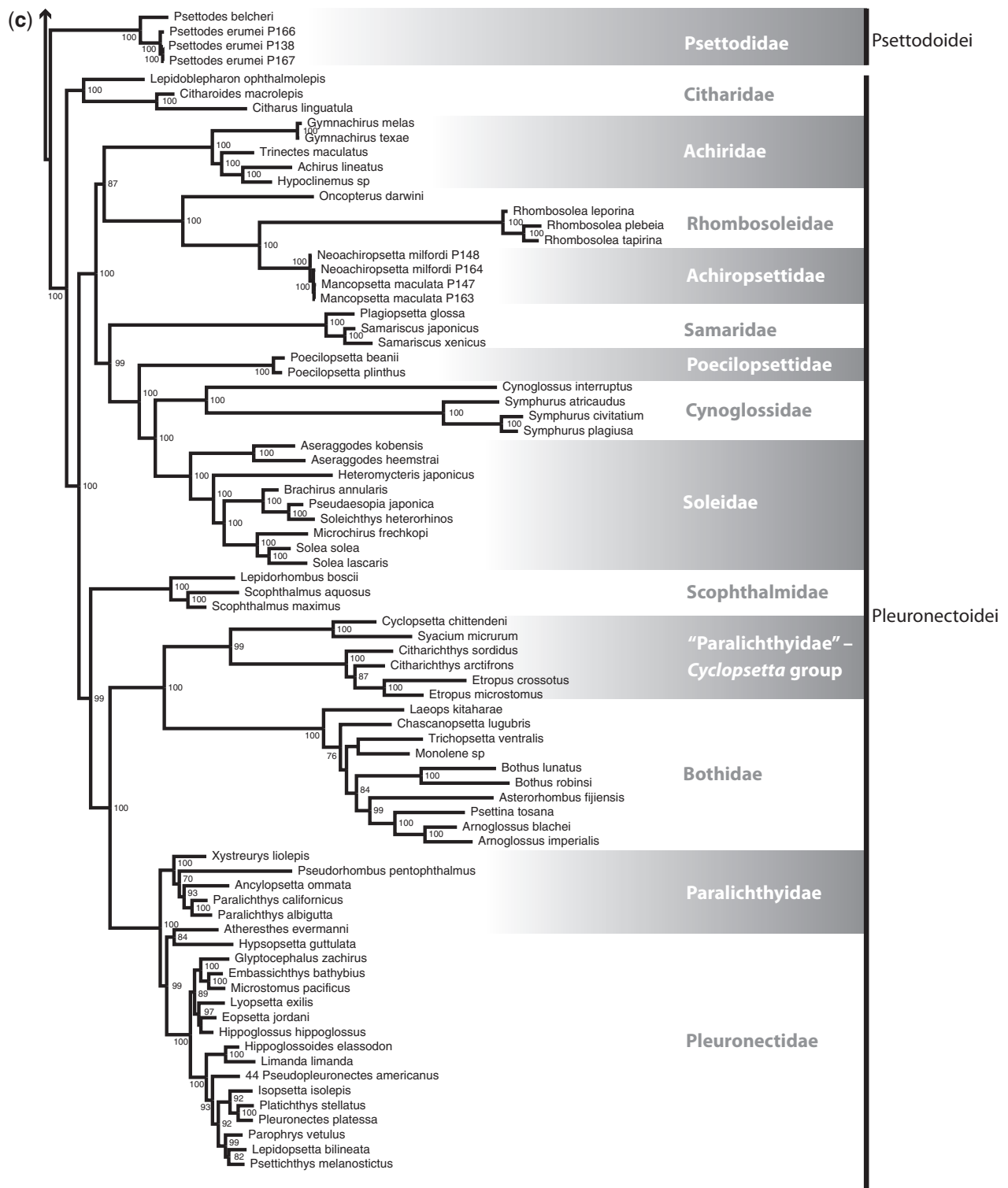


FIGURE 3. Continued

taxa differs from gene to gene, a total evidence analysis combining or concatenating all partitions may correct the systematic bias affecting each individual gene. We test this possibility below (point 4).

2) Does the implementation of *nh* models reduce gene tree incongruence?—We report results using two Bayesian methods (P4 and *nhPhyloBayes*) and a likelihood approach (*nhPhyML*). Mean ESS values obtained across

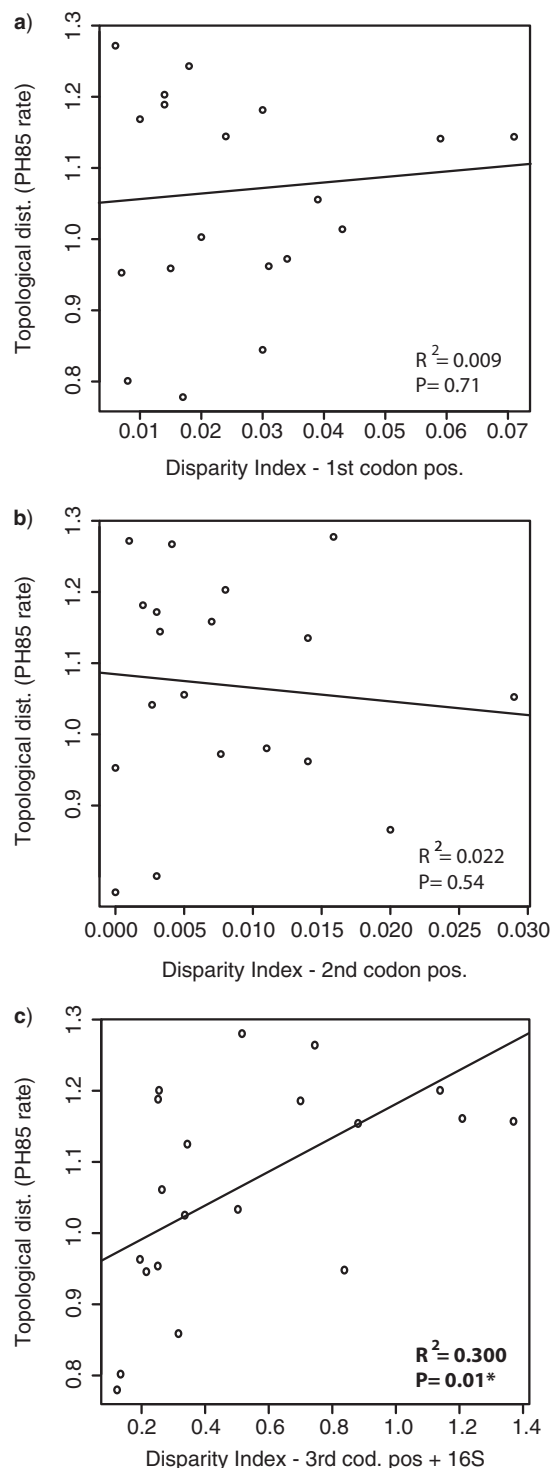


FIGURE 4. Scatterplots of mean disparity index values for each gene partition against topological distance between each gene tree and the concatenation tree (prediction 1). a) First codon positions; b) second codon positions; c) third codon positions plus 16S (first and second codon positions are homogeneous partitions; 16S and third codon positions are not; Fig. 1). Note that scale on the x-axis differs among plots.

all P4 runs was >2000 after 25% burn-in, indicating convergence of most parameters. However, we note that for a minority of the non-stationary partitions (5 out of 20; i.e., third codon positions of GPR61, PLAGL2, RH, RNF213, ZNF536) *ca.* 40% of the compositional parameters had ESS values <100, suggesting the necessity of running the analysis for much longer. These results were obtained after 4 months of continuous runs. We estimate that the computational burden for achieving convergence on those remaining parameters does not justify re-running the analyses or adjusting the models *ad hoc*, on a case-by-case basis. Specification of the best non-stationary model ("model testing") using statistical criteria would be desirable, but resulted computationally intractable. For nhPhyloBayes we obtained 4000–12,000 samples for each chain and discarded 20% of the initial samples as burn-in. The sets of nhPhyloBayes trees resulted in maximum differences of estimated posterior probability of tree-bipartition of <0.1 for 13 genes and 0.1–0.3 for 5 genes, suggesting topological convergence among chains (although it is not possible to assess convergence of equilibrium frequencies). For nhPhyML, we obtained different results depending on the starting tree (RAxML or MrBayes topologies) and thus the tree with the highest likelihood score between the two searches was chosen in each case.

Topological distances among gene trees estimated under homogeneous (RAxML and MrBayes) and most non-homogeneous (P4, nhPhyloBayes, and nhPhyML) models were overall similar (Fig. 6a), but gene tree disparity was slightly higher with the nhPhyloBayes approach (mean disparity: RAxML = 0.72; MrBayes = 0.70; P4 = 0.71; nhPhyloBayes = 0.75; nhPhyML = 0.70; Fig. 6a). Mean posterior probability values across branches for gene trees obtained with MrBayes (65.9%) were slightly higher than those obtained with P4 (63.3%), both of which were in turn substantially higher than the nhPhyloBayes support values (53.9%; Fig. 6b). Monophyly of the five indicator groups was recovered in a greater proportion of cases for gene trees estimated with nhPhyML (61.1%) and MrBayes (61.1%), relative to those obtained with RAxML (55.6%), P4 (56.7%), and nhPhyloBayes (46.7%; see details in Table 3). Taken together these results indicate that while phylogenetic reconstructions using available nh models do not appear to efficiently correct the systematic error induced by non-stationarity in our data set, these may be artifacts due to model misspecification or incomplete searches (most notably for nhPhyML).

3) *Is the discordance between gene trees the result of systematic error or coalescent variance?*—We explore this question indirectly by measuring the effect of methods that attempt to accommodate for one, both, or neither of these factors to estimate the species tree. The magnitude of topological disparity among STAR species trees, obtained from input trees inferred under different models (homogeneous and non-homogeneous) and coding strategies (nucleotide, AGY,

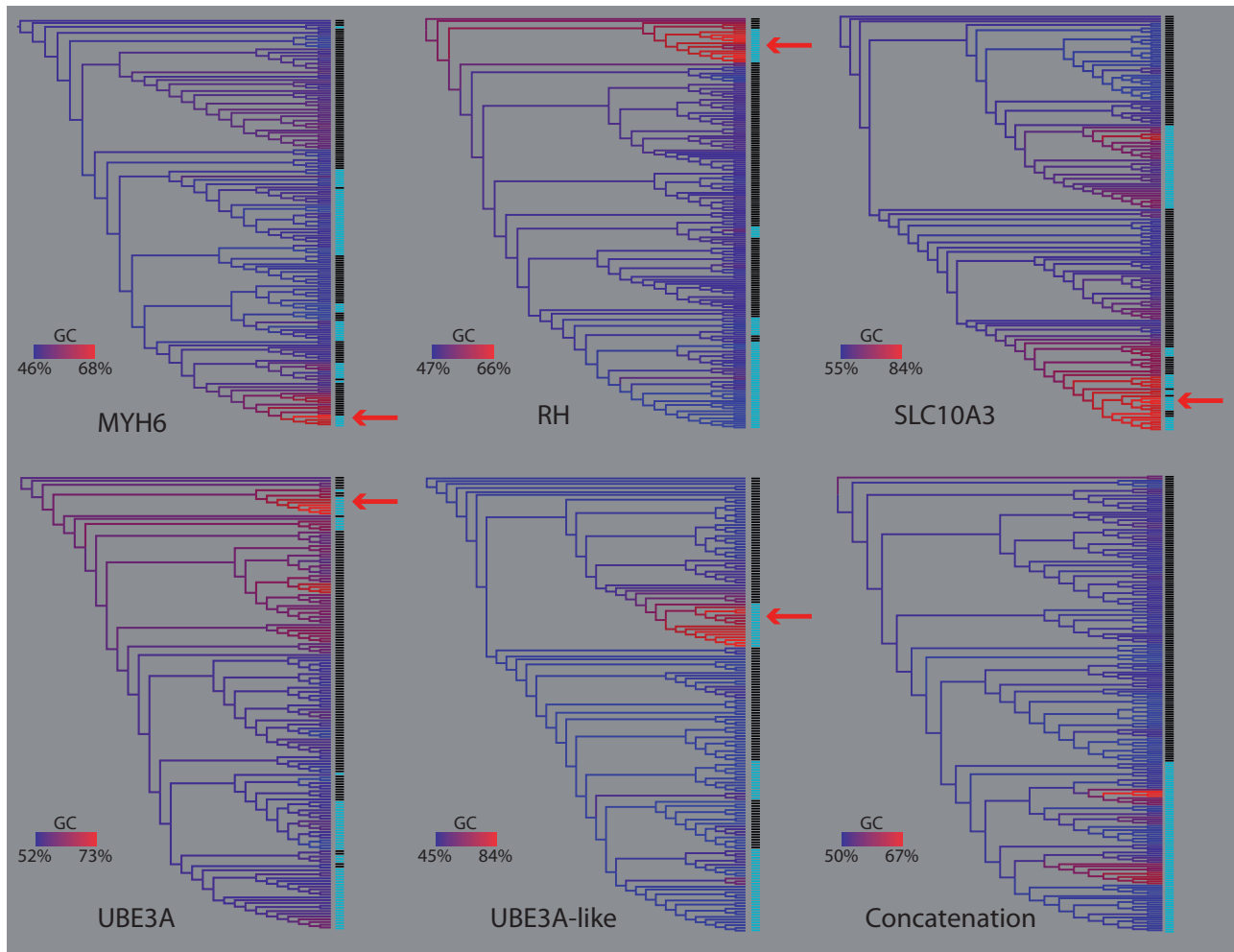


FIGURE 5. Gene tree topologies for five genes with high compositional heterogeneity (Figs. 1 and 2) and the concatenation tree (Fig. 3). GC content is optimized on each branch following a color gradient (prediction 1). Tree topologies were estimated with RAxML and GC content optimization was performed in nhPhyML. The five gene trees display polyphyly for flatfishes, with red arrows indicating clades with high GC content that are composed of both flatfish (cyan terminals) and non-flatfish (black terminals) taxa, indicating compositional attraction. All gene trees are poorly supported; average bootstrap values across nodes: MYH6 = 34.0%, RH = 48.7%, SLC10A3 = 47.3%, UBE3A = 45.5%, UBE3A-like = 47.7%. Color gradient scales for GC content differ among trees.

RY), is similar to the distance observed among STAR and concatenation trees combined (STAR trees mean = 0.255; STAR vs. concatenation trees mean = 0.240; Fig. 6c). Topological distances among concatenation trees alone are substantially smaller (mean = 0.126), but all these trees were inferred under homogenous models only (albeit with and without AGY or RY recoding), and thus include fewer data points (Fig. 6c). Implementation of methods that account for GC biases (recoding) has minor effects on the species trees estimated by concatenation, but causes relatively more discordance among species trees estimated with STAR. Similarly, the level of discrepancy between STAR versus concatenation is not reduced by either recoding methods (AGY or RY on the concatenated or gene trees) or estimation of gene trees under nh models.

Topological comparisons with respect to five key groups were, on the other hand, fully congruent

across methods (concatenation or STAR), with all trees failing to support flatfish monophyly but successfully recovering the remaining clades (Table 3). In conclusion, similar levels of between-method versus within-method variation suggest that the extent to which ILS may explain gene tree disparity in our data set is probably minor relative to systematic error (see Supplementary Fig. 2 for four representative topologies).

4) *Does the multi-locus data set counteract the non-stationarity of discordant genes?*—As shown above, concatenation trees estimated under different coding strategies for the complete data set were highly congruent (mean topological disparity = 0.12) and well supported (mean bootstrap values = 82.6%; Figs. 5e and 6d). Unlike most gene trees, the great majority

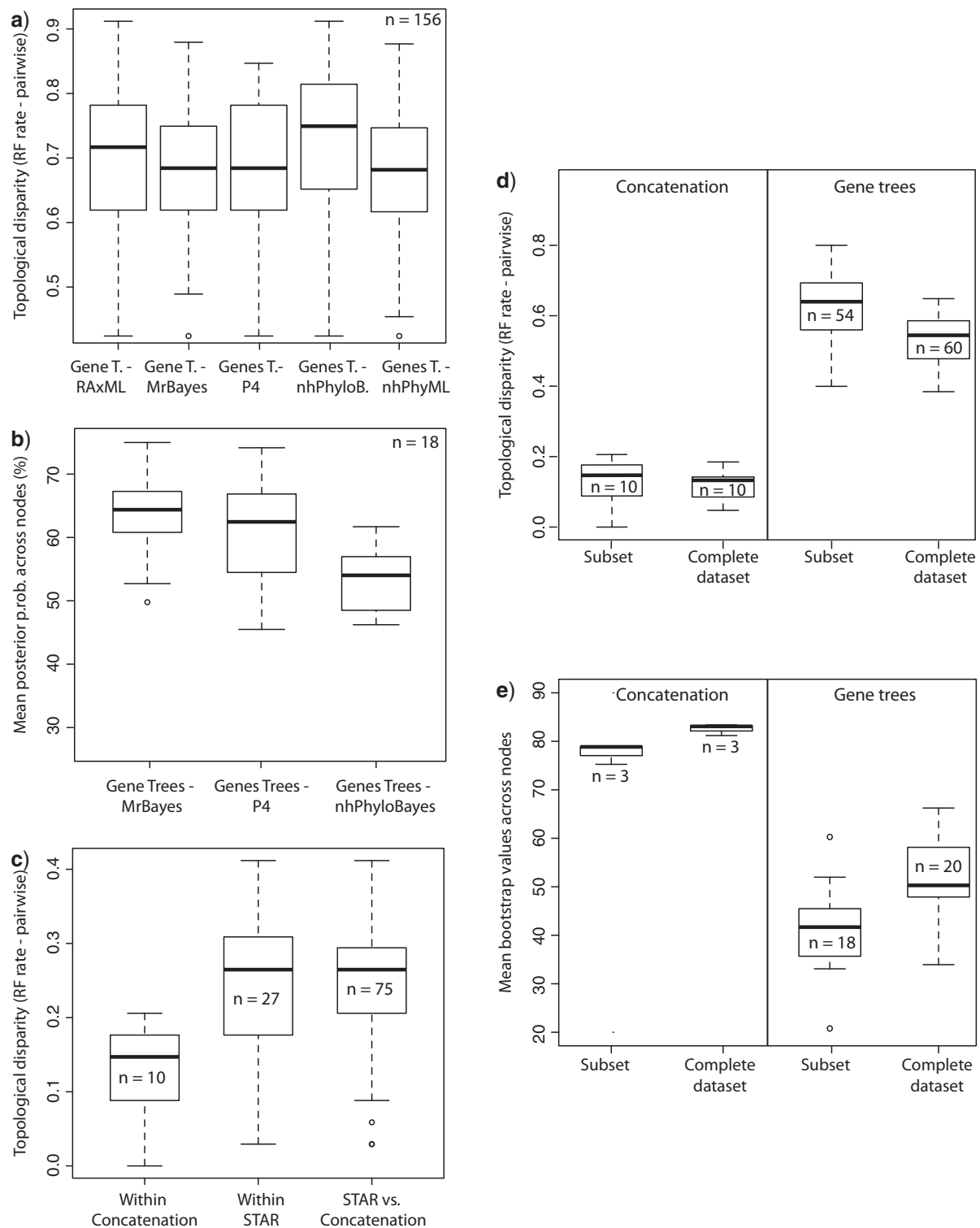


FIGURE 6. Summary of topological discordance and mean nodal support values for phylogenetic trees obtained with different analytical approaches. Whisker plots indicate median, upper and lower quartiles, and ranges; plots (a)–(c) are based on the 37-taxon subset only. a) Pairwise topological distance among gene trees measured by the RF rate (*y*-axis), for trees obtained with different methods (*x*-axis). b) Nodal support measured by average posterior probabilities across nodes (*y*-axis) for gene trees obtained with MrBayes and P4. c) Pairwise topological distance among phylogenetic trees obtained with different coding strategies (nucleotide, AGY, RY) and models (homogenous and non-homogenous) to compare variation within and among concatenation and species tree methods. Panels (d) and (e) illustrate the same approaches as the preceding plots but comparing gene trees and concatenation trees estimated with the subset (37 taxa) and the complete data set (214 taxa); gene trees and bootstrap values were estimated with RAxML; concatenation trees were estimated with RAxML and TreeFinder under different coding strategies.

of concatenation trees support monophyly of the five indicator groups (Table 3).

Relative to the five gene trees with high degree of compositional heterogeneity, ancestral GC content optimization in the concatenation tree (RAxML topology obtained under nucleotide coding) reveals considerably fewer branches with extreme values (Fig. 5 GC content variation among taxa is 17% vs. 19–39 % for individual genes). Most importantly, the few clades (mostly flatfish families) containing relatively high levels of GC, such as bothids, cynoglossids, rhombosoleids, and achiropsettids, appear to adequately group as a result of phylogenetic signal (i.e., most families, as traditionally recognized, are monophyletic) rather than due to convergence in base composition.

By and large, the results obtained with the reduced (37 taxa) data set are similar to those using the full data set, with poorly supported and highly incongruent gene trees but well supported and robust multi-locus trees. However, some differences between the two sampling regimes are evident. First, although most concatenation trees based on the full data set resulted in a monophyletic flatfish group, this clade was not recovered in any of the multi-locus trees estimated with the subset either under concatenation or STAR (Table 3). Second, relative to the gene trees obtained with the 37-taxon data set, topological disparity was slightly lower among gene trees obtained with the complete data set (mean disparity: 0.53 vs. 0.60 for gene trees estimated with the full data set and the subset, respectively; Fig. 6d). Finally, support values were higher in the full data set trees (mean bootstrap values across nodes: 43.0% vs. 52.3% for gene trees, respectively; 79.1% vs. 82.6% for concatenation trees, respectively; Fig. 5e).

Finally, taking the concatenation topology as a reference, we compared likelihood scores between this hypothesis and the topology obtained for each gene with RAxML (Fig. 7; Supplementary Table 4). Likelihood scores were obtained by optimizing each gene partition under nhPhyML for both the gene tree and concatenation topologies using the most general non-stationary model available. The concatenation topology resulted in higher likelihood scores for 9 (of 18) genes, suggesting that our initial nh results for test 2 were compromised by both tree-searching capacity and model misspecification. Thus, non-stationary model optimization favors the concatenation topology for half of individual genes, suggesting that this solution is closer to the optimum than the alternative, and that discordance among gene trees would be much smaller if efficient implementations of nh approaches were available.

Together, these results are consistent with the prediction that genome-wide multi-locus data sets may neutralize the random noise induced by non-stationarity and that the addition of genes and, to some extent, taxa increases signal-to-noise ratio and minimizes systematic error (effectively “converting” systematic error in to sampling error). Although it may be case specific, our phylogenetic reconstructions using the complete data set also suggest that they are robust

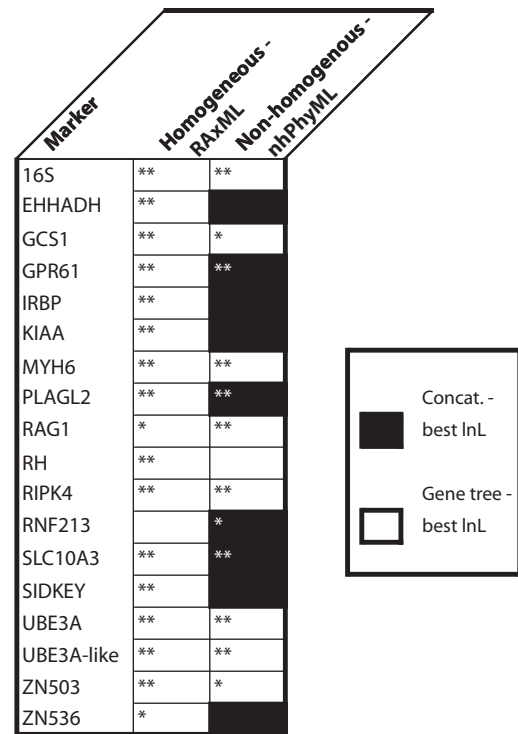


FIGURE 7. Effects of nh model optimization on gene tree inference. For each gene fragment (left column) two topologies were compared using either homogenous models with RAxML (middle column) or nh models in nhPhyML (right column). The topologies compared were the ML gene trees obtained with RAxML (Fig. 2) and the favored phylogenetic hypotheses for the concatenated data set shown in Figure 3 (pruned to include taxa in the 37-taxon subset). Under either optimization (i.e., homogeneous and non-homogenous) we show whether the best likelihood (lnL) score is obtained with the RAxML gene tree (open boxes) or the concatenation topology (black boxes). For all 20 markers the RAxML gene trees resulted in the best lnL scores under homogenous models, whereas under nhPhyML nine markers that optimized the concatenation topology had the best score. Furthermore, in four of those nine cases the score is significantly higher. Further details on this test and full lnL scores are presented in Supplementary Table 4. AU-test * $P = 0.05$; ** $P = 0.01$.

to non-stationarity, even when heterogeneous partitions were analysed under the assumption of homogeneity. For instance, the concatenation trees obtained with AGY- and RY-coding strategies, which show no compositional heterogeneity, are highly similar to those estimated using the uncorrected nucleotide data set (Fig. 6d; Table 3).

DISCUSSION

Our thorough investigation of flatfish phylogeny provides a detailed analysis of potentially adverse effects of base compositional heterogeneity for phylogenetic inference using large multi-locus data sets. Although there are many confounding factors that may exacerbate systematic error, such as model or partition misspecification and heterotachy, it is widely recognized that strong departures from compositional stationarity result in strong phylogenetic biases, and this problem appears to be universal in all major

organismal groups and genome types (Teeling et al. 2000; Jermiin et al. 2004; Phillips et al. 2004; Collins et al. 2005; Gibson et al. 2005; Li and Orti 2007; Song et al. 2010; Nabholz et al. 2011). In spite of this, most studies estimate phylogeny under the assumption of compositional homogeneity without critically assessing whether their data satisfy this condition.

Disregarding recognized issues of non-phylogenetic independence and type II error (Foster 2004), the χ^2 test of PAUP continues to be the most popular approach to diagnose non-stationarity of alignments. Other phylogeny-aware methods have been proposed (e.g., Foster 2004), but these can be computationally demanding, particularly for large data sets with hundreds of taxa. We show that a simple estimation of mean I_D across pairwise comparisons for each marker partition provides a useful metric to rapidly assess their degree of compositional heterogeneity. An obvious advantage of our approach is that it avoids an overly simplistic discretization of the problem (i.e., stationary vs. non-stationary; e.g., Collins et al. 2005). Thus, in combination with the χ^2 test, the I_D provides a powerful tool to rapidly diagnose non-stationarity in large multi-locus data sets (Fig. 1).

We find a strong positive correlation between disparity in gene tree estimation and the degree of compositional heterogeneity (mean I_D) for non-stationary partitions (i.e., third codon positions and 16S; Fig. 4c). For instance, it can be seen that “good” genes such as ZNF536, GCS1, RNF213, KIAA1239, and IRBP, have both the smallest departures from base compositional equilibrium and the smallest topological deviation from the concatenation tree, whereas the opposite is true for RH, MYH6, SLC10A3, UBE3A, and UBE3A-like. This latter group of loci with highest levels of non-stationarity also produces gene trees that show strong evidence of compositional attraction, grouping unrelated taxa with convergent base composition (Fig. 5). These results corroborate the idea that testing for compositional heterogeneity can predict phylogenetic performance of markers (Collins et al. 2005). Thus, given the availability of hundreds to thousands of loci now made possible by genomic resources (e.g., Hittinger et al. 2010; Faircloth et al. 2012; Lemmon et al. 2012), compositional stationarity should be added to the list of criteria to inform marker selection (Jeffroy et al. 2006; Townsend 2007; Lopez-Giraldez and Townsend 2011).

Another recurrent question raised by many authors, with broader implications for phylogenetics, is to determine the number of gene partitions necessary to resolve a particularly hard phylogeny (Ciccarelli et al. 2006; Jeffroy et al. 2006; Rokas and Carroll 2006; Castresana 2007). Emergent phylogenetic support is predicted to grow with the size of the data set and may be quantified by approaches such as random addition concatenation analysis (RADICAL; Narechania et al. 2012). This approach also may be used to find a stable set of genes with generally congruent phylogenetic signal, and to identify outliers that may indicate anomalous or biased historical evidence. In our

case, the non-stationary partitions could fall into this latter category, providing another diagnostic tool for non-stationarity and prompting the need for discarding or recoding genes with conflicting signal (Jeffroy et al. 2006).

Only a fraction of the studies challenged by base compositional heterogeneity implemented data recoding (AGY and RY) and protein translation as a means to mask the bias (Phillips and Penny 2003; Phillips et al. 2004; Gibson et al. 2005; Evans et al. 2010); less frequently, empirical papers have applied non-homogenous models (Galtier and Gouy 1998; Foster 2004; Blanquart and Lartillot 2006; Boussau and Gouy 2006; Gowri-Shankar and Rattray 2007; Dutheil and Boussau 2008). A recent mitogenomic study of 18 beetle taxa found that nh models successfully corrected phylogenetic biases induced by base compositional heterogeneity (Sheffield et al. 2009). Although our necessarily limited implementation of nh approaches for gene tree estimation using the 37-taxon alignments showed no evident improvement relative to traditional approaches using homogeneous models (Fig. 6a,b), optimization of individual gene sequences on the concatenation-tree topology using the most general nh model in nhPhyML resulted in higher likelihood scores for nine out of 18 genes compared to the likelihood values obtained on the RAML-tree topology, four of them significantly higher according to the AU-test (Fig. 7; Supplementary Table 4). These findings suggest that nh models available are computationally demanding and challenging to implement in efficient software applications. Currently available tools for analysis require long running times to converge or are inefficient to correct systematic errors, in particular for large data sets. Future developments to implement nh models based on high performance computing platforms (e.g., Ayres et al. 2012) would be a welcome addition to the phylogenetics toolbox.

The increasing availability of multi-locus sequence data sets has revealed that genealogical histories of individual genes may differ considerably from the underlying organismal phylogeny. Many biological processes have been claimed as the cause of discordance, in particular deep coalescence or ILS (Edwards et al. 2007; Edwards 2009; Kubatko et al. 2009; Liu et al. 2009a, 2009b; Carstens and Dewey 2010). Simulation studies show that coalescent-based species-tree methods may outperform concatenation when ILS is substantial (Carstens and Knowles 2007; Degnan and Rosenberg 2009; Wu 2012). However, their performance or relevance for empirical data sets is not as clear (but see Sen et al. 2012). We have shown that incongruence among gene trees in our data set (Fig. 2) may stem, to a large extent, from systematic error due to base compositional heterogeneity (see above and Figs. 4c and 5; see also Collins et al. 2005; Castresana 2007). To address the effect of ILS, we compared results obtained with concatenated versus species-tree approaches. The topologies estimated with species-tree methods, using input trees obtained with alternative strategies to

account for non-stationarity, were largely similar to those obtained with concatenation (Fig. 6c; Table 3). These findings suggest that the extent to which ILS explains gene tree discordance in our data set is apparently minor relative to non-stationarity.

By extension, the rising popularity of species-tree methods should be tempered with the fact that systematic error, rather than ILS, may be the main factor of discordance, particularly for higher taxonomic-level phylogenetic studies (Huang et al. 2010). Even in the absence of base compositional bias or other systematic artifacts, individual markers often lack sufficient phylogenetic signal to fully resolve gene trees with high support and most tree-based species-tree methods require fully resolved topologies. The fact that DNA-AA analyses resulted in considerable loss of phylogenetic signal implies that our results are dependent on silent-site variation, and also suggests that this example may be an intermediate case between a “low-level” study (where ILS is expected to be a relatively stronger force than base-composition bias) and a “high-level” study, where silent sites are expected to be near enough to substitutional saturation as to be routinely ignored in favor of amino acid analyses (but see Rota-Stabelli et al. 2013). Although some available sequence-based species-tree methods integrate over the gene tree uncertainty (e.g., BEST, *BEAST; Liu 2008; Heled and Drummond 2010), these are inefficient for large-scale data sets and most higher-level phylogenomic studies utilize tree-based methods under the implicit assumption that ILS is the sole cause of discordance (e.g., Faircloth et al. 2012; McCormack et al. 2012; Sen et al. 2012). This is not to say that tree-based species-tree methods are unsuited for phylogenetic reconstruction (see below), but rather that the fundamental assumption that input trees depict the true genealogic history of genes may be unrealistic in many cases.

Finally, our results illustrate the power of combining many independent loci in conjunction with dense taxonomic sampling to address difficult problems, such as base compositional heterogeneity. The cumulative phylogenetic signal of each gene, achieved via concatenation or species-tree approaches, can provide sufficient support for individual nodes to circumvent pervasive but randomly distributed homoplasy (Rokas et al. 2003; Philippe et al. 2011). The addition of taxa also appears to increase both phylogenetic accuracy and signal (Fig. 6d,e; Zwickl and Hillis 2002; Rokas and Carroll 2005; Heath et al. 2008). We also find that trees estimated from the raw nucleotide data sets were highly similar to those inferred with data recoding (Fig. 6d; Table 3), which show no compositional heterogeneity. Thus, the adverse effects of non-stationary partitions appear to have been offset in the analysis of our multi-locus data set. This is expected when convergence in base composition affects gene partitions to different degrees or with different patterns, as seems to be the case in this study (Fig. 5). We note, however, that these results are not necessarily generalizable. Other phylogenomic studies have reported poorly resolved

trees, despite large amounts of data (Rodriguez-Ezpeleta et al. 2007), suggesting that systematic biases (e.g., base compositional heterogeneity) may cause mutual annihilation of phylogenetic and non-phylogenetic signal. Because current implementations of nh models are impractical for large phylogenomic data sets, discarding or recoding non-stationary partitions of the data appears to be the only work-around until more efficient tools are developed. Further studies are required to better understand the adverse effects of base compositional biases in phylogenomics.

Flatfish Monophyly and Interrelationships

The asymmetrical body plan of Pleuronectiformes is certainly one of the most spectacular anatomical specializations among vertebrates. In spite of this, some ichthyologists have challenged the hypothesis of a single evolutionary origin for this unique trait. Kyle 1923 and Chabanaud 1949 regarded flatfishes as polyphyletic, suggesting that “*Psettodes*, the ‘Percoid’, appears to have sprung from a different line of evolution, and is a modern accession to the ranks of Flat-fishes” (Kyle 1923). More recent anatomical studies using modern approaches, however, support the monophyly of flatfishes on the basis of three synapomorphies: 1) cranial asymmetry associated with ocular migration; 2) anterior insertion of the dorsal fin over the cranium; and 3) presence of a recessus orbitalis, a muscular sac-like evagination of the orbit that allows eye protrusion (Hensley and Ahlstrom 1984; Chapleau 1993; Munroe 2005; see Friedman 2012 for a review of other diagnostic features).

Claims disputing the hypothesis of flatfish monophyly appear to receive some support from previous molecular phylogenetic studies (Smith and Wheeler 2006; Li et al. 2009; Li et al. 2011), which fail to group *Psettodes* with the remaining pleuronectiforms. However, most previous studies examined a handful of markers and included a limited taxonomic sampling. In contrast, the rich taxonomic and molecular data set compiled for this study has been designed to test this hypothesis with confidence. Here, the majority of concatenation trees derived from the complete data set (>200 taxa) provide evidence of pleuronectiform monophyly (Fig. 3; Table 3), although nodal support was low in all cases (<50%). These analyses also resolved with high confidence the monophyly of flatfish suborders (*Psettodoidei* and *Pleuronectoidei*; 100% nodal support) and of most families, except *Rhombosoleidae* and *Paralichthyidae*, which were deemed non-monophyletic. Among the latter, two paralichthyid clades, one including the *Cyclopsetta* group and the other containing the remaining taxa (Fig. 3), also were recognized by previous morphological studies (Chapleau 1993; Munroe 2005). Although addressing flatfish interfamilial relationships is outside the scope of this study, our multi-locus topologies are, in this regard, remarkably more congruent with previous morphological phylogenies (e.g., Chapleau 1993; Hensley 1997; Munroe 2005) than

with those inferred with mitochondrial sequence data alone (e.g., Berendzen and Dimmick 2002; Pardo et al. 2005; Azevedo et al. 2008). For instance, the sister-group relationship of Citharidae to all other pleuronectoids is supported by morphological and our multi-locus hypotheses, whereas the mitochondrial phylogenies tend to place this family in a nested position within Pleuronectoidei.

We also note that a minority of the analyses based on the complete data set as well as all those using the 37-taxon set resulted in non-monophyletic Pleuronectiformes, with varying placements for *Psettodes* and pleuronectoids among carangimorphs (Table 3; Supplementary Fig. 2). Preliminary analyses during early stages of the study based on fewer genes (e.g., <11) resulted in similar patterns. Although applying approaches to determine the number of loci or taxa required to obtain flatfish monophyly with strong support is beyond the scope of this study (or the possibilities afforded by our data set), our results emphasize the need for large genome-scale data sets to resolve the so-called “percomorph bush” (Nelson 1989; Smith and Craig 2007; Betancur-R et al. 2013).

Our analyses also confirm recent hypotheses regarding flatfish interrelationships. We find compelling evidence for the monophyly of Carangimorpha sensu Li et al. (2009) or the subseries Carangimorphariae sensu Betancur-R et al. (2013), and provide improved understanding of its taxonomic limits (Fig. 8). In addition to flatfishes, carangimorphs comprise a diverse array of percomorph taxa, whose close relatedness was never detected on the basis of traditional anatomical analyses (e.g., Regan 1912; Norman 1934; Chapleau 1993; Johnson and Patterson 1993; Hensley 1997); these include: Centropomidae sensu Greenwood (snook, barramundi, and Nile perch; see Li et al. 2011), Sphyrinae (barracudas), Toxotidae (archerfishes), Leptobramidae (Beachsalmon), Menidae (moonfishes), Polynemidae (threadfins), Carangoidea sensu Smith-Vaniz 1984; including the families Carangidae, Echeneidae, Coryphaenidae, Rachycentridae, and Nematistidae), and Xiphiidae (swordfish) plus Istiophoridae (billfishes) to the exclusion of scombrids (see also Orrell et al. 2006; Little et al. 2010). Guided by results from molecular studies, two recent studies provided a reappraisal of the morphological evidence supporting Carangimorphariae (Little et al. 2010; Friedman 2012). Little et al. 2010 found that most flatfish, billfish, and carangid taxa share a relatively low number of vertebrae, lack a fusion of the 5th hypural to the uroneural, have multiple pterygiophores inserting before the second neural spine, and lack supraneurals, among others. However, according to (Friedman 2012), some of these characters are symplesiomorphies while others are absent in several carangimorph groups. It thus seems paradoxical that despite the apparent lack of morphological synapomorphies for carangimorphs there is such a strong molecular signal supporting their monophyly, whereas the opposite is true for pleuronectiforms (see above).

Our investigation of carangimorph relationships revealed unexpected results regarding the monophyly and intrarelationships of the order Carangiformes, which comprises the families Nematistidae, Coryphaenidae, Rachycentridae, Echeneidae, and Carangidae (Smith-Vaniz 1984; Wiley and Johnson 2010; Betancur-R et al. 2013). In all analyses, *Nematistius* failed to group with the remaining carangiform families and, except for the RAxML analysis based on the DNA-AA data set (Supplementary Fig. 1), carangids were rendered non-monophyletic, with the tribes Scomberoidini and Trachinotini (Smith-Vaniz 1984; Reed et al. 2002) falling outside (Fig. 3). We suspect that these results are artifacts of the reconstruction given the relatively poor taxonomic representation for carangid taxa in our data set and the short internodes leading to major carangiform groups, a pattern consistent with an evolutionary radiation (Figs. 3 and 7).

Unraveling the sister group of flatfishes in the carangimorph radiation is also a difficult question that remains unanswered. Friedman (2012) inferred a close affinity between centropomids and flatfishes based on a several anatomical features shared by extant members of the centropomid subfamily Latinae and the flatfish fossil *Heteronectes*; e.g., an expansion of the second neural arch and spine, an extension of the lateral line into the caudal fin, and a spur-like processes of the mesethmoid. Despite low nodal support, we note that some of our multi-locus topologies corroborate Friedman’s hypothesis. Finally, given the exuberant diversity in Carangimorphariae and their apparent signature of early lineage accumulation (Fig. 6) it would be interesting to investigate whether this signal is associated with possible morphological innovations and elevated rates of morphological disparity.

CONCLUSIONS

Base composition heterogeneity has been largely neglected in empirical studies as a potential cause of systematic bias in phylogenetic inference. Instead, most recent attention has focused on development of species-tree methods to model deep coalescent events. Accounting for all putative causes of systematic error is challenging but methodical approaches should be established to identify and deal efficiently with the main sources of error in any particular data set. We show a strong association between base compositional bias (non-stationarity) and topological discordance among individual gene partitions and their inferred trees. This problem can be exacerbated in large data sets because current implementations of nh models do not scale up when sampling large number of taxa. In particular cases, however, the combination of phylogenetic information from multiple genetic partitions may help to overcome potential biases present in varying degrees and patterns among single partitions. Finally, the data set analysed for this study provides support for monophyly of flatfishes and their close affinity with other carangimorph groups.

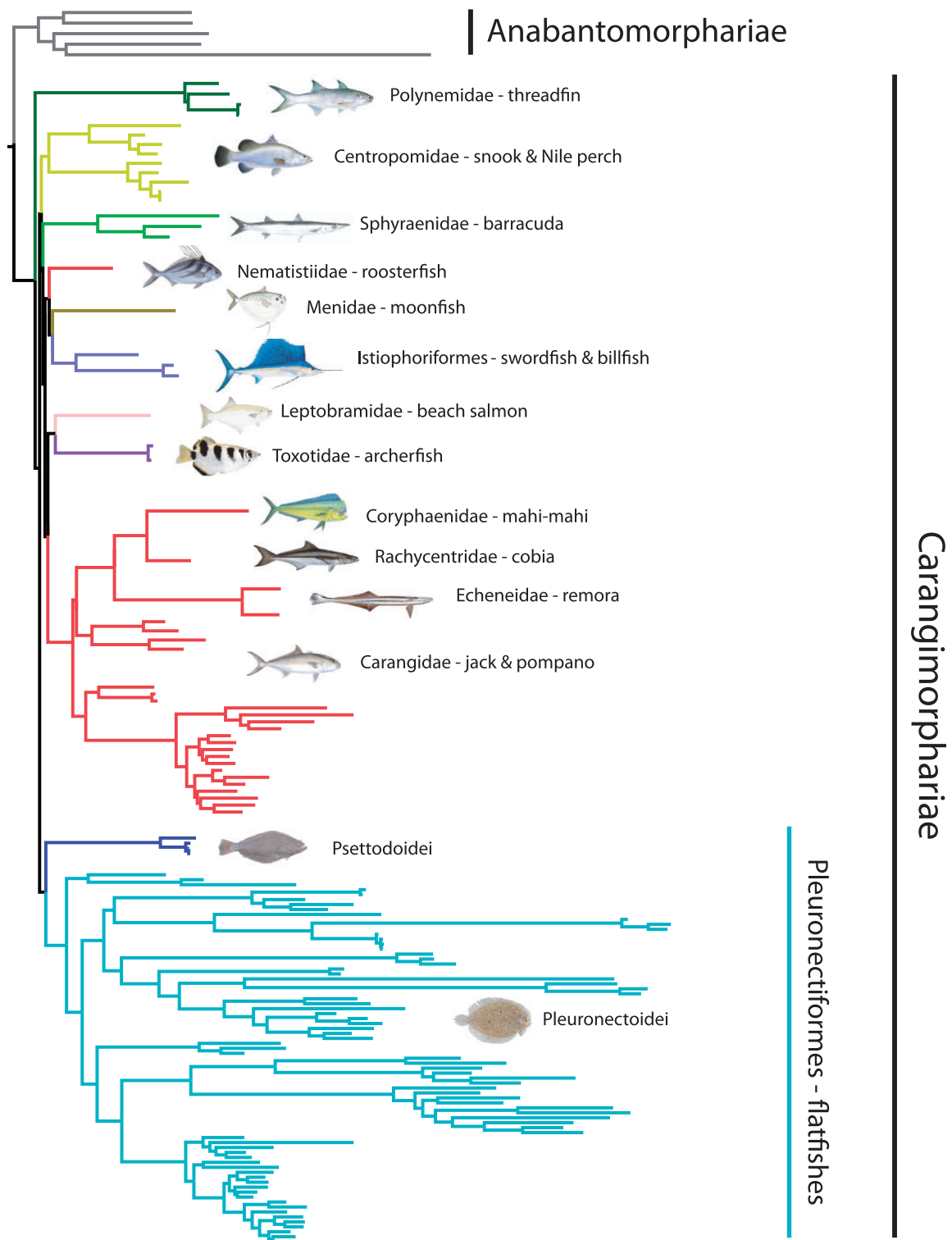


FIGURE 8. Flatfish interrelationships based on the phylogeny estimated with TreeFinder using the GRT3 model for AGY-recorded partitions (see also Fig. 3). Major clades mentioned in the text are indicated (i.e., Anabantomorphariae and Carangimorphariae).

SUPPLEMENTARY MATERIAL

Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository at <http://datadryad.org>, doi:10.5061/dryad.dj31p. See also <http://datadryad.org/review?wfid=11972&token=65b90899-3dd5-492e-8ddf-b0d1156b98d4>.

FUNDING

This research was partially funded by GWU Selective Excellence in Diversity of Life program (to R.B.R.); and the National Science Foundation [grant DEB-1004765 (Euteleost Tree of Life)] to G.O. and C.L.J.

ACKNOWLEDGMENTS

We are grateful to A. Roa, K. Mullaney, S. Willis, C. Borden, D. Arcila, and other members of the Orti lab for providing laboratory assistance. We also thank R. Broughton, K. Carpenter, and members of their labs for contributing sequences that were generated as part of the Euteleost Tree of Life initiative. L. Smith and A. Bentley provided tissue samples. P. Foster and S. Blanquart helped setting the P4 and nhPhyloBayes analyses, respectively. We also thank E. Wiley and L. Smith for insightful discussions. M. Miya, G. Lecointre, an anonymous reviewer, Associate Editor M. Holder, and Editor R. DeBry provided thoughtful suggestions to improve the quality of this study.

REFERENCES

- Abascal F., Zardoya R., Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 21:2104–5.
- Ahlstrom E.H., Amaoka K., Hensley D.A., Moser H.G., Sumida B. 1984. Pleuronectiformes: development. In: Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall A.W. Jr., Richardson S.L., editors. *Ontogeny and systematics of fishes*. Lawrence (KS): Special Publication/American Society of Ichthyologists and Herpetologists. p. 640–70.
- Ayres D.L., Darling A., Zwickl D.J., Beerli P., Holder M.T., Lewis P.O., Huelsenbeck J.P., Ronquist F., Swofford D.L., Cummings M.P., Rambaut A., Suchard M.A. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* 61:170–3.
- Azevedo M.F.C., Oliveira C., Pardo B.G., Martinez P., Foresti F. 2008. Phylogenetic analysis of the order Pleuronectiformes (Teleostei) based on sequences of 12S and 16S mitochondrial genes. *Genet. Mol. Biol.* 31:284–92.
- Baker R.H., DeSalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *Syst. Biol.* 46:654–73.
- Berendzen P.B., Dimmick W.W. 2002. Phylogenetic relationships of Pleuronectiformes based on molecular evidence. *Copeia* 2002: 642–52.
- Betancur-R. R., Broughton R.E., Wiley E.O., Carpenter K., Lopez J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton J., et al. 2013. The tree of life and a new classification of bony fishes. *PLoS Curr. Tree Life*. (in press) doi: 10.1371/currents.tol.53ba26640df0c9ae75bb165c8c26288.
- Blanquart S., Lartillot N. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol. Biol. Evol.* 23:2058–71.
- Boussau B., Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. *Syst. Biol.* 55:756–68.
- Broughton R.E., Betancur-R. R., Li C., Arratia G., Orti G. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLOS Curr. Tree Life*. (in press) doi: 10.1371/currents.tol.2ca8041495ffad0c92756e75247483e.
- Carstens B.C., Dewey T.A. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. *Syst. Biol.* 59:400–14.
- Carstens B.C., Knowles L.L. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Syst. Biol.* 56:400–11.
- Castresana J. 2007. Topological variation in single-gene phylogenetic trees. *Genome Biol.* 8:216.
- Chabanaud P. 1949. Le probleme de la phylogenese des Heterosomata. *Bull. Inst. Oceanogr. (Monaco)*, 950:1–24.
- Chapleau F. 1993. Pleuronectiform relationships - a cladistic reassessment. *Bull. Mar. Sci.* 52:516–40.
- Chen W., Bonillo C., Lecointre G. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Mol. Phylogenet. Evol.* 26:262–88.
- Ciccarelli F.D., Doerks T., von Mering C., Creevey C.J., Snel B., Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–7.
- Collins T.M., Fedrigo O., Naylor G.J. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Syst. Biol.* 54:493–500.
- Conant G.C., Lewis P.O. 2001. Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. *Mol. Biol. Evol.* 18:1024–33.
- Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24:332–40.
- Delsuc F., Phillips M.J., Penny D. 2003. Comment on “Hexapod origins: monophyletic or paraphyletic?”. *Science* 301:1482; author reply 1482.
- Dettai A., Lecointre G. 2005. Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. *C. R. Biol.* 328:674–89.
- Dettai A., Lecointre G. 2008. New insights into the organization and evolution of vertebrate IRBP genes and utility of IRBP gene sequences for the phylogenetic study of the Acanthomorpha (Actinopterygii: Teleostei). *Mol. Phylogenet. Evol.* 48:258–69.
- Dutheil J., Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol. Biol.* 8:255.
- Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63:1–19.
- Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. *Proc. Natl Acad. Sci. USA*. 104:5936–41.
- Evans J., Sullivan J. 2012. Generalized mixture models for molecular phylogenetic estimation. *Syst. Biol.* 61:12–21.
- Evans N.M., Holder M.T., Barbeitos M.S., Okamura B., Cartwright P. 2010. The phylogenetic position of myxozoa: exploring conflicting signals in phylogenomic and ribosomal data sets. *Mol. Biol. Evol.* 27:2733–46.
- Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61:717–26.
- Foster P. 2004. Modeling compositional heterogeneity. *Syst. Biol.* 53:485–95.
- Friedman M. 2008. The evolutionary origin of flatfish asymmetry. *Nature* 454:209–12.
- Friedman M. 2012. Osteology of †*Heteronectes chaneti* (Acanthomorpha, Pleuronectiformes), an Eocene stem flatfish, with a discussion of flatfish sister-group relationships. *J. Vertebr. Paleontol.* 32:735–56.
- Galtier N., Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15:871–9.

- Gibson A., Gowri-Shankar V., Higgs P.G., Rattray M. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Mol. Biol. Evol.* 22:251–64.
- Gowri-Shankar V., Rattray M. 2007. A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. *Mol. Biol. Evol.* 24:1286–99.
- Guindon S., Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.
- Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* 46:239–57.
- Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27:570–80.
- Hensley D.A. 1997. An overview of the systematics and biogeography of the flatfishes. *J. Sea Res.* 37:187–94.
- Hensley D.A., Ahlstrom E.H. 1984. Pleuronectiformes: relationships. In: Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall A.W. Jr., Richardson S.L., editors. *Ontogeny and systematics of fishes*. Special Publication/American Society of Ichthyologists and Herpetologists. p. 670–88.
- Hittinger C.T., Johnston M., Tossberg J.T., Rokas A. 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *PNAS* 107:1476–81.
- Huang H., He Q., Kubatko L.S., Knowles L.L. 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Syst. Biol.* 59:573–83.
- Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22:225–231.
- Jermiin L., Ho S.Y., Ababneh F., Robinson J., Larkum A.W. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.* 53:638–43.
- Jobb G., von Haeseler A., Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* 4:18.
- Johnson G.D., Patterson C. 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. *Bull. Mar. Sci.* 52:554–626.
- Kawahara R., Miya M., Mabuchi K., Lavoue S., Inoue J.G., Satoh T.P., Kawaguchi A., Nishida M. 2008. Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): a new perspective based on whole mitogenome sequences from 75 higher teleosts. *Mol. Phylogenet. Evol.* 46:224–36.
- Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics.* 25:971–3.
- Kumar S., Gadagkar G. 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences (vol 158, pg 1321, 2001). *Genetics* 159:913–4.
- Kyle H.M. 1923. The asymmetry, metamorphosis and origin of flatfishes. *Philos T R Soc Lond B.* 211:75–129.
- Lake J.A. 1994. Reconstructing evolutionary trees from DNA and protein sequences - paralinear distances. *Proc. Natl Acad. Sci. USA.* 91:1455–9.
- Lanfear R., Calcott B., Ho S.Y., Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29:1695–701.
- Lartillot N., Brinkmann H., Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* 7(1 Suppl):S4.
- Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61:727–44.
- Li B., Dettai A., Cruaud C., Couloux A., Desoutter-Meniger M., Lecointre G. 2009. RNF213, a new nuclear marker for acanthomorph phylogeny. *Mol. Phylogenet. Evol.* 50:345–63.
- Li C., Betancur-R. R., Smith W.L., Orti G. 2011. Monophyly and interrelationships of Snook and Barramundi (Centropomidae sensu Greenwood) and five new markers for fish phylogenetics. *Mol. Phylogenet. Evol.* 60:463–71.
- Li C., Lu G., Orti G. 2008. Optimal data partitioning and a test case for ray-finned fishes (actinopterygii) based on ten nuclear Loci. *Syst. Biol.* 57:519–39.
- Li C., Orti G. 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 44:386–98.
- Li C., Orti G., Zhang G., Lu G. 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol. Biol.* 7:44.
- Li C., Orti G., Zhao J. 2010. The phylogenetic placement of siniperid fishes ("Perciformes") revealed by 11 nuclear loci. *Mol. Phylogenet. Evol.* 56:1096–104.
- Little A.G., Loughheed S.C., Moyes C.D. 2010. Evolutionary affinity of billfishes (Xiphiidae and Istiophoridae) and flatfishes (Pleuronectiformes): independent and trans-subordinal origins of endothermy in teleost fishes. *Mol. Phylogenet. Evol.* 56:897–904.
- Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics.* 24:2542–3.
- Liu K., Raghavan S., Nelesen S., Linder C.R., Warnow T. 2009a. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science.* 324:1561–4.
- Liu L., Yu L., Pearl D.K., Edwards S.V. 2009b. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* 58:468–77.
- Lockhart P.J., Howe C.J., Bryant D.A., Beanland T.J., Larkum A.W.D. 1992. Substitutional bias confounds inference of cyanobacterial origins from sequence data. *J. Mol. Evol.* 34:153–62.
- Lockhart P., Novis P., Milligan B.G., Riden J., Rambaut A., Larkum T. 2006. Heterotachy and tree building: a case study with plastids and eubacteria. *Mol. Biol. Evol.* 23:40–5.
- Lockhart P.J., Steel M.A., Hendy M.D., Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* 11:605–12.
- López J.A., Chen W.J., Orti G. 2004. Esociform phylogeny. *Copeia.* 2004:449–64.
- López P., Casane D., Philippe H. 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19:1–7.
- Lopez-Giraldez F., Townsend J.P. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evol. Biol.* 11:152.
- Maddison W.P. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–36.
- McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty R.T., Brumfield R.T., Glenn S.M. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species tree analysis. *Genome Res.* 22:746–54.
- Miya M., Satoh T., Nishida M. 2005. The phylogenetic position of toadfishes (order Batrachoidiformes) in the higher ray-finned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. *Biol. J. Linn. Soc. Lond.* 85:289–306.
- Miya M., Takeshima H., Endo H., Ishiguro N., Inoue J., Mukai T., Satoh T., Yamaguchi M., Kawaguchi A., Mabuchi K., Shirai S., Nishida M. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 26:121–38.
- Munroe T.A. 2005. Chapter 2: Systematic diversity of the Pleuronectiformes. In: Gibson R.N., editor. *Flatfishes: biology and exploitation*. Oxford: Blackwell Science Ltd. p. 10–41.
- Nabholz B., Kunstner A., Wang R., Jarvis E.D., Ellegren H. 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Mol. Biol. Evol.* 28:2197–210.
- Narechania A., Baker R.H., Sit R., Kolokotronis S.O., DeSalle R., Planet P.J. 2012. Random Addition Concatenation Analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. *Genome Biol. Evol.* 4:30–43.
- Nelson G. 1989. Phylogeny of major fish groups. In: Fernholm B., Bremer K., Brundin L., Jörnvall H., Rutberg L., Wanntorp H.E., editors. *The hierarchy of life*. Amsterdam: Elsevier Science. p. 325–36.
- Norman J.R. 1934. A systematic monograph of the flatfishes (Heterosomata). London: British Museum of Natural History.
- Nylander J. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

- Orrell T.M., Collette B.B., Johnson G.D. 2006. Molecular data support separate scombroid and xiphioid clades. *Bull. Mar. Sci.* 79:505–19.
- Orti G., Li C. 2009. Phylogeny and classification. In: Jamieson B.M.G., editor. *Reproductive biology and phylogeny of fish (Agnatha and Osteichthyes)*. Enfield (NH): Science Publishers.
- Pagel M., Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 53:571–81.
- Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–90.
- Pardo B.G., Machordom A., Foresti F., Porto-Foresti F., Azevedo M.F.C., Bañón R., Sánchez L., Martínez P. 2005. Phylogenetic analysis of flatfish (Order Pleuronectiformes) based on mitochondrial 16S rDNA sequences. *Sci. Mar.* 69:531–43.
- Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Worheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9.
- Phillips M.J., Delsuc F., Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* 21:1455–8.
- Phillips M.J., McLenachan P.A., Down C., Gibb G.C., Penny D. 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. *Syst. Biol.* 55:122–37.
- Phillips M.J., Penny D. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* 28:171–85.
- R Development Core Team. 2011. R: a Language and environment for statistical computing.
- Reed D.L., Carpenter K., deGravelle M.J. 2002. Molecular systematics of the Jacks (Perciformes: Carangidae) based on mitochondrial cytochrome b sequences using parsimony, likelihood, and Bayesian approaches. *Mol. Phylogenet. Evol.* 23:513–24.
- Regan C. 1912. The classification of the teleostean fishes of the order *Pediculati*. *Ann. Mag. Nat. Hist.* 8:277–89.
- Rodriguez-Ezpeleta N., Brinkmann H., Roure B., Lartillot N., Lang B.F., Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* 56:389–99.
- Rokas A., Carroll S.B. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol. Biol. Evol.* 22:1337–44.
- Rokas A., Carroll S.B. 2006. Bushes in the tree of life. *PLoS Biol.* 4:1899–1904.
- Rokas A., Williams B.L., King N., Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*. 425:798–804.
- Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–4.
- Rota-Stabelli O., Lartillot N., Philippe H., Pisani D. 2013. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. *Syst. Biol.* 62:121–33.
- Roure B., Philippe H. 2011. Site-specific time heterogeneity of the substitution process and its impact on phylogenetic inference. *BMC Evol. Biol.* 11:17.
- Sen S., Liu L., Edwards S.V., Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *PNAS*. 109:14942–7.
- Sheffield N.C., Song H., Cameron S.L., Whiting M.F. 2009. Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. *Syst. Biol.* 58:381–94.
- Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics (Oxf.)*. 17:1246–7.
- Smith W.L., Craig M.T. 2007. Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percoid fishes. *Copeia*. 2007:35–55.
- Smith W.L., Wheeler W.C. 2006. Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. *J. Hered.* 97:206–17.
- Smith-Vaniz W.F. 1984. Carangidae: relationships. In: Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall A.W. Jr., Richardson S.L., editors. *Ontogeny and systematics of fishes*. Special Publication/American Society of Ichthyologists and Herpetologists. p. 640–70.
- Song H., Sheffield N.C., Cameron S.L., Miller K.B., Whiting M.F. 2010. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Syst. Entomol.* 35:429–48.
- Stamatakis A. 2006. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–90.
- Sul S.J., Williams T.L. 2008. An experimental analysis of Robinson-Foulds distance matrix algorithms. *European Symposium on Algorithms*.
- Swofford D.L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.0 Beta. Sunderland (MA): Sinauer Associates.
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731–9.
- Teeling E.C., Scally M., Kao D.J., Romagnoli M.L., Springer M.S., Stanhope M.J. 2000. Molecular evidence regarding the origin of echolocation and flight in bats. *Nature*. 403:188–92.
- Townsend J.P. 2007. Profiling phylogenetic informativeness. *Syst. Biol.* 56:222–31.
- Van Den Bussche R.A., Baker R.J., Huelsenbeck J.P., Hillis D.M. 1998. Base compositional bias and phylogenetic analyses: a test of the “flying DNA” hypothesis. *Mol. Phylogenet. Evol.* 10: 408–16.
- Weisburg W.G., Giovannoni S.J., Woese C.R. 1989. The *Deinococcus-Thermus* Phylum and the effect of ribosomal-RNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* 11: 128–34.
- Wiley E.O., Johnson G.D. 2010. A teleost classification based on monophyletic groups. In: Nelson J.S., Schultze H.P., Wilson M.V.H., editors. *Origin and phylogenetic interrelationships of teleosts*. München, Germany: Verlag Dr. Friedrich Pfeil. p. 123–82.
- Wu Y. 2012. Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. *Evolution*. 66:763–75.
- Zwickl D.J., Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51:588–98.