Dissecting Incongruence between Concatenation- and Quartet-Based Approaches in Phylogenomic Data

XING-XING SHEN1,2,*, JACOB L. STEENWYK3 AND ANTONIS ROKAS3

1State Key Laboratory of Rice Biology and Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Department of Plant Protection, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China; 2Department of Biological Sciences, Vanderbilt University, VU Station B #35-1634, Nashville, TN 37235, USA

Abstract.—Topological conflict or incongruence is widespread in phylogenomic data. Concatenation- and coalescent-based approaches produce conflicting topologies, but the causes of this conflict are often unclear. Here we dissect incongruence stemming from conflict between the likelihood-based signal (quantified by the difference in gene-wise log-likelihood score or AGLS) and quartet-based topological signal (quantified by the difference in gene-wise quartet score or AQSQ) for every gene in three phylogenomic studies in animals, fungi, and plants, which were chosen because their concatenated IQ-TREE (T1) and quartet-based ASTRAL (T2) phylogenies are known to produce eight conflicting internal branches (bipartitions). By comparing the types of phylogenetic signal for all genes in these three data matrices, we found that 30–36% of genes in each data matrix are inconsistent, that is, each of these genes has a higher log-likelihood score for T1 versus T2 (i.e., AGLS > 0) whereas its T1 topology has lower quartet score than its T2 topology (i.e., AQSQ < 0) or vice versa. Comparison of inconsistent and consistent genes using a variety of metrics (e.g., evolutionary rate, gene tree discordance) showed that inconsistent genes are more likely to recover neither T1 nor T2 and have higher levels of gene tree discordance than consistent genes. Simulation analyses demonstrate that the removal of inconsistent genes from data sets with low levels of incomplete lineage sorting (ILS) and low and medium levels of gene tree estimation error (GTEE) reduced incongruence and increased accuracy. In contrast, removal of inconsistent genes from data sets with medium and high ILS levels and high GTEE levels eliminated or extensively reduced incongruence, but the resulting congruent species phylogenies were not always topologically identical to the true species trees.

Advances in genome sequencing technologies have greatly facilitated the construction of phylogenomic data matrices for inference of species phylogenies from genomic and transcriptomic data (Mishof et al. 2014; One Thousand Plant Transcriptomes Initiative et al. 2019; Shen et al. 2020b). These phylogenomic data matrices are typically analyzed using concatenation- and coalescent-based approaches (Fig. 1a). However, phylogenetic investigations in animals (Prum et al. 2015; Irsari et al. 2017), fungi (e.g., Shen et al. 2016, 2018; Prasanna et al. 2020), and plants (e.g., Wickett et al. 2014; Wu et al. 2018) have reported topologies inferred from concatenation and coalescent-based approaches with conflicting internal bipartitions. For example, 93% of studies published in Systematic Biology in the last 5 years (from July 2015 to June 2020) that used both concatenation-based maximum likelihood and quartet-based ASTRAL approaches reported one or more conflicting internal branches between the species trees generated used the two approaches (Supplementary Table S1 available on Dryad at https://dx.doi.org/10.5061/dryad.9b8c28w5c). The presence of incongruence between concatenation- and coalescent-based approaches poses a big challenge for estimating robust species phylogenies from phylogenomic data (Kubatko and Degnan 2007; Blom et al. 2017; Bravo et al. 2019).

The concatenation-based approach is a “total evidence” approach that combines all gene alignments into a supermatrix (e.g., Rokas et al. 2003), which can then be analyzed by specifying site-homogeneous and site-heterogeneous models using programs such as IQ-TREE (Nguyen et al. 2015), RAxML/RAxML-NG (Stamatakis 2014; Kozlov et al. 2019), PhyML/nhPhyML (Boussau and Gouy 2006; Guindon et al. 2010), MrBayes (Ronquist et al. 2012), RevBayes (Hoehna et al. 2016), ExaBayes (Aberer et al. 2014), and PhyloBayes (Lartillot and Philippe 2004; Lartillot et al. 2009). The major weakness of this approach is its assumption that all genes have the same evolutionary history. This assumption can be violated due to various biological processes that cause gene histories to differ from each other and from the species phylogeny (Degnan and Rosenberg 2009; Edwards 2009; Nakhleh 2013), such as hidden paralogy (e.g., Salichos and Rokas 2011; Rasmussen and Kellis 2012), horizontal gene transfer (HGT) (e.g., Lapierre et al. 2014; Davidson et al. 2015), and incomplete lineage sorting (ILS) (e.g., Mirarab et al. 2016; Scornavacca and Galtier 2017).

The coalescent-based approach employs the multispecies coalescent model to infer the species phylogeny while accounting for the presence of ILS in individual gene trees (Kingman 1982; Maddison 1997;unnamed user on 27 March 2021

© The Author(s) 2021. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
DOI:10.1093/sysbio/syab011
For permissions, please email: journals.permissions@oup.com
Syst. Biol. 0(0) 1–3, 2021
Gene-wise log-likelihood scores (GLS)

<table>
<thead>
<tr>
<th>Gene</th>
<th>GLS</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6321</td>
<td>0.5478</td>
<td>1.5451</td>
</tr>
<tr>
<td>2</td>
<td>0.1857</td>
<td>-1.4538</td>
<td>-1.562</td>
</tr>
<tr>
<td>3</td>
<td>1752</td>
<td>1584</td>
<td>-1988</td>
</tr>
<tr>
<td>4</td>
<td>1562</td>
<td>-2016</td>
<td>-2016</td>
</tr>
<tr>
<td>5</td>
<td>1562</td>
<td>-2016</td>
<td>-2016</td>
</tr>
</tbody>
</table>

Gene-wise quartet scores (GQS)

<table>
<thead>
<tr>
<th>Gene</th>
<th>GQS</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1752</td>
<td>1584</td>
<td>-1988</td>
</tr>
<tr>
<td>2</td>
<td>1562</td>
<td>-2016</td>
<td>-2016</td>
</tr>
<tr>
<td>3</td>
<td>1562</td>
<td>-2016</td>
<td>-2016</td>
</tr>
<tr>
<td>4</td>
<td>1562</td>
<td>-2016</td>
<td>-2016</td>
</tr>
<tr>
<td>5</td>
<td>1562</td>
<td>-2016</td>
<td>-2016</td>
</tr>
</tbody>
</table>

Gene support consistency assessment

Consistent genes: Gene 1, Gene 2, Gene 3
Inconsistent genes: Gene 4, Gene 5

Figure 1. Schematic representation of dissecting incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL approaches in phylogenomic data. a) The concatenation-based IQ-TREE and quartet-based ASTRAL approaches, two widely used approaches for species phylogeny inference from phylogenomic data, often yield species phylogenies (T1 and T2) that differ from each other. b) Calculation of the difference in the gene-wise log-likelihood score (ΔGLS) for T1 versus T2 for each gene based on the supermatrix and substitution model that were used to infer the T1 (left panel); Calculation of the difference in the gene-wise quartet score (ΔGQS) for T1 versus T2 for each gene based on the set of individual gene trees that were used to infer the T2 (right panel). c) Assessment of the consistency of each gene’s support between ΔGLS and ΔGQS. Genes whose ΔGLS > 0 and ΔGQS > 0 (i.e., recovering T1 labeled in blue) and genes whose ΔGLS < 0 and ΔGQS < 0 (i.e., recovering T2 labeled in orange) are considered consistent, whereas genes whose ΔGLS ≤ 0 and ΔGQS ≤ 0 or vice versa (labeled in half-blue and half-orange) are considered inconsistent.
Several previous studies have argued for the relative merits of each approach (e.g., Kubatko and Degnan 2007; Gatesy and Springer 2014; Edwards et al. 2016; Springer and Gatesy 2016; Jiang et al. 2020). Several previous studies have argued for the relative merits of each approach (e.g., Kubatko and Degnan 2007; Gatesy and Springer 2014; Edwards et al. 2016) and evaluation of their relative performance using simulations has yielded mixed results (e.g., Leaché and Rannala 2011; Mirarab et al. 2014). Consequently, most empirical phylogenomic studies employ both and explicitly discuss the observed incongruence between the phylogenies inferred using the two approaches (e.g., Boachon et al. 2018; Steenwyk et al. 2019; Roycroft et al. 2020).

Gene tree variation due to various biological processes across loci (e.g., deep coalescence, gene duplication and loss, and ILS) (e.g., Kubatko and Degnan 2007; Edwards et al. 2016; Bravo et al. 2019; Jiang et al. 2020) and gene tree estimation error due to lack of resolution (e.g., undersampling of informative characters) and multispecies coalescent (MSC) model inadequacy (e.g., recombination within locus) (e.g., Gatesy and Springer 2013; Simmons and Gatesy 2015; Springer and Gatesy 2016) are part of reasons why analyses of phylogenetic data matrices using concatenation- and coalescent-based approaches may give rise to incongruence. However, one potential and little-explored source of this incongruence may be that certain genes exhibit inconsistent support between concatenation- and coalescent-based approaches. In the presence of incongruence between species phylogenies inferred using a concatenation-based approach (IQ-TREE in our case) yielding topology T1 and a coalescent-based approach (the quartet-based ASTRAL approach in our case) yielding topology T2 (Fig. 1a), we define inconsistent genes as those whose gene-wise, likelihood-based signal recovers T1 whereas their gene-wise, quartet-based topological signal recovers T2 or vice versa. In contrast, consistent genes are those whose gene-wise, likelihood-based signal and gene-wise, quartet-based topological signal both recover T1 or both recover T2 (Fig. 1b). Several previous studies have quantified the distribution of the gene-wise phylogenetic signal to dissect incongruence under the concatenation-based maximum likelihood approach (Lee and Hugall 2003; Gatesy and Baker 2005; Evans et al. 2010; Kimball et al. 2013; Shen et al. 2017; Walker et al. 2018; Smith et al. 2020) or under the quartet-based ASTRAL approach (Gatesy et al. 2017, 2019), but little is known about how often these two metrics of gene-wise phylogenetic signal conflict in phylogenomic data matrices (giving rise to inconsistent genes) and the underlying causes of the observed incongruence.

To address these questions, we developed a workflow that can i) quantify the distribution of likelihood-based signal (measured by the difference in gene-wise log-likelihood score or ΔGLS) and quartet-based topological signal (measured by the difference in gene-wise quartet score or ΔGQS) for every gene in a phylogenomic data matrix and ii) identify genes that exhibit inconsistent support between concatenation- and quartet-based approaches (Fig. 1). We applied our workflow in three data matrices from recent phylogenomic studies in animals (Roycroft et al. 2020), fungi (Steenwyk et al. 2019), and plants (Boachon et al. 2018) to determine the proportion of genes in each data matrix that are inconsistent. We then used standard metrics (e.g., evolutionary rate, distribution of branch lengths, hidden paralogy, and gene tree discordance) and four measures of phylogenetic signal (likelihood-based signals, ΔGLS and normalized ΔGLS, and quartet-based topological signal, ΔGQS and normalized ΔGQS) to examine the characteristics of inconsistent and consistent genes. Lastly, we tested whether the removal of inconsistent genes can ameliorate incongruence and increase accuracy in empirical and simulated data matrices.

**MATERIALS AND METHODS**

**Measuring Gene-Wise Likelihood-Based Signal and Quartet-Based Topological Signal**

Our workflow consists of five steps (Fig. 1).

**Step 1:** Our analysis begins when two conflicting topologies, T1 (denoted by blue dot; Fig. 1a) and T2 (denoted by orange dot; Fig. 1a), are reconstructed by concatenation-based IQ-TREE analysis (Nguyen et al. 2013) and quartet-based ASTRAL analysis (Misof et al. 2014), respectively.

**Step 2:** For incongruent internal bipartition(s) between T1 and T2, we define a concatenation-based gene-wise phylogenetic signal as the difference in gene-wise log-likelihood score (ΔGLS) for T1 versus T2 and quartet-based gene-wise phylogenetic signal as the difference in gene-wise quartet score (ΔGQS) for T1 versus T2, respectively (Fig. 1b). The Perl scripts for calculating ΔGLS and ΔGQS for every gene are provided in the Dryad repository.

**ΔGQS estimation (Fig. 1b, left panel):** We first estimate the site-wise log-likelihood values for both T1 and T2, whose branch lengths and substitution model...
parameters were jointly optimized among sites based on the concatenation data matrix and model of sequence evolution used in each of the original studies, with IQ-TREE multithread version 1.6.8 (options “-z” and “-wsl”) (Nguyen et al. 2015). We then calculate the difference in gene-wise log-likelihood scores (ΔAGLS) between T1 and T2 for every gene i in the concatenation data matrix (Lee and Hugall 2003; Gatesy and Baker 2005; Evans et al. 2010; Kimball et al. 2013; Shen et al. 2017; Walker et al. 2018; Smith et al. 2020):

$$\Delta \text{AGLS}_i = \ln L(G_i, T1) - \ln L(G_i, T2).$$  \hspace{1cm} (1)

**ΔGQS estimation** (Fig. 1b, right panel): We first estimate the quartet score (i.e., the number of induced quartets in a single-gene tree that can be mapped on the species phylogeny) for both T1 and T2 using ASTRAL-III version 5.5.11 (options “-s” and “-w”) (Misof et al. 2014; Zhang et al. 2018) for every gene tree (Gatesy et al. 2017, 2019). We then calculate the difference in gene-wise quartet scores (ΔGQS) between T1 and T2 for every gene tree i using the equation:

$$\Delta \text{GQS}_i = \text{QS}(G_i, T1) - \text{QS}(G_i, T2).$$  \hspace{1cm} (2)

**Step 3:** According to equations (1) and (2), ΔAGLS and ΔGQS values can be positive, negative, or zero. We assess the consistency of gene-wise phylogenetic signal calculated by the two measures for every gene (Fig. 1c). Genes whose ΔAGLS > 0 and ΔGQS > 0 (i.e., recovering T1) and genes whose ΔAGLS < 0 and ΔGQS < 0 (i.e., recovering T2) are considered consistent (dots in solid blue or solid orange; Fig. 1c). In contrast, genes whose ΔAGLS ≥ 0 and ΔGQS ≤ 0 or vice versa are considered inconsistent (dots that are half blue and half orange; Fig. 1c).

### Analysis of Three Empirical Phylogenetic Data Matrices

**Data Set Collection.**—To illustrate the utility of our workflow, we used the data matrices from three recent phylogenomic studies in animals (Roycroft et al. 2020), fungi (Steinwyk et al. 2019), and plants (Boachon et al. 2018). Specifically, the animal data matrix was comprised of 1245 exons with an average length of 970 nucleotide sites and gene occupancy of 99% from 37 rodent taxa (Roycroft et al. 2020). The fungal data matrix was comprised of 1034 protein-coding genes with an average length of 506 amino acid sites and gene occupancy of 100% from 25 bipolar budding yeasts and four outgroups (Steinwyk et al. 2019). The plant data matrix was comprised of 520 protein-coding genes with an average length of 1274 nucleotide sites and gene occupancy of 83% from 48 Lamiales species and 4 outgroups (Boachon et al. 2018). The number of incongruent internal branches between the species phylogenies inferred from concatenation-based IQ-TREE analysis and quartet-based ASTRAL analysis in the three data matrices is one (animals), three (fungi), and four (plants) (Table 1). All data matrices, single-gene trees, and species trees are provided in the Dryad repository. A recent study suggested that releasing of the log file of each analysis, which contains a record of the values of all these key parameters (e.g., alignment, program name, number of tree searches, substitution model, number of threads, and random starting seed), can increase the reproducibility of phylogenetic analyses (Shen et al. 2020a). Hence, we also provided the log files in the Dryad repository.

### Identifying Inconsistent Genes Between Concatenation-based IQ-TREE and Quartet-based ASTRAL Approaches

We applied our workflow (Fig. 1) to quantify the distribution of likelihood-based signal and quartet-based topological signal for every gene in three empirical phylogenomic

---

**Table 1.** T1 and T2 hypotheses for the eight incongruent internal branches observed between concatenation- and coalescent-based approaches in three phylogenomic data matrices from animals, fungi, and plants

<table>
<thead>
<tr>
<th>Study</th>
<th>Branch</th>
<th>Concatenation-based tree hypothesis (T1)</th>
<th>Coalescent-based tree hypothesis (T2)</th>
<th>P-value of AU test*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>Coccynys</td>
<td>Coccynys as sister to SHL</td>
<td>Coccynys as sister to Mallomyops + Mammelomyops + Xenaromyops</td>
<td>1.4 × 10⁻²⁴</td>
</tr>
<tr>
<td>Fungi</td>
<td>H. nectariphila</td>
<td>H. nectariphila as sister to H. uvarum</td>
<td>H. nectariphila as sister to H. meyeri + H. clemontiae and H. uvarum</td>
<td>6.35 × 10⁻⁶⁴</td>
</tr>
<tr>
<td></td>
<td>H. uvarum 34-9</td>
<td>H. uvarum 34-9 as sister to H. uvarum</td>
<td>H. uvarum 34-9 as sister to all other three strains of H. uvarum</td>
<td>9.2 × 10⁻³</td>
</tr>
<tr>
<td>Plants</td>
<td>AWR13580</td>
<td>AWR13580 as sister to H. uvarum CBS314</td>
<td>AWR13580 as sister to H. uvarum CBS314 + H. uvarum DSM2768</td>
<td>3.1 × 10⁻³³</td>
</tr>
<tr>
<td></td>
<td>Premnoideae</td>
<td>Premnoideae is paraphyletic</td>
<td>Premnoideae is paraphyletic</td>
<td>2.2 × 10⁻¹</td>
</tr>
<tr>
<td></td>
<td>Peronematideae</td>
<td>Peronematideae as sister to Lamiodeae</td>
<td>Peronematideae as sister to Lamiodeae</td>
<td>1.9 × 10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>Lycopus</td>
<td>Lycopus as sister to Prunella</td>
<td>Lycopus as sister to a clade consisting of Prunella + Nepeta + Agastache + Origanum + Mentha</td>
<td>3.2 × 10⁻³⁴</td>
</tr>
<tr>
<td>Nepeta + Agastache</td>
<td>Nepeta + Agastache as sister to Origanum + Mentha</td>
<td>Nepeta + Agastache as sister to Origanum + Mentha</td>
<td>Nepeta + Agastache as sister to Origanum + Mentha</td>
<td>9.1 × 10⁻³⁴</td>
</tr>
</tbody>
</table>

*For each branch, a constrained ML for T2 hypothesis was inferred using IQ-TREE and can therefore be used to conduct the topological test between T1 and T2 using the approximately unbiased (AU) test (Shimodaira 2002), as implemented in IQ-TREE with 10,000 bootstrap replicates. Asterisks (*) indicate cases in which T1 is significantly different from T2 (P-value < 0.05).
studies (Fig. 2, Supplementary Figs. S1 and S2 available on Dryad).

Examining the Underlying Causes of Inconsistent Genes.—To identify factors that likely contribute to genes with inconsistent support between concatenation-based IQ-TREE and quartet-based ASTRAL approaches, we compared inconsistent and consistent genes in:

1. Ten standard sequence- and tree-based metrics. These metrics are number of taxa in gene alignment, alignment length, percentage of parsimony-informative sites in gene alignment, GC-content rate determined by pairwise sequence similarity in gene alignment, relative composition frequency variability (RCFV) (Phillips and Penny 2003) in gene alignment, average bootstrap support (ABS) in single-gene tree, proportion of sum of internal branch lengths over sum of all branch lengths across single-gene tree (Treeness) (Phillips and Penny 2003), degree of violation of a molecular clock (DVMC) (Liu et al. 2017) in a single-gene tree, and signal-to-noise ratio (ratio of Treeness to RCFV);

2. Two measures of likelihood-based signal and two measures of quartet-based topological signal. These were absolute difference in gene-wise log-likelihood score (absolute ΔGLS) for T1 versus T2, normalized absolute ΔGLS by gene alignment length, the absolute difference in gene-wise quartet score (absolute ΔQQS) for T1 versus T2, and normalized absolute ΔQQS by total number of quartets in single-gene tree (total number of quartets is \( n(n−1)+(n−2)+(n−3)/24 \), where \( n \) is number of tips in the gene tree);

3. Hidden paralogy and gene tree discordance. Hidden paralogy was quantified as ortholog certainty by considering the frequency of the most prevalent ortholog against that of the second most prevalent ortholog, following the concept of internode certainty (Salichos and Rokas 2013; Salichos et al. 2014). In a given gene alignment, we used each sequence to search the reference genome (Rattus norvegicus for the rodent data set; Saccharomyces cerevisiae for the bipolar budding yeast data set; and Rattus norvegicus for the Lamiaceae data set) and kept the reference gene name returned by the best hit. The ortholog certainty is \( 1 + P1\log_2(P1) + P2\log_2(P2) \), where P1 and P2 are fractions of the two most prevalent orthologs. If all sequences identify the same gene in the reference genome as their best hit, ortholog certainty is 1; if the best hits of 50% of sequences are gene A in the reference genome and the best hits of the rest are gene B in the reference genome, ortholog certainty is 0.

Gene tree discordance was measured by the normalized Robinson–Foulds (RF) distance between estimated gene trees and the concatenation-based ML (T1) tree (note that we observed similar levels of gene tree discordance between estimated gene trees and the quartet-based ASTRAL tree (T2)). For estimated gene trees with partial taxon sampling, the T1 was pruned to match the taxon set in the gene tree.

Detailed values of gene properties for every gene are given in Supplementary Tables S2 and S3 available on Dryad.

Using Simulated Data to Examine the Effect of Varying Levels of Gene Tree Discordance on Inconsistent Genes

Gene tree discordance due to incomplete lineage sorting (ILS) and gene tree estimation error (GTEE) could contribute to inconsistent support in phylogenomic data matrices. Since it is challenging to distinguish whether gene tree discordance observed in empirical phylogenomic data matrices is due to ILS or due to GTEE, we simulated data matrices for both ILS and GTEE.

For different ILS levels (low: species tree height = 10M, medium: species tree height = 5M, and high: species tree height = 1M), we simulated 15 data sets (five replicates for each level). Specifically, we simulated the species tree of 100 taxa and one outgroup using SimPhy v1.02 (Mallo et al. 2016) for five replicates for each level, using a different species tree in each replicate. The species trees were simulated under the birth-only process with birth rate \( 10^{-7} \) per generation, a fixed population size of 400,000, and fixed species tree height (10M, 5M, or 1M generations). Trees with higher heights exhibited lower levels of ILS and lower levels of gene tree discordance (average normalized Robinson–Foulds distance between true species tree and true gene trees are 0.16, 0.21, and 0.27 for species tree heights of 10M, 5M, and 1M, respectively; see Supplementary Fig. S3 available on Dryad). For each replicate, we simulated 1000 gene trees under the multispecies coalescent model, which were then used to simulate 1000 gap-free nucleotide gene alignments with randomly varying lengths from 300 to 1500 base pairs using Indelible v1.03 (Fletcher and Yang 2009) under the GTR + G4 model (sequence transition rates, shape for the gamma rate heterogeneity, and equal state frequency are estimated on the animal data set). Detailed commands and parameters are given in Supplementary Text and Table S4 available on Dryad.

For different GTEE levels (low: \( X = 0.1 \), medium: \( X = 0.07 \), and high: \( X = 0.05 \)), we simulated three data sets, each level with one replicate (note that having more replicates is not necessary because we constrained the reference gene tree to have the same topology for all gene alignments). Specifically, we first used the concatenation-based animal ML tree (Fig. 2a) to generate three gene trees with same topology, but different branch lengths, each of which was scaled by \( X \) (\( X = 0.1, 0.07, 0.05 \)), respectively. Next, each gene tree of 37 animal taxa was used to generate 1000 gap-free nucleotide gene alignments with varying length (randomized to be between 300 and 1500 base pairs) using Seq-Gen v1.3.2.
Figure 2. Dissecting incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL approaches in the animal phylogenomic data matrix. a) Concatenation-based IQ-TREE species phylogeny (left) and quartet-based ASTRAL species phylogeny (right) were inferred by analysis of 1245 exons (1,207,638 nucleotide sites) from 37 rodent taxa (Roycroft et al. 2020). Branch support values denote rapid bootstrap support (BS) and local posterior probability (LPP), respectively. Only support values smaller than 100% are shown. One conflicting internal branch (SHL [Sahul Hydromyini excluding Antimyza, early branching New Guinea, and Coccymys] clade shown in red) between concatenation-based IQ-TREE (T1) and quartet-based ASTRAL (T2) is found. b) Distribution of \( \Delta GLS \) and \( \Delta GQS \) for T1 (blue bars) and T2 (orange bars) across 1245 genes. \( \Delta GLS \) (above y-axis) and \( \Delta GQS \) (below y-axis) values were calculated by measuring the difference in gene-wise log-likelihood scores and the difference in gene-wise quartet scores for T1 versus T2, respectively. The number of genes that exhibited consistent support between \( \Delta GLS \) and \( \Delta GQS \) measures is 794 (64%), while the number of genes that exhibited inconsistent support between two measures is 451 (36%). Dissecting incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL approaches in fungal and plant phylogenomic data matrices are given in Supplementary Figs. S1 and S2 available on Dryad.)
under the GTR+G4 model (sequence transition rates, shape for the gamma rate heterogeneity, and equal state frequency are estimated on the animal data set).

For each simulated data matrix, we inferred the concatenation-based IQ-TREE species tree under a single GTR+G4 model using IQ-TREE multi-thread version 1.6.8 (Nguyen et al. 2015). We also inferred the quartet-based species tree with ASTRAL-III version 5.5.11 (Misof et al. 2014; Zhang et al. 2018) using the set of individual ML gene trees inferred using IQ-TREE multi-thread version 1.6.8 with options “–runs 10 -nt 2 -st DNA -m GTR+G4+F -me 0.0001 -bb 1000”. The reliability of each internal branch was evaluated using 100 ultrafast bootstrap replicates (Minh et al. 2013) and local posterior probability (LPP) (Sayyari and Mirarab 2016), in the concatenation- and quartet-based species trees, respectively. If the concatenation-based IQ-TREE species tree topologically differed from the quartet-based ASTRAL species tree, we applied the workflow (Fig. 1) to quantify the distribution of likelihood-based signal and quartet-based topological signal for every gene and identify inconsistent genes. The characteristics of ten standard gene- and tree-based metrics (e.g., evolutionary rate, distribution of branch lengths) for every simulated gene were examined as described above and are provided in the Supplementary Tables S5 and S6 available on Dryad.

RESULTS AND DISCUSSION

Incongruence between Concatenation-Based IQ-TREE and Quartet-Based ASTRAL Phylogenies in Phylogenomic Studies

We analyzed three empirical phylogenomic matrices (Boachon et al. 2018; Steenwyk et al. 2019; Roycroft et al. 2020) and found that all concatenation-based IQ-TREE (T1) and quartet-based ASTRAL (T2) phylogenies reported were reproducible, albeit with slightly different branch support values (Fig. 2a, Supplementary Figs. S1a and S2a available on Dryad). Comparison of each concatenation-based IQ-TREE phylogeny with its corresponding quartet-based ASTRAL phylogeny showed the presence of one, three, and four incongruent internal branches in the animal, fungal, and plant studies (see Fig. 2a, Supplementary Figs. S1a and S2a available on Dryad), respectively. Furthermore, we found that T1 hypotheses were significantly different from T2 hypotheses under the approximately unbiased (AU) test (Shimodaira 2002) implemented in IQ-TREE for 6/8 incongruent internal branches (Table 1). The phylogenetic relationships for the eight incongruent internal branches observed between the two approaches are given below.

Animal data set (Roycroft et al. 2020). The concatenation-based IQ-TREE species phylogeny recovered the genus Coccymys as sister to SHL (Sahul Hydromyini excluding Anisomys, early branching New Guinea, and Coccymys) with high bootstrap value (BS =98), while the quartet-based ASTRAL phylogeny recovered the genus Coccymys as sister to the genera Mallomys, Mammelomys, and Xenuromys with high local posterior probability (LPP =0.95) (Fig. 2a).

Fungal data set (Steenwyk et al. 2019). We observed three conflicting branches between concatenation-based IQ-TREE and quartet-based ASTRAL phylogenies (Supplementary Fig. S1a available on Dryad). The first concerned the placement of Hanseniaspora nectarophila; the concatenation-based IQ-TREE tree recovered H. nectarophila as sister to a clade of four different strains (CBS304, AWRI3580, 34-9, and DSM2768) of Hanseniaspora uvarum (BS =100), whereas the quartet-based ASTRAL phylogeny recovered H. nectarophila as sister to H. meyeri + H. cleromiae and H. uvarum (LPP =1.00). The other two conflicting branches concerned the relative placement of four different strains (CBS304, AWRI3580, 34-9, and DSM2768) of H. uvarum and exhibited moderate support values (BS =87; LPP =0.69) (Supplementary Fig. S1a available on Dryad).

Plant data set (Boachon et al. 2018). We observed four conflicting branches between concatenation-based IQ-TREE and quartet-based ASTRAL phylogenies (Supplementary Fig. S2a available on Dryad). The first concerned the paraphyly of the subfamily Premnoideae recovered by the concatenation-based IQ-TREE phylogeny (BS =76), while the quartet-based ASTRAL phylogeny strongly supported the monophyly of Premnoideae (LPP =0.99). The second concerned the subfamily Perenomatoideae, which was recovered as the sister group to either Lamioideae + Ajugoideae in the concatenation-based IQ-TREE phylogeny (BS =100) or to Lamioideae in the quartet-based ASTRAL phylogeny (LPP =0.72). The third and fourth concerned a clade consisting of Lycopus and Prunella (BS =99) as the sister group to a clade consisting of Nepeta + Agastache + Hyssopus + Glechoma and Origanum + Thymus + Mentha + Monarda (BS =72) in the concatenation-based IQ-TREE phylogeny, while the quartet-based ASTRAL phylogeny recovered the lineage Lycopus as early branching (LPP =0.9) and the genus Prunella as sister group to a clade consisting of Origanum + Thymus + Mentha + Monarda (LPP =0.95) (Supplementary Fig. S2a available on Dryad).

Approximately One-Third of Genes Exhibit Inconsistent Behavior between Concatenation-Based IQ-TREE and Quartet-Based ASTRAL Approaches

For each of three empirical phylogenomic data matrices from animals (Roycroft et al. 2020), fungi (Steenwyk et al. 2019), and plants (Boachon et al. 2018), we applied our workflow (Fig. 1) to quantify the distribution of concatenation-based IQ-TREE phylogenetic signal (i.e., the difference in gene-wise quartet score [ΔGQS] for T1 vs. T2) and quartet-based ASTRAL phylogenetic signal (i.e., difference in gene-wise quartet score [ΔGQS] for T1 vs. T2) for
every gene. \(\Delta GLS\) and \(\Delta GQS\) values for every gene in the three phylogenomic data matrices are given in Supplementary Table S2 available on Dryad.

Our distributions showed that the proportion of genes (54\% in animals; 53\% in fungi; 55\% in plants) recovering the concatenation-based IQ-TREE phylogeny (T1 in blue) is generally higher than that of genes (46\% in animals; 47\% in fungi; 45\% in plants) recovering quartet-based ASTRAL phylogeny (T2 in orange) when analyzed in a concatenation-based IQ-TREE framework (Fig. 2b, Supplementary Figs. S1b and S2b available on Dryad). In contrast, the proportion of genes (51\% in animals; 54\% in fungi; 54\% in plants) recovering the quartet-based ASTRAL phylogeny (T2 in orange) is generally higher than that of genes (49\% in animals; 46\% in fungi; 46\% in plants) recovering the concatenation-based IQ-TREE phylogeny (T1 in blue) when analyzed in a quartet-based ASTRAL framework (Fig. 2b, Supplementary Figs. S1b and S2b available on Dryad).

Examination of the distribution of gene-wise phylogenetic signal between concatenation-based IQ-TREE and quartet-based ASTRAL approaches showed that 794/1245 (64\%) genes in animals, 683/1034 (66\%) genes in fungi, and 363/520 (70\%) genes in plants were consistent, that is their \(\Delta GLS\) and \(\Delta GQS\) values had the same signs, while 451/1245 genes (36\%) in animals, 351/1034 genes (34\%) in fungi, and 157/520 genes (30\%) in plants were inconsistent, that is their \(\Delta GLS\) and \(\Delta GQS\) values had opposite signs (Fig. 2b, Supplementary Figs. S1b and S2b available on Dryad). Interestingly, we found that proportions of inconsistent genes with \(\Delta GLS > 0\) (recovering T1) and \(\Delta GQS < 0\) (recovering T2) are generally lower than those of inconsistent genes with \(\Delta GLS < 0\) (recovering T2) and \(\Delta GQS > 0\) (recovering T1) in three empirical data matrices (animals: 196/451 genes [43\%] vs. 255/451 genes [57\%]; fungi: 142/351 genes [40\%] vs. 209/351 genes [60\%]; plants: 55/157 genes [35\%] vs. 102/157 genes [65\%]). However, we found that proportions of inconsistent genes with \(\Delta GLS > 0\) and \(\Delta GQS < 0\) are not always lower than those of inconsistent genes with \(\Delta GLS < 0\) and \(\Delta GQS > 0\) in different types of simulated data matrices (on average, simulated data matrices with different ILS levels: 642/1234 genes [52\%] vs. 592/1234 genes [48\%]; simulated data matrices with different GTEE levels: 111/1162 genes [10\%] vs. 1051/1162 genes [90\%]). Last, compared to inconsistent genes, consistent genes had significantly higher (Wilcoxon rank-sum test) values of absolute \(\Delta GLS\), normalized absolute \(\Delta GLS\), absolute \(\Delta GQS\), and normalized absolute \(\Delta GQS\) in all three empirical phylogenomic data matrices (Fig. 3a).

**Genes that Exhibit Inconsistent Behavior between Concatenation- and Quartet-Based Approaches are More Likely to Recover Neither T1 Nor T2**

By examining the support for individual unconstrained ML gene trees, we found that proportions of inconsistent genes recovering T1 (concatenation-based IQ-TREE tree) or T2 (quartet-based ASTRAL tree) were generally lower than those of consistent genes recovering T1 or T2 in five out of eight conflicting internal branches (Fig. 4a). For the remaining three conflicting internal branches, their proportions of inconsistent genes recovering T1 or T2 are similar to or slightly higher than those of consistent genes recovering T1 or T2. Specifically, proportions of inconsistent genes recovering T1 or T2 are 0.30 for the branch *Coccymys* in animals, 0.47 for the branch *H. nectarophila* in fungi, 0.47 for the branch *H. varum* AWR13580 in fungi, 0.18 for the branch Premnoideae in plants, 0.17 for the branch Peronematoideae in plants, 0.36 for the branch *Lycopus* in plants, and 0.31 for the branch *Nepeta + Agastache* in plants, while proportions of consistent genes recovering T1 or T2 are 0.57 for the branch *Coccymys* in animals, 0.50 for the branch *H. nectarophila* in fungi, 0.46 for the branch *H. varum* AWR13580 in fungi, 0.26 for the branch Premnoideae in plants, 0.35 for the branch *Lycopus* in plants, and 0.44 for the branch *Nepeta + Agastache* in plants. These results suggest that inconsistent genes are more likely to recover neither T1 nor T2.

In addition to unconstrained ML gene trees, we also constrained ML gene trees to the topologies T1 or T2 and examined the distributions of the lengths of eight conflicting internal branches between inconsistent genes and consistent genes. For each of the eight conflicting internal branches (see Table 1), we calculated the length of its corresponding internal branch with respect to T1 when we constrained and optimized a single-gene tree to recover T1 (note that we observed a similar pattern of internal branch lengths between inconsistent genes and consistent genes when we constrained individual gene trees to recover T2; see Supplementary Fig. S4 available on Dryad). Our results show that in seven out of eight conflicting branches, the average length of internal branches of inconsistent genes with respect to T1 are generally shorter than those of consistent genes (Fig. 4b). Specifically, we found that inconsistent genes exhibited a 4.44-fold shorter length for the branch *Coccymys* in animals (on average, 0.00036 substitutions per site across inconsistent genes vs. 0.00071 substitutions per site across consistent genes), 1.08-fold shorter length for the branch *H. nectarophila* in fungi (0.0071 vs. 0.0077), 1.28-fold shorter length for the branch *H. varum* 34-9 (0.00025 vs. 0.00032), 1.05-fold longer length for the branch *H. varum* AWR13580 (0.00061 vs. 0.00058), 1.15-fold shorter length for the branch Premnoideae in plants (0.0019 vs. 0.0022), 1.03-fold shorter length for the branch *Lycopus* in plants (0.0062 vs. 0.0063), and 1.18-fold shorter length for the branch *Nepeta + Agastache* in plants (0.0022 vs. 0.0026), compared with consistent genes. Our results suggest that inconsistent genes tend to recover neither T1 nor T2, and
Gene Tree Discordance Likely Contributes to Inconsistent Behavior between Concatenation- and Quartet-Based Approaches

Phylogenomic inference is a linear workflow that includes a series of separate steps (e.g., data sampling, genome/transcriptome assembly, orthology identification, multiple sequence alignment, alignment trimming, model selection, phylogenetic inference, and sensitivity analyses) (Anisimova et al. 2013; Guang et al. 2016; Philippe et al. 2017), where each step relies on previous steps and influences subsequent ones. Each step can introduce noise or bias, such as sequence contaminants (e.g., Laurin-Lemay et al. 2012), alignment error (e.g., Di Franco et al. 2019), uncertainty in trimming (e.g., Tan et al. 2015), mis-specified model parameters (e.g., Brown and Thomson 2018; Yang and Zhu 2018), and suboptimal tree search parameters (e.g., Zhang et al. 2018; Shen et al. 2020a). In addition to these sources of noise and bias in the linear workflow, hidden paralogy and gene tree discordance (Degnan and Rosenberg 2009; Nakhleh 2013; Liu et al. 2015b) can also contribute to incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL approaches.
Figure 4. Characteristics of individual gene trees in three phylogenomic data sets. Examination of the distribution of gene-wise support values between concatenation-based IQ-TREE and quartet-based ASTRAL approaches showed that 451 genes in animals, 351 genes in fungi, and 157 genes in plants were inconsistent, while 794 genes in animals, 683 genes in fungi, and 363 genes in plants were consistent. For each of eight conflicting internal branches (see Table 1), a) we checked each ML gene tree’s support: T1: (O, (A, (B, C))) (i.e., concatenation-based IQ-TREE tree in blue), T2: (O, ((A, B), C)) (i.e., quartet-based ASTRAL tree in orange), T3: (O, ((A, C), B)) (i.e., the third hypothesis in black), or others in gray. We found that in five out of eight conflicting branches proportions of inconsistent genes recovering T1 or T2 are generally lower than those of consistent genes recovering T1 or T2. b) We also constrained single-gene trees to recover T1 and calculated the length of their corresponding internal branches with respect to T1. Note that we observed similar patterns of internal branch lengths between inconsistent and consistent genes when we constrained them to recover T2 (see Supplementary Fig. S4 available on Dryad). We found that in seven out of eight conflicting branches the average length of internal branches of inconsistent genes optimized under the T1 constraint are generally shorter than those of internal branches of consistent genes when we constrained them to recover T1 or T2. The Wilcoxon rank-sum test was used to test if inconsistent and consistent genes exhibited significantly different patterns (**P value ≤ 0.01; *P value ≤ 0.05; NS = not significant).
To explore factors that likely contribute to different gene-wise support values between 959 inconsistent genes and 1840 consistent genes from the three empirical phylogenomic data sets, we extensively examined ten sequence- (e.g., GC content, evolutionary rate) and tree-based (e.g., distribution of branch lengths or Transness, degree of violation of a molecular clock or DVMC) metrics, hidden paralogy, and gene tree discordance for every gene (see Materials and Methods for details of calculation for each metric).

Comparisons of inconsistent genes and consistent genes across ten sequence- and tree-based metrics showed that inconsistent genes did not exhibit any significant differences in ten sequence- and tree-based metrics than consistent genes in the fungal and plant phylogenomic data matrices (Fig. 3b), but not in the animal data matrix. For example, we observed very similar patterns of GC content, evolutionary rate, Transness, and DVMC between inconsistent and consistent genes (on average, GC content: 32.5% vs. 31.7%; evolutionary rate: 85% vs. 84.5%; Transness: 0.360 vs. 0.365; DVMC: 0.13 vs. 0.14) (Fig. 3b and Supplementary Table S2 available on Dryad). In the animal data matrix, inconsistent genes exhibited significant differences in three of the ten metrics (alignment length, percentage of parsimony-informative sites in gene alignment, and average bootstrap support, which are associated with gene informativeness) compared to consistent genes (Fig. 3c).

We also did not identify significant differences in the level of hidden paralogy between the 959 inconsistent genes and the 1840 consistent genes (on average, ortholog certainty: 0.90 vs. 0.89) (Fig. 3a; Supplementary Table S3 available on Dryad). In contrast, we found that inconsistent genes had significantly higher levels of gene tree discordance (measured by the normalized RF distance between estimated gene trees and the concatenation-based IQ-TREE tree; note that we observed a similar level of gene tree discordances between estimated gene trees and quartet-based ASTRAL tree) than consistent genes in the animal and fungal phylogenomic data matrices (on average, normalized RF distance in the animal data set: 0.35 vs. 0.51; normalized RF distance in the fungal data set: 0.28 vs. 0.25), but not in the plant data matrix (on average, normalized RF distance in the plant data set: 0.373 vs. 0.369) (Fig. 3b). These results suggest that gene tree discordance is likely a contributor to inconsistent genes.

Several recent studies have shown that one or a few outlier genes with the very strong phylogenetic signal can have a major influence on the results of concatenation-based phylogenetic inference (e.g., Di Franco et al. 2019; Shen et al. 2017; Walker et al. 2018, 2020) and quartet-based ASTRAL phylogenetic inference (Gatesy et al. 2017, 2019). However, examination of the distributions of ΔGLS and ΔCQS values in the three phylogenomic data matrices suggests that none of the data matrices contains any obvious outlier genes (Supplementary Table S2 available on Dryad).

The Effect of Varying Levels of Gene Tree Discordance on Incongruence between Concatenation- and Quartet-Based Approaches

Since gene tree discordance is a likely contributor to inconsistent genes between concatenation-based IQ-TREE and quartet-based ASTRAL approaches in three empirical phylogenomic data sets, we used simulations to examine the effect of different levels of gene tree discordance due to ILS or GTEE (Degnan and Rosenberg 2009; Liu et al. 2015a; Roch and Warnow 2015; Mirarab et al. 2016; Springer and Gatesy 2016) on incongruence between the two approaches (see Materials and Methods for details). Because it is very challenging to distinguish whether gene tree discordance is due to ILS or GTEE, we simulated data matrices for both ILS and GTEE. For ILS, the simulated true species trees and true gene trees are different in topology and branch length. For GTEE, we first
simulated 1000 DNA gene alignments under the same topology and the same branch lengths based on the concatenation-based animal ML tree (Fig. 2a), and found that inferred concatenation-based ML and quartet-based ASTRAL trees both recovered the true species tree (Fig. 2a), which means that no inconsistent genes could be identified between the two approaches (See scenario 1X in Table 3). Hence, we examined the simulations under the same topology but different branch lengths.

After applying the workflow to quantify the distribution of likelihood-based signal ($\Delta$GLS) and quartet-based topological signal ($\Delta$Q5) for every gene for each simulated data matrix, we found that:

1. The fraction of inconsistent genes between concatenation-based IQ-TREE and quartet-based ASTRAL approaches increased with increasing levels of ILS and GTEE (on average, the numbers of inconsistent genes were 398/1000 (39.8%), 412/1000 (41.2%), and 423/1000 (42.3%) for species trees with low, medium, and high levels of ILS, respectively; the numbers of inconsistent genes were 349/1000 (34.9%), 384/1000 (38.4%), and 429/1000 (42.9%) for low, medium, and high levels of GTEE, respectively; Fig. 6a and d),

2. Inconsistent genes had higher levels of gene tree discordance than consistent genes and their gene tree discordances increased with increasing levels of ILS and GTEE (average normalized Robinson–Foulds distances between estimated gene trees and true gene tree were 0.30, 0.34, and 0.38 for species trees with low, medium, and high levels of ILS, respectively; average normalized Robinson–Foulds distances between estimated gene trees and true gene tree were 0.83, 0.87, and 0.90 for low, medium, and high levels of GTEE, respectively; see Fig. 6b and e),

3. the total number of conflicting branches between the species phylogenies inferred by concatenation-based IQ-TREE and quartet-based ASTRAL approaches increased with increasing levels of ILS and GTEE (total numbers of conflicting branches were 4, 6, and 7 for species trees with low, medium, and high levels of ILS, respectively; total numbers of conflicting branches were 4, 8, and 9 for low, medium, and high levels of GTEE, respectively; see Fig. 6c and f).

Examination of 10 sequence- and tree-based metrics in the simulated data matrices show that most of the 10 metrics exhibited similar characteristics between inconsistent and consistent genes. Specially, in ILS-simulated data matrices, none of the ten metrics (except average bootstrap support in the medium ILS data set) exhibited significant differences between inconsistent and consistent genes (Supplementary Fig. S5 and Table S5 available on Dryad). In GTEE-simulated data matrices, inconsistent and consistent genes exhibited significant differences in three out of the ten metrics (gene alignment length, percentage of parsimony-informative sites in gene alignment, and Treeenss) (Supplementary Fig. S6 and Table S6 available on Dryad).

How Should Inconsistent Genes be Handled in Phylogenomic Analyses?

Having identified the sets of genes that exhibited inconsistent support between concatenation-based IQ-TREE and quartet-based ASTRAL approaches in both empirical and in simulated phylogenomic data matrices, we next tested whether removal of inconsistent genes can ameliorate the observed incongruence.

In empirical data matrices, we found that removal of inconsistent genes eliminated the one incongruent branch in the animal data set, all four incongruent branches in the plant data set, as well as reducing the number of incongruent branches from three to one in the fungal data set (Fig. 7, Supplementary Figs. S7 and S8 available on Dryad). Among the seven eliminated incongruent branches, five recovered T1 (animal data set: Cocymys; fungal data set: H. nectarophila and H. urarum; plant data set: Premnoideae and Peronematoideae) and two recovered T2 (plant data set: Lycopus and Nepeta + Agastache). For example, the removal of 451 inconsistent genes eliminated the incongruence between the phylogenies obtained using concatenation-based IQ-TREE and quartet-based ASTRAL approaches on the full data matrix (Fig. 2).

Following inconsistent genes removal, concatenation-based IQ-TREE and quartet-based ASTRAL approaches placed the genus Cocymys as the sister group to the SHL clade (T1). Although the removal of inconsistent genes eliminated or extensively reduced incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL approaches in all three empirical data matrices, we do not know whether the congruent species phylogenies inferred from these reduced data matrices are closer to the true species trees.

Thus, we next used the simulated data matrices to test whether the removal of inconsistent genes both ameliorated the observed incongruence and yielded congruent phylogenies that are closer to the true species trees. In simulated data matrices with low levels of ILS and low and medium levels of GTEE (Fig. 6c and f and Tables 2 and 3), we found that the removal of inconsistent genes eliminated incongruence and recovered the true species trees. However, in simulated data sets with medium and high levels of ILS and high levels of GTEE, the removal of inconsistent genes eliminated or extensively reduced incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL trees, but these congruent IQ-TREE and ASTRAL phylogenies inferred from the reduced data matrices were not always topologically identical to the true species trees (Fig. 6c and f and Tables 2 and 3). Consistent with previous results (Mollay and Warnow 2018), concatenation is more
accurate when GTEE levels are high and ILS levels are low, whereas coalescence is better when GTEE levels are low and ILS levels are high. If both GTEE and ILS levels are high (or low), the two approaches have similar performance (Tables 2 and 3). Taken together, these results suggest that the removal of inconsistent genes may be helpful in phylogenies with low levels of ILS and GTEE but more problematic in the presence of high levels of ILS and GTEE.

Given that the level of gene tree discordance due to ILS and GTEE is positively correlated with the number of inconsistent genes between concatenation- and quartet-based approaches on simulated data matrices, we make the following recommendations for the handling of inconsistent genes in empirical phylogenomic studies: i) Estimation of ILS: accurately estimating ILS is very challenging, but we can adopt an alternative way that quantifies the lengths of conflicting internal branches in single-gene trees. If inconsistent genes exhibit substantially shorter internal branches (e.g., due to a rapid radiation) than consistent genes, we recommend removing these inconsistent genes and
Figure 7. Removal of inconsistent genes eliminates the incongruence between the concatenation-based IQ-TREE species tree and the quartet-based ASTRAL species tree on the animal phylogenomic data set. The concatenation-based IQ-TREE species tree (left panel) and the quartet-based ASTRAL species tree (right panel) were inferred using a data set of the 794 genes that exhibited consistent support between $\Delta$GLS and $\Delta$GQS measures. Only support values smaller than 100% are shown. Our results show that removal of 451 inconsistent eliminates the incongruence observed between the concatenation- and quartet-based phylogenies when the full data matrix is analyzed (Fig. 2). Phylogenies inferred using reduced fungal and plant data matrices are provided in Supplementary Figs. S7 and S8 available on Dryad.

Table 2. Summary of numbers of conflicting branches between concatenation-based ML species tree, quartet-based ASTRAL species tree, and true species tree on simulated data sets with different levels of incomplete lineage sorting (ILS).

<table>
<thead>
<tr>
<th>Species tree height</th>
<th>Level of ILS</th>
<th>Replicate</th>
<th>All genes</th>
<th>Consistent genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>IQ-TREE vs. true species tree</td>
<td>ASTRAL vs. true species tree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1M High</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1M High</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1M High</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1M High</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1M High</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5M Medium</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5M Medium</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5M Medium</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5M Medium</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5M Medium</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10M Low</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10M Low</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10M Low</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10M Low</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10M Low</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that NA means concatenation-based ML and quartet-based ASTRAL trees both recovered the true species tree when analyzing full data sets so that our approach is not applicable to dissect incongruence between two approaches. Values in bold denote different phylogenies between concatenation-based ML tree, quartet-based ASTRAL tree, and true species tree.

Table 2. Summary of numbers of conflicting branches between concatenation-based ML species tree, quartet-based ASTRAL species tree, and true species tree on simulated data sets with different levels of incomplete lineage sorting (ILS).
branches receive support from this gene, while others do not. Hence, quantifying phylogenetic signal (e.g., by measuring internode certainty or bootstrap support) on individual internal branches will be informative. If inconsistent genes exhibit substantially lower levels of phylogenetic signal for the conflicting internal branch than consistent genes, removing these inconsistent genes is encouraged. iii) Estimation of tree discordance: considering that ILS and GTEE, as well as other factors, can jointly act to contribute to the observed gene tree discordance, we can quantify gene tree discordance by calculating the topological distance between estimated gene trees and the concatenation-based species phylogeny or the quartet-based species phylogeny. If inconsistent genes exhibit substantially higher levels of gene tree discordance for the conflicting internal branch than consistent genes, removing these inconsistent genes is encouraged. In contrast, we suggest caution for the removal of inconsistent genes when the differences between inconsistent and consistent genes are not significant or when the numbers of retained (i.e., consistent) genes is too small.

## Conclusion

The results presented here show that approximately one-third of genes exhibited inconsistent support between concatenation-based IQ-TREE analysis (T1) and quartet-based ASTRAL analysis (T2) in three representative phylogenomic studies. Compared to consistent genes, inconsistent genes often had similar characteristics in typical sequence- and tree-based metrics, but were more likely to recover neither T1 nor T2 and had higher levels of gene tree discordance, likely due to suffering from higher levels of ILS and/or GTEE. Though a number of ILS-aware algorithms (e.g., quartet-based ASTRAL) putatively account for the bias in the concatenation-based analysis (e.g., Drummond and Rambaut 2007; Liu et al. 2010; Misof et al. 2014), they are not free from gene tree estimation error (e.g., Springer and Gaty 2016; Blom et al. 2017; Mirarab 2019). Given that there is a considerable number of genes that exhibited inconsistent support between concatenation-based IQ-TREE and quartet-based ASTRAL approaches and the weaknesses of two approaches, the dilemma is whether to use an approach that reduces stochastic error caused by using short gene alignments (concatenation-based IQ-TREE) or one that takes into account ILS (quartet-based ASTRAL); answering whether concatenation-based or the coalescent-based approach is more appropriate for phylogenomic inference in general remains challenging because both gene tree estimation error and ILS vary across data sets.

The practical strategy presented here will be useful for dissecting incongruence stemming from the use of two major phylogenomic approaches and for examining the underlying causes of this incongruence. Our results showed that the removal of inconsistent genes from three empirical data sets and simulated data sets eliminated or extensively reduced incongruence between concatenation- and quartet-based approaches. However, it should be noted that the removal of inconsistent genes in simulated data sets with medium or high gene tree discordance levels reduced incongruence but did not always recover the true species phylogeny.

## Supplementary Material

Data available from the Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.9p8cz8wc5.

## Funding

This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University; The National Natural Science Foundation of China [No. 32071665] and the Thousand Youth Talents Program (No. appending) to X.-X.S.; the Howard Hughes Medical Institute through the James H. Gilman Fellowships for Advanced Study program to J.L.S. and A.R. Research on phylogenomics in A.R.’s lab is also supported by the National Science Foundation [DEB-1442113], the Guggenheim Foundation, and the Burnoughs Wellcome Fund.

## Acknowledgments

We thank members of the Rokas lab for helpful discussions and comments. We also thank Emily

## Table 3

Summary of numbers of conflicting branches between concatenation-based ML tree, quartet-based ASTRAL tree, and true species tree on simulated animal data sets with different levels of gene tree estimation error (GTEE)

<table>
<thead>
<tr>
<th>Level of GTEE</th>
<th>All genes</th>
<th>Consistent genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ-TREE vs. true species tree</td>
<td>ASTRAL vs. true species tree</td>
</tr>
<tr>
<td></td>
<td>IQ-TREE vs. true species tree</td>
<td>ASTRAL vs. true species tree</td>
</tr>
<tr>
<td>0.05 x</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>0.05 x</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>0.1 x</td>
<td>Low</td>
<td>0</td>
</tr>
<tr>
<td>1 x</td>
<td>Control</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that NA means concatenation-based ML and quartet-based ASTRAL species trees both recovered the true species tree when analyzing full data sets so that our approach is not applicable to dissect incongruence between two approaches. Values in bold denote different phylogenies between concatenation-based ML species tree, quartet-based ASTRAL species tree, and true species tree.
Roycroft for kindly providing the rodent phylogenomic data matrix. This study stems from an invitation to present our work in a session entitled “New measures of phylogenetic support for the genomic era” in the 2019 Evolution conference. We thank the three symposium organizers (Jeremy Brown, Robert Thomson, and Robert Lanfear) for the invitation and three anonymous reviewers for insightful comments that improved this work.

REFERENCES


