

**Article type:** *Research Article*

**Running Head:** Disentangling incongruence in phylogenomics

**Title:** Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data

Xing-Xing Shen<sup>1, 2\*</sup>, Jacob L. Steenwyk<sup>3</sup>, and Antonis Rokas<sup>3</sup>

<sup>1</sup>State Key Laboratory of Rice Biology and Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Zhejiang University, Hangzhou 310058, China

<sup>2</sup>Institute of Insect Sciences, Zhejiang University, Hangzhou 310058, China

<sup>3</sup>Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

\*Correspondence: xingxingshen@zju.edu.cn (S.X.X.)

**Keywords:** Phylogenomics, Phylogenetics, Conflict, Phylogenetic signal, Tree of Life, Gene Tree

© The Author(s) 2021. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved. For Permissions, please email: journals.permissions@oup.com

## Abstract

Topological conflict or incongruence is widespread in phylogenomic data. Concatenation- and coalescent-based approaches often result in incongruent topologies, but the causes of this conflict can be difficult to characterize. We examined incongruence stemming from conflict between likelihood-based signal (quantified by the difference in gene-wise log likelihood score or  $\Delta\text{GLS}$ ) and quartet-based topological signal (quantified by the difference in gene-wise quartet score or  $\Delta\text{GQS}$ ) for every gene in three phylogenomic studies in animals, fungi, and plants, which were chosen because their concatenation-based 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 higher log likelihood score for T1 versus T2 (i.e.,  $\Delta\text{GLS} > 0$ ) whereas its T1 topology has lower quartet score than its T2 topology (i.e.,  $\Delta\text{GQS} < 0$ ) or vice versa. Comparison of inconsistent and consistent genes using a variety of metrics (e.g., evolutionary rate, gene tree topology, distribution of branch lengths, hidden paralogy, and 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 removal of inconsistent genes from datasets 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 datasets 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.

## Introduction

Advances in genome sequencing technologies have greatly facilitated the construction of phylogenomic data matrices for inference of species phylogenies from genomic and transcriptomic data (Misof et al. 2014; One Thousand Plant Transcriptomes Initiative 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; Irisarri 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 five 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). 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 (Höhna 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 multi-species coalescent model to infer the species phylogeny while accounting for the presence of ILS in individual gene trees (Kingman 1982; Maddison 1997; Rannala and Yang 2003; Liu and Pearl 2007; Edwards 2009). Current scalable (i.e., computationally tractable for genome-wide data) implementations of the coalescent-based approach rely on inference of the individual gene trees, which are then used as input data to estimate a coalescent-based species tree using “summary” species tree programs, such as ASTRAL (Mirarab et al. 2014) and MP-EST (Liu et al. 2010). The major weakness of this approach is that it can be substantially affected by gene tree estimation errors; these can stem from low-quality individual gene alignments, such as those with few informative sites and uneven distributions of missing sequence data (e.g.,

Springer and Gatesy 2016; Blom et al. 2017; Mirarab 2019; Simmons and Kessenich 2020).

Whether the concatenation-based approach or the coalescent-based approach is more appropriate for phylogenomic inference is vigorously debated (e.g., Roch and Warnow 2015; Simmons and Gatesy 2015; 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; Roch and Steel 2015; 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 phylogenomic 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 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 a) quantify the distribution of likelihood-based signal (measured by the difference in gene-wise log likelihood score or  $\Delta\text{GLS}$ ) and quartet-based topological signal (measured by the difference in gene-wise quartet score or  $\Delta\text{GQS}$ ) for every gene in a phylogenomic data matrix, and b) 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,  $\Delta$ GLS and normalized  $\Delta$ GLS, and quartet-based topological signal,  $\Delta$ GQS and normalized  $\Delta$ GQS) to examine characteristics of inconsistent and consistent genes. Lastly, we tested whether 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. 2015) and quartet-based ASTRAL analysis (Mirarab et al. 2014), respectively.

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

**$\Delta\text{GLS}$  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 multi-thread version 1.6.8 (options “-z” and “-wsl”) (Nguyen et al. 2015). We then calculate the difference in gene-wise log-likelihood scores ( $\Delta\text{GLS}$ ) 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):

$$\Delta\text{GLS}_i = \ln L(G_i, T1) - \ln L(G_i, T2) \quad (1)$$

**$\Delta\text{GQS}$  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 “-i” and “-q”) (Mirarab 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 ( $\Delta\text{GQS}$ ) between T1 and T2 for every gene tree  $i$  using the equation:

$$\Delta\text{GQS}_i = \text{QS}(G_i, T1) - \text{QS}(G_i, T2) \quad (2)$$

**Step 3:** According to equations (1) and (2),  $\Delta\text{GLS}$  and  $\Delta\text{GQS}$  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  $\Delta\text{GLS} > 0$  and  $\Delta\text{GQS} > 0$  (i.e., recovering T1) and genes whose  $\Delta\text{GLS} < 0$  and  $\Delta\text{GQS} < 0$  (i.e., recovering T2) are considered consistent (dots in solid blue or solid orange; Fig. 1c). In contrast, genes whose  $\Delta\text{GLS} \geq 0$  and  $\Delta\text{GQS} \leq 0$  or vice versa are considered inconsistent (dots that are half blue and half orange; Fig. 1c).

## Analysis of three empirical phylogenomic data matrices

**Dataset 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 (Steenwyk et al. 2019), and plants (Boachon et al. 2018). Specifically, the animal data matrix was comprised of 1,245 exons with 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 1,034 protein-coding genes with average length of 506 amino acid sites and gene occupancy of 100% from 25 bipolar budding yeasts and 4 outgroups (Steenwyk et al. 2019). The plant data matrix was comprised of 520 protein-coding genes with average length of 1,274 nucleotide sites and gene occupancy of 83% from 48 Lamiaceae 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 studies (Fig.2, Supplementary Figs. S1 and S2).

***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:

- a) 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 (%), evolutionary 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);
- b) 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  $\Delta$ GLS) for T1 versus T2, normalized absolute  $\Delta$ GLS by gene alignment length, absolute difference in

gene-wise quartet score (absolute  $\Delta$ GQS) for T1 versus T2, and normalized absolute  $\Delta$ GQS 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);

c) 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 dataset; *Saccharomyces cerevisiae* for the bipolar budding yeast dataset; and *Rattus norvegicus* for the Lamiaceae dataset) 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 datasets (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). For each replicate, we simulated 1,000 gene trees under the multi-species coalescent model, which were then used to simulate 1,000 gap-free nucleotide gene alignments with randomly varying lengths from 300 to 1,500 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 dataset). Detailed commands and parameters are given in Supplementary Text and Supplementary Table S4.

For different GTEE levels (low:  $X = 0.1$ , medium:  $X = 0.07$ , and high:  $X = 0.05$ ), we simulated three datasets, 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 1,000 gap-free nucleotide gene alignments with varying length (randomized to be between 300 and 1,500 base pairs) using Seq-Gen.v1.3.2 under the GTR +G4 model (sequence transition rates, shape for the gamma rate heterogeneity, and equal state frequency are estimated on the animal dataset).

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 (Mirarab 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 1,000 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.

## 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). 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), 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 two approaches are given below:

**Animal dataset** (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 dataset** (Steenwyk et al. 2019). We observed three conflicting branches between concatenation-based IQ-TREE and quartet-based ASTRAL phylogenies (Supplementary Fig. S1a). 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 (CBS314, 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. clermontiae* and *H. uvarum* (LPP=1.00). The other two conflicting branches concerned the relative placement of four different strains (CBS314, AWRI3580, 34-9 and DSM2768) of *H. uvarum* and exhibited moderate support values (BS=87; LPP=0.69) (Supplementary Fig. S1a).

**Plant dataset** (Boachon et al. 2018). We observed four conflicting branches between concatenation-based IQ-TREE and quartet-based ASTRAL phylogenies (Supplementary Fig. S2a). 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 Peronematoideae, 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 genus *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).

### **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 [that is, difference in gene-wise log likelihood score ( $\Delta$ GLS) for T1 versus T2] and quartet-based ASTRAL phylogenetic signal [that is, difference in gene-wise quartet score ( $\Delta$ GQS) for T1 versus 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). 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).

Examination of the distribution of gene-wise phylogenetic signal between concatenation-based IQ-TREE and quartet-based ASTRAL approaches showed that 794 / 1,245 (~64%) genes in animals, 683 / 1,034 (~66%) genes in fungi, and 363 / 520 (~70%) genes in plants were consistent, that is their  $\Delta\text{GLS}$  and  $\Delta\text{GQS}$  values had the same signs, while 451 / 1,245 genes (~36%) in animals, 351 / 1,034 genes (~34%) in fungi, and 157 / 520 genes (~30%) in plants were inconsistent, that is their  $\Delta\text{GLS}$  and  $\Delta\text{GQS}$  values had opposite signs (Fig. 2b, Supplementary Figs. S1b and S2b). Interestingly, we found that proportions of inconsistent genes with  $\Delta\text{GLS} > 0$  (recovering T1) and  $\Delta\text{GQS} < 0$  (recovering T2) are generally lower than those of inconsistent genes with  $\Delta\text{GLS} < 0$  (recovering T2) and  $\Delta\text{GQS} > 0$  (recovering T1) in three empirical data matrices (animals: 196 / 451 genes [43%] versus 255 / 451 genes [57%]; fungi: 142 / 351 genes [40%] versus 209 / 351 genes [60%]; plants: 55 / 157 genes [35%] versus 102 / 157 genes [65%]). However, we found that proportions of inconsistent genes with  $\Delta\text{GLS} > 0$  and  $\Delta\text{GQS} < 0$  are not always lower than those of inconsistent genes with  $\Delta\text{GLS} < 0$  and  $\Delta\text{GQS} > 0$  in different types of simulated data matrices (on average, simulated data matrices with different ILS levels: 642 / 1,234 genes [52%] versus 592 / 1,234 genes [48%]; simulated data matrices with different GTEE levels: 111 / 1,162 genes [10%] versus 1051 / 1,162 genes [90%]). Last, compared to inconsistent genes, consistent genes had significantly higher (Wilcoxon rank-sum test) values of absolute  $\Delta\text{GLS}$ , normalized absolute  $\Delta\text{GLS}$ , absolute  $\Delta\text{GQS}$ , and normalized absolute  $\Delta\text{GQS}$  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 0.10 for the branch *Coccymys* in animals, 0.17 for the branch *H. nectarophila* in fungi, 0.47 for the branch *H. uvarum* 34-9 in fungi, 0.47 for the branch *H. uvarum* AWRI3580 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 pants, 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. uvarum* 34-9 in fungi, 0.47 for the branch *H. uvarum* AWRI3580 in fungi, 0.26 for the branch Premnoideae in plants, 0.28 for the branch Peronematoideae in plants, 0.35 for the branch *Lycopus* in pants, 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 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 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). 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, 0.00016 substitutions per site across inconsistent genes versus 0.00071 substitutions per site across consistent genes), 1.08-fold shorter length for the branch *H. nectarophila* in fungi (0.0071 versus 0.0077), 1.28-fold shorter length for the branch *H. uvarum* 34-9 (0.00025 versus 0.00032), 1.05-fold longer length for the branch *H. uvarum* AWRI3580 (0.00061 versus 0.00058), 1.15-fold shorter length for the branch Premnoideae in plants (0.0019 versus 0.0022), 1.03-fold shorter length for the branch Peronematoideae in plants (0.00038 versus 0.00039), 1.02-fold shorter length for the branch *Lycopu* in plants (0.0062 versus 0.0063), and 1.18-fold shorter length for the branch *Nepeta* + *Agastache* in plants (0.0022 versus 0.0026), compared with consistent genes. Our results suggest that inconsistent genes tend to recover neither T1 nor T2, and are more likely to have short internal branches when we constrain them to recover T1 or T2.

## **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, 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., Zhou 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.

To explore factors that likely contribute to different gene-wise support values between 959 inconsistent genes and 1,840 consistent genes from the three empirical phylogenomic datasets, we extensively examined ten sequence- (e.g., GC content, evolutionary rate) and tree-based (e.g., distribution of branch lengths or Treeness, 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, Treeness, and DVMC between inconsistent and consistent genes (on average, GC content: 32.5% versus 31.7%; evolutionary rate: 85% versus 84.5%; Treeness: 0.360 versus 0.365; DVMC: 0.13 versus 0.14) (Fig. 3b and Supplementary Table S2). 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 to gene informativeness) compared to consistent genes (Fig. 3b).

We also did not identify significant differences in level of hidden paralogy between the 959 inconsistent genes and the 1,840 consistent genes (on average, ortholog certainty: 0.90 versus 0.89) (Fig. 5a; Supplementary Table S3). 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 dataset: 0.55 versus 0.51;



normalized RF distance in the fungal dataset: 0.28 versus 0.25), but not in the plant data matrix (on average, normalized RF distance in the plant dataset: 0.373 versus 0.369) (Fig. 5b). 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 very strong phylogenetic signal can have a major influence on the results of concatenation-based phylogenetic inference (e.g., Brown and Thomson 2017; 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  $\Delta$ GLS and  $\Delta$ GQS values in the three phylogenomic data matrices suggests that none of the data matrices contains any obvious outlier genes (Supplementary Table S2).

### **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 datasets, we used simulations to examine the effect of different levels of gene tree discordance due to ILS and 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 1,000 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 length.

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

- a) 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 / 1,000 (39.8%), 412 / 1,000 (41.2%), and 423 / 1,000 (42.3%) for species trees with low, medium, and high levels of ILS, respectively; the numbers of inconsistent genes were 349 / 1,000 (34.9%), 384 / 1,000 (38.4%), and 429 / 1,000 (42.9 %) for low, medium, and high levels of GTEE, respectively; Fig. 6a and d),
- b) 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), and

c) the total number of conflicting branches between the species phylogenies inferred by concatenation-based IQ-TREE and quartet-based ASTRAL approaches increased with increased 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 ten sequence- and tree-based metrics in the simulated data matrices show that most of the ten 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 dataset) exhibited significant differences between inconsistent and consistent genes (Supplementary Fig. S5 and Supplementary Table S5). 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 Treeness) (Supplementary Fig. S6 and Supplementary Table S6).

## 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 dataset, all four incongruent branches in the plant dataset, as well as reducing the number of incongruent branches from three to one in the fungal dataset (Fig. 7, Supplementary Figs. S7 and S8). Among the seven eliminated incongruent branches, five recovered T1 (animal dataset: *Coccymys*; fungal dataset: *H. nectarophila* and *H. uvarum* AWRI3580; plant dataset: Premnoideae and Peronematoideae) and two recovered T2 (plant dataset: *Lycopus* and *Nepeta* + *Agastache*). For example, 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 *Coccymys* as the sister group to the SHL clade (T1). Although 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 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 removal of inconsistent genes eliminated incongruence and recovered the true species trees. However, in simulated datasets with medium and high levels of ILS and high levels of GTEE, 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 (Molloy 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 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: 1) 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 testing whether inference based on the remaining genes results in a congruent species tree inferred by both approaches. 2) Estimation of GTEE: lack of phylogenetic signal is one of most typical factors that influence the accuracy of gene tree estimation. Previous studies pointed that a gene's phylogenetic signal can be unevenly or disproportionately distributed across the branches of its gene tree (Townsend and Leuenberger 2011; Chen et al. 2015; Shen et al. 2017), which means that in concatenation- or quartet-based analyses, some internal 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. 3) Estimation of gene 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 concatenation-based analysis (e.g., Drummond and Rambaut 2007; Liu et al. 2010; Mirarab et al. 2014), they are not free from gene tree estimation error (e.g., Springer and Gatesy 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 datasets.

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 removal of inconsistent genes from three empirical datasets and simulated datasets eliminated or extensively reduced incongruence between concatenation- and quartet-based approaches. However, it should be noted that removal of inconsistent genes in simulated datasets 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. X.-X.S. was supported by the National Natural Science Foundation of China (No. 32071665) and the Thousand Youth Talents Program (No. appending). J.L.S. and A.R. were supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program.

Research on phylogenomics in A.R.'s lab is also supported by the National Science



Foundation (DEB-1442113), the Guggenheim Foundation, and the Burroughs Wellcome Fund.

## **Acknowledgments**

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

- Aberer A.J., Kobert K., Stamatakis A. 2014. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. *Mol. Biol. Evol.* 31:2553–2556.
- Anisimova M., Liberles D.A., Philippe H., Provan J., Pupko T., von Haeseler A. 2013. State-of the art methodologies dictate new standards for phylogenetic analysis. *BMC Evol. Biol.* 13:161.
- Blom M.P.K., Bragg J.G., Potter S., Moritz C. 2017. Accounting for uncertainty in gene tree estimation: Summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Syst. Biol.* 66:352–366.
- Boachon B., Buell C.R., Crisovan E., Dudareva N., Garcia N., Godden G., Henry L., Kamileen M.O., Kates H.R., Kilgore M.B., Lichman B.R., Mavrodiev E. V., Newton L., Rodriguez-Lopez C., O'Connor S.E., Soltis D., Soltis P., Vaillancourt B., Wiegert-Rininger K., Zhao D. 2018. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Mol. Plant.* 11:1084–1096.
- Boussau B., Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. *Syst. Biol.* 55:756–768.
- Bravo G.A., Antonelli A., Bacon C.D., Bartoszek K., Blom M.P.K., Huynh S., Jones G., Knowles L.L., Lamichhaney S., Marcussen T., Morlon H., Nakhleh L.K., Oxelman B., Pfeil B., Schliep A., Wahlberg N., Werneck F.P., Wiedenhoeft J., Willows-Munro S., Edwards S. V. 2019. Embracing heterogeneity: coalescing the Tree of Life and the

- future of phylogenomics. *PeerJ*. 7:e6399.
- Brown J.M., Thomson R.C. 2017. Bayes Factors Unmask Highly Variable Information Content, Bias, and Extreme Influence in Phylogenomic Analyses. *Syst. Biol.* 66:517–530.
- Brown J.M., Thomson R.C. 2018. Evaluating Model Performance in Evolutionary Biology. *Annu. Rev. Ecol. Evol. Syst.* 49:95–114.
- Chen M.-Y., Liang D., Zhang P. 2015. Selecting Question-Specific Genes to Reduce Incongruence in Phylogenomics: A Case Study of Jawed Vertebrate Backbone Phylogeny. *Syst. Biol.* 64:1104–1120.
- Davidson R., Vachaspati P., Mirarab S., Warnow T. 2015. Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. *BMC Genomics*. 16:S1.
- Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24:332–340.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Edwards S. V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Mol. Phylogenet. Evol.* 94:447–62.
- Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging?

- Evolution (N. Y). 63:1–19.
- 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.* 27:2733–2746.
- Fletcher W., Yang Z. 2009. INDELible: A Flexible Simulator of Biological Sequence Evolution. *Mol. Biol. Evol.* 26:1879–1888.
- Di Franco A., Poujol R., Baurain D., Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol. Biol.* 19:21.
- Gatesy J., Baker R.H. 2005. Hidden Likelihood Support in Genomic Data: Can Forty-Five Wrongs Make a Right? *Syst. Biol.* 54:483–492.
- Gatesy J., Meredith R.W., Janecka J.E., Simmons M.P., Murphy W.J., Springer M.S. 2017. Resolution of a concatenation/coalescence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. *Cladistics.* 33:295–332.
- Gatesy J., Sloan D.B., Warren J.M., Baker R.H., Simmons M.P., Springer M.S. 2019. Partitioned coalescence support reveals biases in species-tree methods and detects gene trees that determine phylogenomic conflicts. *Mol. Phylogenet. Evol.* 139:106539.
- Gatesy J., Springer M.S. 2013. Concatenation versus coalescence versus “concatalescence”. *Proc. Natl. Acad. Sci. U. S. A.* 110:E1179.
- Gatesy J., Springer M.S. 2014. Phylogenetic Analysis at Deep Timescales: Unreliable Gene Trees, Bypassed Hidden Support, and the Coalescence/Concatalence Conundrum.

- Mol. Phylogenet. Evol. 80:231–266.
- Guang A., Zapata F., Howison M., Lawrence C.E., Dunn C.W. 2016. An Integrated Perspective on Phylogenetic Workflows. *Trends Ecol. Evol.* 31:116–126.
- Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
- Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Syst. Biol.* 65:726–736.
- Irisarri I., Baurain D., Brinkmann H., Delsuc F., Sire J.-Y., Kupfer A., Petersen J., Jarek M., Meyer A., Vences M., Philippe H. 2017. Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nat. Ecol. Evol.* 1:1370–1378.
- Jiang X., Edwards S. V, Liu L. 2020. The Multispecies Coalescent Model Outperforms Concatenation Across Diverse Phylogenomic Data Sets. *Syst. Biol.* 69:795–812.
- Kimball R.T., Wang N., Heimer-McGinn V., Ferguson C., Braun E.L. 2013. Identifying localized biases in large datasets: a case study using the avian tree of life. *Mol. Phylogenet. Evol.* 69:1021–1032.
- Kingman J.F.C. 1982. The coalescent. *Stoch. Process. their Appl.* 13:235–248.
- Kozlov A.M., Darriba D., Flouri T., Morel B., Stamatakis A. 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 35:4453–4455.

- Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56:17–24.
- Lapierre P., Lasek-Nesselquist E., Gogarten J.P. 2014. The impact of HGT on phylogenomic reconstruction methods. *Brief. Bioinform.* 15:79–90.
- Lartillot N., Lepage T., Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics.* 25:2286–2288.
- Lartillot N., Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21:1095–1109.
- Laurin-Lemay S., Brinkmann H., Philippe H. 2012. Origin of land plants revisited in the light of sequence contamination and missing data. *Curr. Biol.* 22:R593–R594.
- Leaché A.D., Rannala B. 2011. The Accuracy of Species Tree Estimation under Simulation: A Comparison of Methods. *Syst. Biol.* 60:126–137.
- Lee M.S.Y., Hugall A.F. 2003. Partitioned likelihood support and the evaluation of data set conflict. *Syst. Biol.* 52:15–22.
- Liu L., Pearl D.K. 2007. Species Trees from Gene Trees: Reconstructing Bayesian Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions. *Syst. Biol.* 56:504–514.
- Liu L., Xi Z., Davis C.C. 2015a. Coalescent Methods Are Robust to the Simultaneous Effects of Long Branches and Incomplete Lineage Sorting. *Mol. Biol. Evol.* 32:791–805.
- Liu L., Xi Z., Wu S., Davis C.C., Edwards S. V. 2015b. Estimating phylogenetic trees from genome-scale data. *Ann. N. Y. Acad. Sci.* 1360:36–53.

- Liu L., Yu L., Edwards S. V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* 10:302.
- Liu L., Zhang J., Rheindt F.E., Lei F., Qu Y., Wang Y., Zhang Y., Sullivan C., Nie W., Wang J., Yang F., Chen J., Edwards S. V., Meng J., Wu S. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proc. Natl. Acad. Sci. U. S. A.* 114:E7282–E7290.
- Maddison W.P. 1997. Gene Trees in Species Trees. *Syst. Biol.* 46:523.
- Mallo D., De Oliveira Martins L., Posada D. 2016. SimPhy : Phylogenomic Simulation of Gene, Locus, and Species Trees. *Syst. Biol.* 65:334–344.
- Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* 30:1188–1195.
- Mirarab S. 2019. Species Tree Estimation Using ASTRAL: Practical Considerations. :1–21.
- Mirarab S., Bayzid M.S., Warnow T. 2016. Evaluating Summary Methods for Multilocus Species Tree Estimation in the Presence of Incomplete Lineage Sorting. *Syst. Biol.* 65:366–380.
- Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics.* 30:i541–i548.
- Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J., Aberer A.J., Aspöck U., Aspöck H., Bartel D., Blanke A., Berger S., Böhm A.,

- Buckley T.R., Calcott B., Chen J., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S., Huang Y., Jermini L.S., Kawahara A.Y., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y., Li Z., Li J., Lu H., Machida R., Mashimo Y., Kapli P., McKenna D.D., Meng G., Nakagaki Y., Navarrete-Heredia J.L., Ott M., Ou Y., Pass G., Podsiadlowski L., Pohl H., von Reumont B.M., Schütte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W., Su X., Szucsich N.U., Tan M., Tan X., Tang M., Tang J., Timelthaler G., Tomizuka S., Trautwein M., Tong X., Uchifune T., Walz M.G., Wiegmann B.M., Wilbrandt J., Wipfler B., Wong T.K., Wu Q., Wu G., Xie Y., Yang S., Yang Q., Yeates D.K., Yoshizawa K., Zhang Q., Zhang R., Zhang W., Zhang Y., Zhao J., Zhou C., Zhou L., Ziesmann T., Zou S., Li Y., Xu X., Zhang Y., Yang H., Wang J., Wang J., Kjer K.M., Zhou X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* (80-. ). 346:763–767.
- Molloy E.K., Warnow T. 2018. To Include or Not to Include: The Impact of Gene Filtering on Species Tree Estimation Methods. *Syst. Biol.* 67:285–303.
- Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol. Evol.* 28:719–728.
- Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature.* 574:679–685.



- Philippe H., Vienne D.M. de, Ranwez V., Roure B., Baurain D., Delsuc F. 2017. Pitfalls in supermatrix phylogenomics. *Eur. J. Taxon.*:1–25.
- Phillips M.J., Penny D. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* 28:171–185.
- Prasanna A.N., Gerber D., Kijpornyongpan T., Aime M.C., Doyle V.P., Nagy L.G. 2020. Model Choice, Missing Data, and Taxon Sampling Impact Phylogenomic Inference of Deep Basidiomycota Relationships. *Syst. Biol.* 69:17–37.
- Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature.* 526:569–573.
- Rannala B., Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics.* 164:1645–1656.
- Rasmussen M.D., Kellis M. 2012. Unified modeling of gene duplication, loss, and coalescence using a locus tree. *Genome Res.* 22:755–765.
- Roch S., Steel M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor. Popul. Biol.* 100:56–62.
- Roch S., Warnow T. 2015. On the Robustness to Gene Tree Estimation Error (or lack thereof) of Coalescent-Based Species Tree Methods. *Syst. Biol.* 64:663–676.
- 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., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu

- L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–542.
- Roycroft E.J., Moussalli A., Rowe K.C. 2020. Phylogenomics Uncovers Confidence and Conflict in the Rapid Radiation of Australo-Papuan Rodents. *Syst. Biol.* 69:431–444.
- Salichos L., Rokas A. 2011. Evaluating ortholog prediction algorithms in a yeast model clade. *PLoS One.* 6:e18755.
- Salichos L., Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature.* 497:327–331.
- Salichos L., Stamatakis A., Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol. Biol. Evol.* 31:1261–1271.
- Sayyari E., Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Mol. Biol. Evol.* 33:1654–1668.
- Scornavacca C., Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Syst. Biol.* 66:112–120.
- Shen X.-X., Hittinger C.T., Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* 1:0126.
- Shen X.-X., Li Y., Hittinger C.T., Chen X., Rokas A. 2020a. An investigation of irreproducibility in maximum likelihood phylogenetic inference. *Nat. Commun.* 11:6096.
- Shen X.-X., Opulente D.A., Kominek J., Zhou X., Steenwyk J.L., Buh K. V., Haase M.A.B., Wisecaver J.H., Wang M., Doering D.T., Boudouris J.T., Schneider R.M., Langdon

- Q.K., Ohkuma M., Endoh R., Takashima M., Manabe R., Čadež N., Libkind D., Rosa C.A., DeVirgilio J., Hulfachor A.B., Groenewald M., Kurtzman C.P., Hittinger C.T., Rokas A. 2018. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*. 175:1533-1545.e20.
- Shen X.-X., Steenwyk J.L., LaBella A.L., Opulente D.A., Zhou X., Kominek J., Li Y., Groenewald M., Hittinger C.T., Rokas A. 2020b. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci. Adv.* 6:eabd0079.
- Shen X.-X., Zhou X., Kominek J., Kurtzman C.P., Hittinger C.T., Rokas A. 2016. Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3 Genes|Genomes|Genetics*. 6:3927–3939.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51:492–508.
- Simmons M.P., Gatesy J. 2015. Coalescence vs. concatenation: sophisticated analyses vs. first principles applied to rooting the angiosperms. *Mol. Phylogenet. Evol.* 91:98–122.
- Simmons M.P., Kessenich J. 2020. Divergence and support among slightly suboptimal likelihood gene trees. *Cladistics*. 36:322–340.
- Smith S.A., Walker-Hale N., Walker J.F., Brown J.W. 2020. Phylogenetic Conflicts, Combinability, and Deep Phylogenomics in Plants. *Syst. Biol.* 69:579–592.
- Springer M.S., Gatesy J. 2016. The gene tree delusion. *Mol. Phylogenet. Evol.* 94:1–33.
- Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of

large phylogenies. *Bioinformatics*. 30:1312–1313.

Steenwyk J.L., Opulente D.A., Kominek J., Shen X.-X., Zhou X., Labella A.L., Bradley N.P.,

Eichman B.F., Čadež N., Libkind D., DeVirgilio J., Hulfachor A.B., Kurtzman C.P.,

Hittinger C.T., Rokas A. 2019. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLOS Biol.* 17:e3000255.

Tan G., Muffato M., Ledergerber C., Herrero J., Goldman N., Gil M., Dessimoz C. 2015.

Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Syst. Biol.* 64:778–791.

Townsend J.P., Leuenberger C. 2011. Taxon Sampling and the Optimal Rates of Evolution for Phylogenetic Inference. *Syst. Biol.* 60:358–365.

Walker J.F., Brown J.W., Smith S.A. 2018. Analyzing Contentious Relationships and Outlier Genes in Phylogenomics. *Syst. Biol.* 67:916–924.

Walker J.F., Shen X., Rokas A., Smith S.A., Moyroud E. 2020. Disentangling biological and analytical factors that give rise to outlier genes in phylogenomic matrices. *bioRxiv*. doi:10.1101/2020.04.20.049999.

Wickett N.J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N., Ayyampalayam S., Barker M.S., Burleigh J.G., Gitzendanner M. a, Ruhfel B.R., Wafula E., Der J.P., Graham S.W., Mathews S., Melkonian M., Soltis D.E., Soltis P.S., Miles N.W., Rothfels C.J., Pokorny L., Shaw A.J., DeGironimo L., Stevenson D.W., Surek B., Villarreal J.C., Roure B., Philippe H., DePamphilis C.W., Chen T., Deyholos M.K., Baucom R.S., Kutchan T.M., Augustin M.M., Wang J., Zhang Y., Tian Z., Yan Z., Wu X., Sun X.,

- Wong G.K.-S., Leebens-Mack J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci.* 111:E4859–E4868.
- Wu M., Kostyun J.L., Hahn M.W., Moyle L.C. 2018. Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol. Ecol.* 27:3301–3316.
- Yang Z., Zhu T. 2018. Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. *Proc. Natl. Acad. Sci.* 115:1854–1859.
- Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 19:153.
- Zhou X., Shen X.-X., Hittinger C.T., Rokas A. 2018. Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *Mol. Biol. Evol.* 35:486–503.

## Figure Legends

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 ( $\Delta\text{GLS}$ ) 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 ( $\Delta\text{GQS}$ ) 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  $\Delta\text{GLS}$  and  $\Delta\text{GQS}$ . Genes whose  $\Delta\text{GLS} > 0$  and  $\Delta\text{GQS} > 0$  (i.e., recovering T1 labeled in blue) and genes whose  $\Delta\text{GLS} < 0$  and  $\Delta\text{GQS} < 0$  (i.e., recovering T2 labeled in orange) are considered consistent, whereas genes whose  $\Delta\text{GLS} \geq 0$  and  $\Delta\text{GQS} \leq 0$  or vice versa (labeled in half-blue and half-orange) are considered inconsistent.

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 1,245 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 *Anisomys*, 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 1,245 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).

Figure 3. Characteristics of 14 metrics between inconsistent and consistent genes in three phylogenomic datasets. 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 gene, **a)** we calculated absolute difference in gene-wise log likelihood score (absolute  $\Delta$ GLS) for T1 versus T2, normalized absolute  $\Delta$ GLS by gene alignment length, absolute difference in gene-wise quartet score (absolute  $\Delta$ GQS) for T1 versus T2, and normalized absolute  $\Delta$ GQS

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 single fully resolved gene tree). **b)** We also calculated number of taxa in gene alignment, alignment length, percentage of parsimony-informative sites in gene alignment, GC content (%), evolutionary rate determined by pairwise similarity in gene alignment, relative composition frequency variability (RCFV) 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), degree of violation of a molecular clock (DVMC) in a single gene tree, and signal-to-noise ratio (ratio of Treeness to RCFV). Each bar denotes the average value across genes; error bar denotes the standard deviation. The Wilcoxon rank-sum test was used to test if inconsistent and consistent genes exhibited significantly different patterns. The list of gene-wise characteristics is given in Supplementary Table S2.

Figure 4. Characteristics of individual gene trees in three phylogenomic datasets.

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 single gene trees to recover T2 (see Supplementary Fig. S4). 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 optimized under the same constraint. By examining the unconstrained gene trees and the constrained gene trees between inconsistent and consistent genes in three empirical data matrices, we found that inconsistent genes identified by our approach tended to recover neither T1 nor T2 and were more likely to have shorter internal branches 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  $\leq$  0.001; \*\*: P-value  $\leq$  0.01; \*: P-value  $\leq$  0.05; NS: not significant)

Figure 5. Characteristics of hidden paralogy and gene tree discordance between inconsistent and consistent genes in three phylogenomic datasets. 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 gene, we examined: **a**) hidden paralogy determined by considering the frequency of the most prevalent ortholog against that of the second most prevalent ortholog in gene alignment (ortholog certainty), and **b**) gene tree discordance quantified by the amount of discordance between estimated gene tree and concatenation-based IQ-TREE species tree (note that we observed a similar pattern of gene tree discordance between estimated gene trees and the quartet-based ASTRAL species tree). Each bar denotes the average value across genes; error bar denotes the standard deviation. The Wilcoxon rank-sum test was used to test if inconsistent and consistent genes exhibited significantly different patterns (\*\*\*:  $P\text{-value} \leq 0.001$ ; NS: not significant). The list of gene-wise characteristics of two metrics is given in Supplementary Table S3.

Figure 6. Dissecting incongruence in simulated datasets with varying levels of gene tree discordance due to incomplete lineage sorting (ILS) and gene tree estimation error (GTEE). For ILS (left panel), we simulated 15 data sets with low, medium, and high levels of ILS. For GTEE (right panel), we simulated 3 data sets with low, medium, and high levels of GTEE (see Methods for details). **a** and **d**) Average percentage of genes that exhibited inconsistent (in red) and consistent (in green) support between concatenation-based IQ-TREE and quartet-based ASTRAL approaches. **b** and **e**) Comparison of mean gene tree discordance (measured by normalized Robinson-Foulds distance between estimated gene tree and true gene tree). **c** and **f**) The effect of removal of inconsistent genes on reducing incongruence between concatenation-based IQ-TREE and quartet-based ASTRAL approaches. In addition to

comparison of the concatenation-based IQ-TREE species tree and the quartet-based ASTRAL species tree, comparisons of the concatenation-based IQ-TREE species tree and the quartet-based ASTRAL species tree against the true species tree are given in the Tables 2 and 3. The Wilcoxon rank-sum test was used to test if the sets of values are significantly different (\*\*\*:  $P\text{-value} \leq 0.001$ ; \*\*:  $P\text{-value} \leq 0.01$ ; \*:  $P\text{-value} \leq 0.05$ ; NS: not significant)

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 dataset. The concatenation-based IQ-TREE species tree (left panel) and the quartet-based ASTRAL species tree (right panel) were inferred using a dataset of the 794 genes that exhibited consistent support between  $\Delta\text{GLS}$  and  $\Delta\text{GQS}$  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).

## Tables

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.

Study	Branch	Concatenation-based tree hypothesis (T1)	Coalescent-based tree hypothesis (T2)	P-value of AU test <sup>^</sup>
Animals	<i>Coccymys</i>	<i>Coccymys</i> as sister to SHL	<i>Coccymys</i> as sister to <i>Mallomys</i> + <i>Mammelomys</i> + <i>Xenuromys</i>	1.4 x 10 <sup>-2</sup> *
Fungi	<i>H. nectarophila</i>	<i>H. nectarophila</i> as sister to <i>H. uvarum</i>	<i>H. nectarophila</i> as sister to <i>H. meyeri</i> + <i>H. clermontiae</i> and <i>H. uvarum</i>	6.35 x 10 <sup>-56</sup> *
	<i>H. uvarum</i> 34-9	<i>H. uvarum</i> 34-9 as sister to <i>H. uvarum</i> DSM2768	<i>H. uvarum</i> 34-9 as sister to all other three strains of <i>H. uvarum</i>	9.2 x 10 <sup>-2</sup>
	<i>H. uvarum</i> AWRI3580	<i>H. uvarum</i> AWRI3580 as sister to <i>H. uvarum</i> CBS314	<i>H. uvarum</i> AWRI3580 as sister to <i>H. uvarum</i> CBS314 + <i>H. uvarum</i> DSM2768	3.1 x 10 <sup>-3</sup> *
Plants	Premnoideae	Premnoideae is paraphyletic	Premnoideae is monophyletic	2.2 x 10 <sup>-1</sup>
	Peronematoideae	Peronematoideae as sister to Lamioideae + Ajugoideae	Peronematoideae as sister to Lamioideae	1.9 x 10 <sup>-4</sup> *
	<i>Lycopus</i>	<i>Lycopus</i> as sister to <i>Prunella</i>	<i>Lycopus</i> as sister to a clade consisting of <i>Prunella</i> + <i>Nepeta</i> + <i>Agastache</i> + <i>Origanum</i> + <i>Mentha</i>	3.2 x 10 <sup>-34</sup> *
	<i>Nepeta</i> + <i>Agastache</i>	<i>Nepeta</i> + <i>Agastache</i> as sister to <i>Origanum</i> + <i>Mentha</i>	<i>Nepeta</i> + <i>Agastache</i> as sister to <i>Prunella</i> and <i>Origanum</i> + <i>Mentha</i>	9.1 x 10 <sup>-3</sup> *

<sup>^</sup>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).

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 datasets with different levels of incomplete lineage sorting (ILS).

Species tree height	Level of ILS	Replicate	All genes			Consistent genes		
			IQ-TREE vs True species tree	ASTRAL vs True species tree	IQ-TREE vs ASTRAL	IQ-TREE vs True species tree	ASTRAL vs True species tree	IQ-TREE vs ASTRAL
1M	High	1	0	<b>1</b>	<b>1</b>	0	0	0
1M	High	2	0	<b>1</b>	<b>1</b>	0	0	0
1M	High	3	<b>2</b>	<b>1</b>	<b>2</b>	0	0	0
1M	High	4	<b>1</b>	<b>1</b>	<b>2</b>	<b>1</b>	0	<b>1</b>
1M	High	5	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>	0
5M	Medium	1	0	<b>2</b>	<b>2</b>	0	0	0
5M	Medium	2	0	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	0
5M	Medium	3	0	<b>1</b>	<b>1</b>	0	0	0
5M	Medium	4	1	0	1	0	0	0
5M	Medium	5	0	0	0	NA	NA	NA
10M	Low	1	<b>1</b>	0	<b>1</b>	0	0	0
10M	Low	2	0	0	0	NA	NA	NA
10M	Low	3	0	<b>1</b>	<b>1</b>	0	0	0
10M	Low	4	<b>2</b>	<b>2</b>	<b>2</b>	0	0	0
10M	Low	5	0	0	0	NA	NA	NA

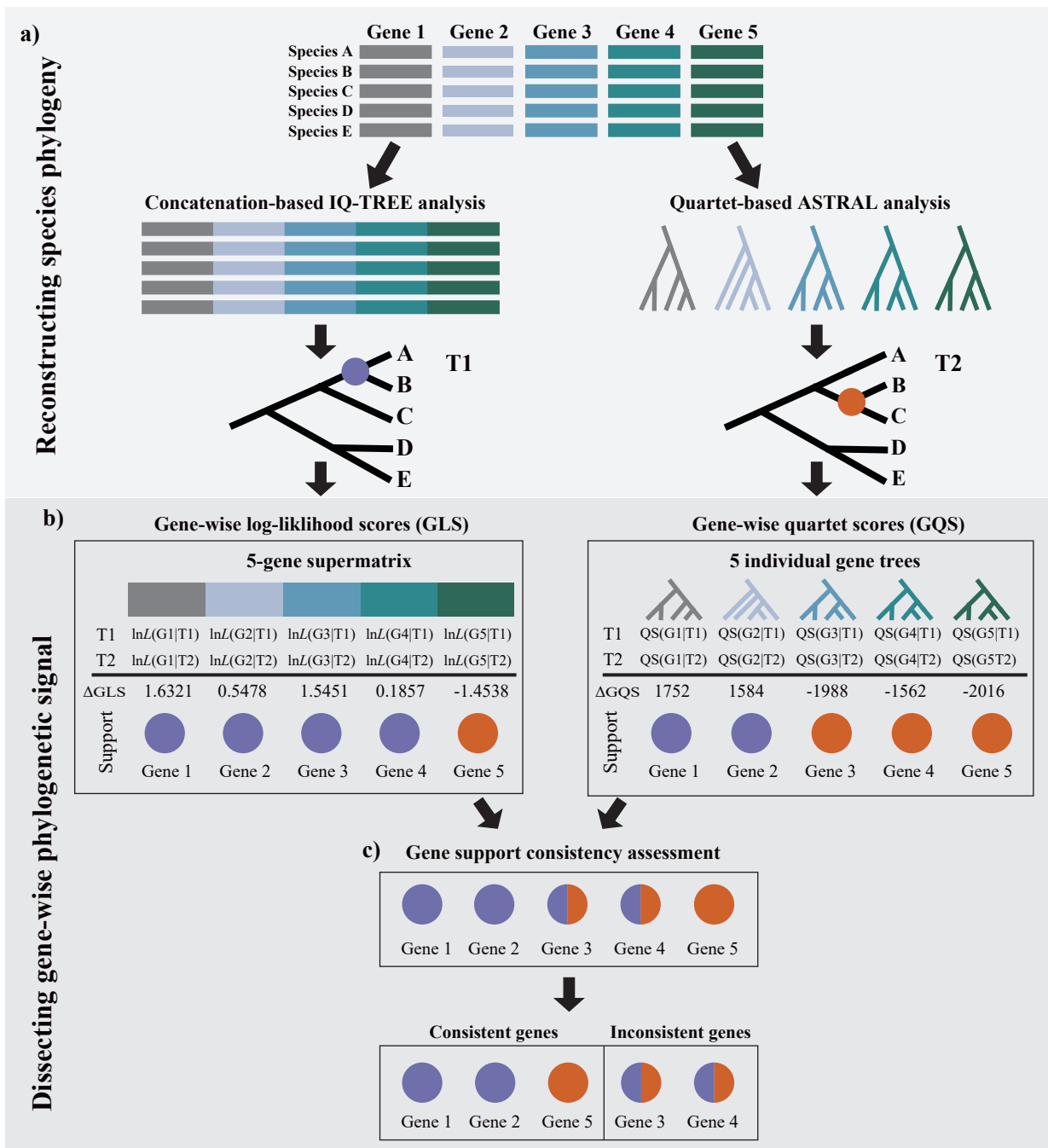
Note that NA means concatenation-based ML and quartet-based ASTRAL trees both recovered the true species tree when analyzing full datasets so that our approach is not applicable to dissect incongruence between two approaches. Values in bold denote different phylogenies between concatenation-based ML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

Table 3. Summary of numbers of conflicting branches between concatenation-based ML tree, quartet-based ASTRAL tree, and true species tree on simulated animal datasets with different levels of gene tree estimation error (GTEE).

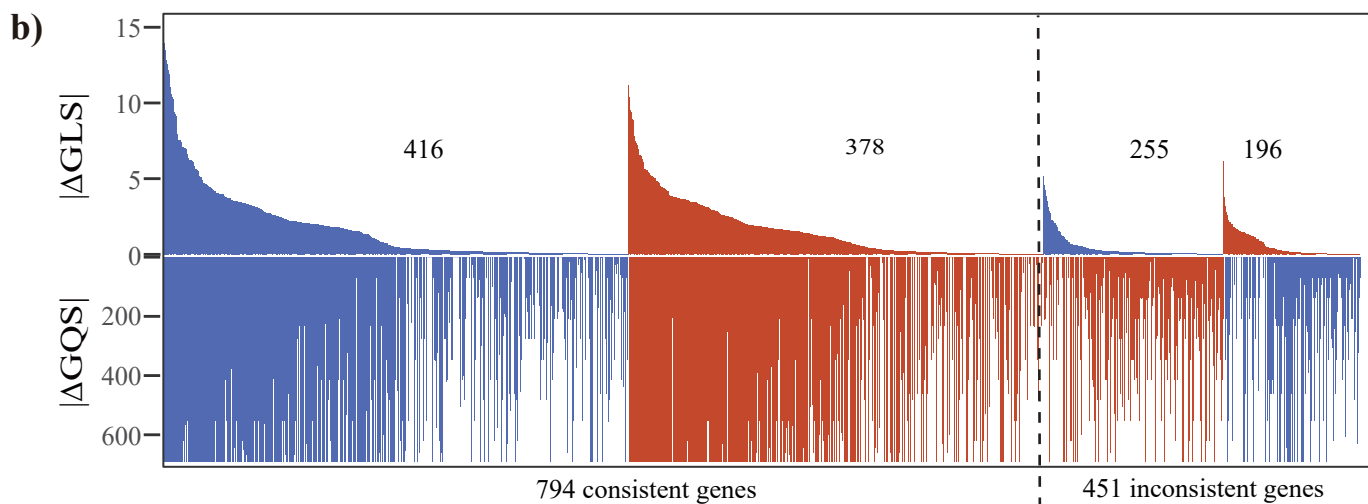
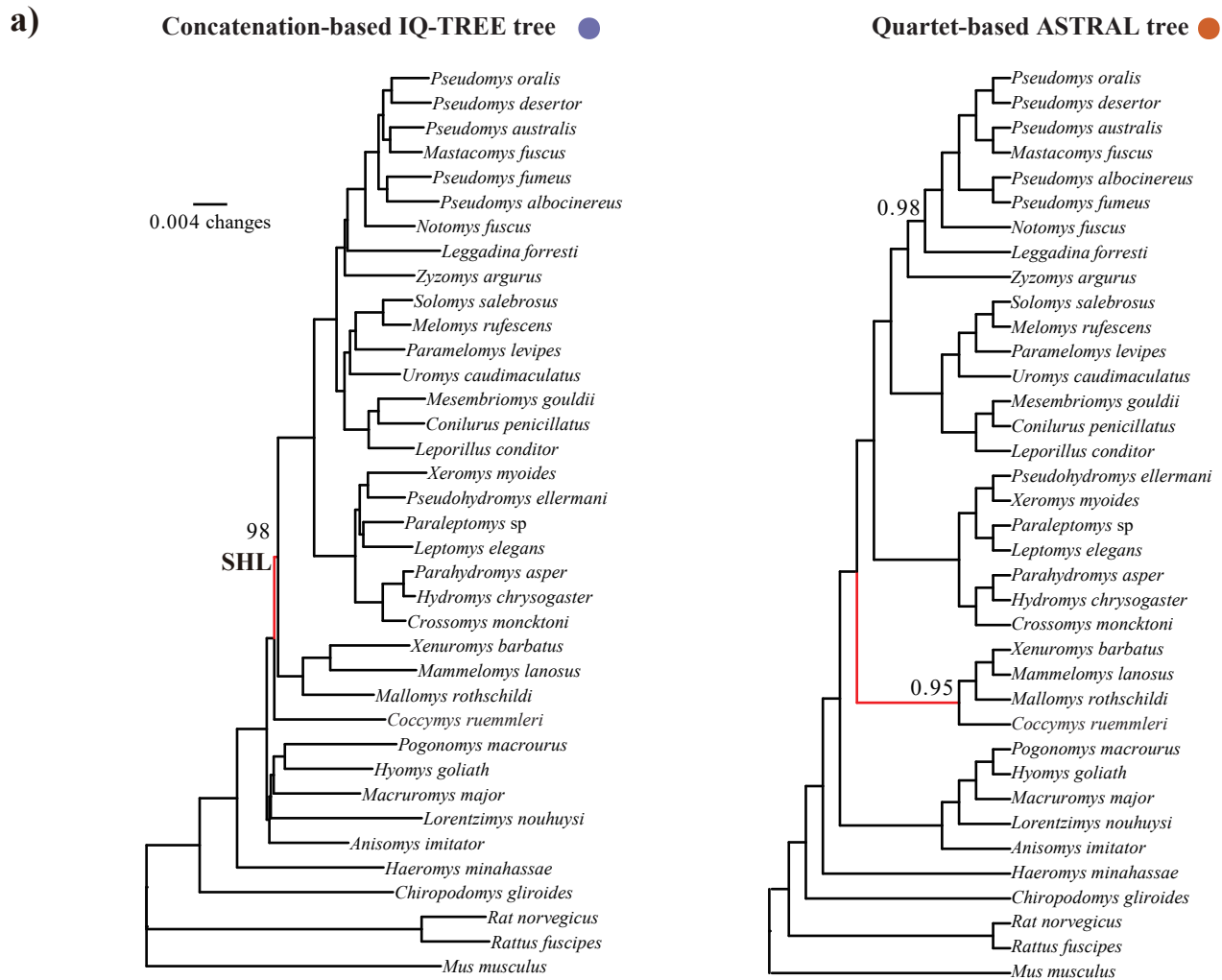
Scale	Level of GTEE	All genes			Consistent genes		
		IQ-TREE vs True species tree	ASTRAL vs True species tree	IQ-TREE vs ASTRAL	IQ-TREE vs True species tree	ASTRAL vs True species tree	IQ-TREE vs ASTRAL
0.05X	High	0	<b>9</b>	<b>9</b>	0	<b>1</b>	<b>1</b>
0.07X	Medium	0	<b>8</b>	<b>8</b>	0	0	0
0.1X	Low	0	<b>4</b>	<b>4</b>	0	0	0
1X	Control	0	0	0	NA	NA	NA

Note that NA means concatenation-based ML and quartet-based ASTRAL species trees both recovered the true species tree when analyzing full datasets 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.

Figure 1

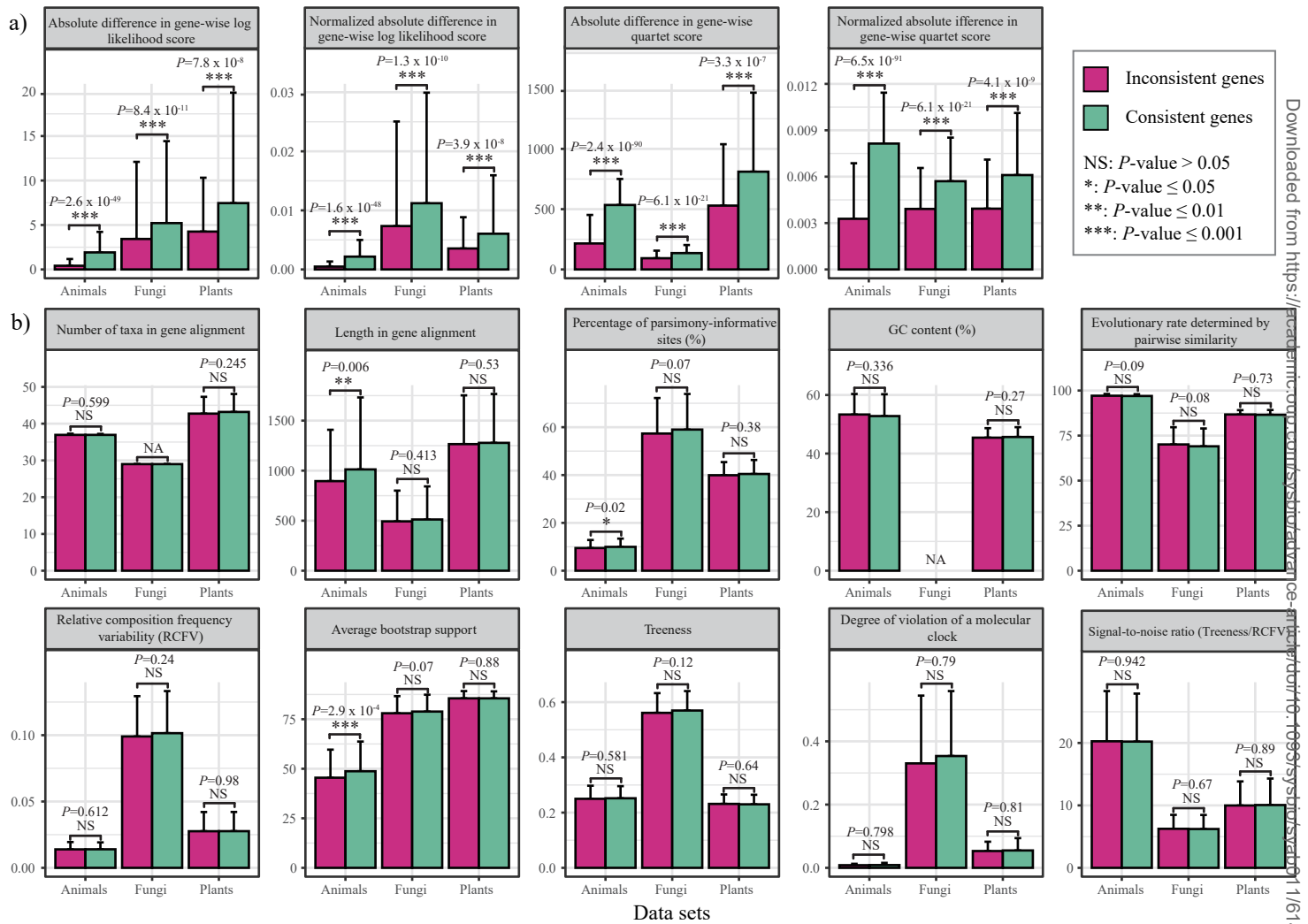


**Figure 2**

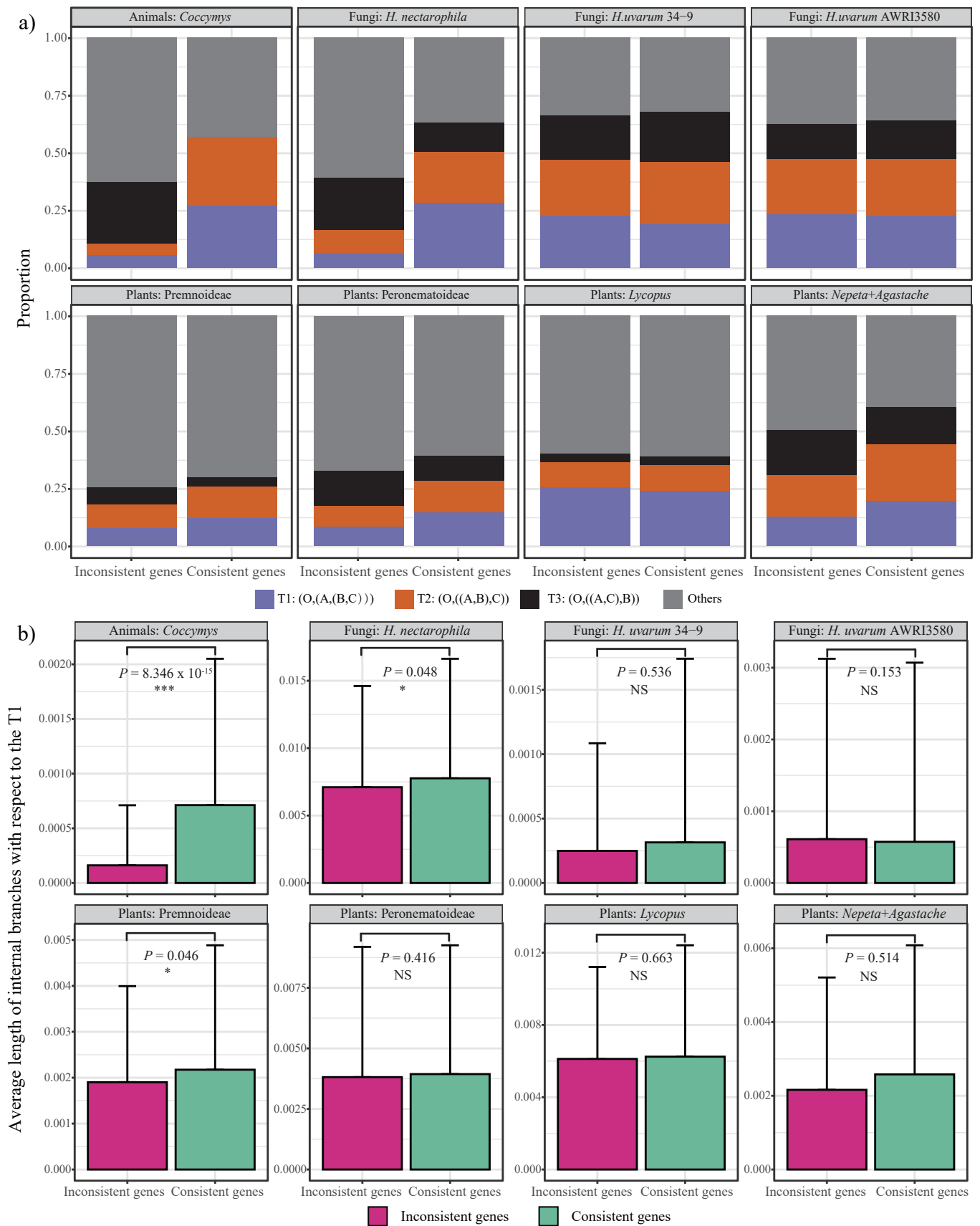




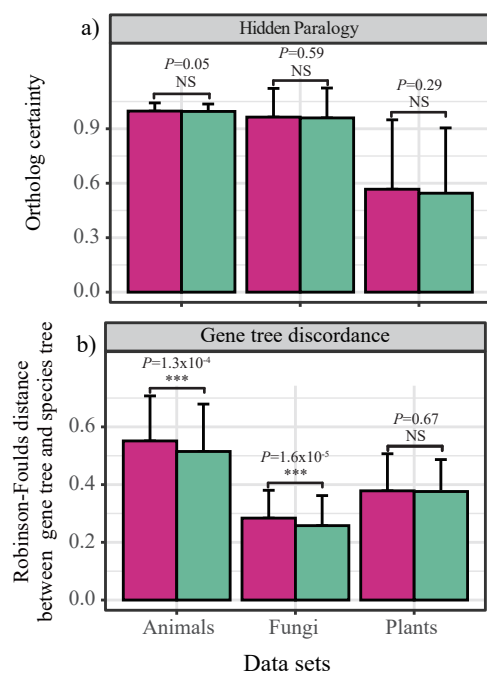
**Figure 3**



**Figure 4**



**Figure 5**



**Figure 6**

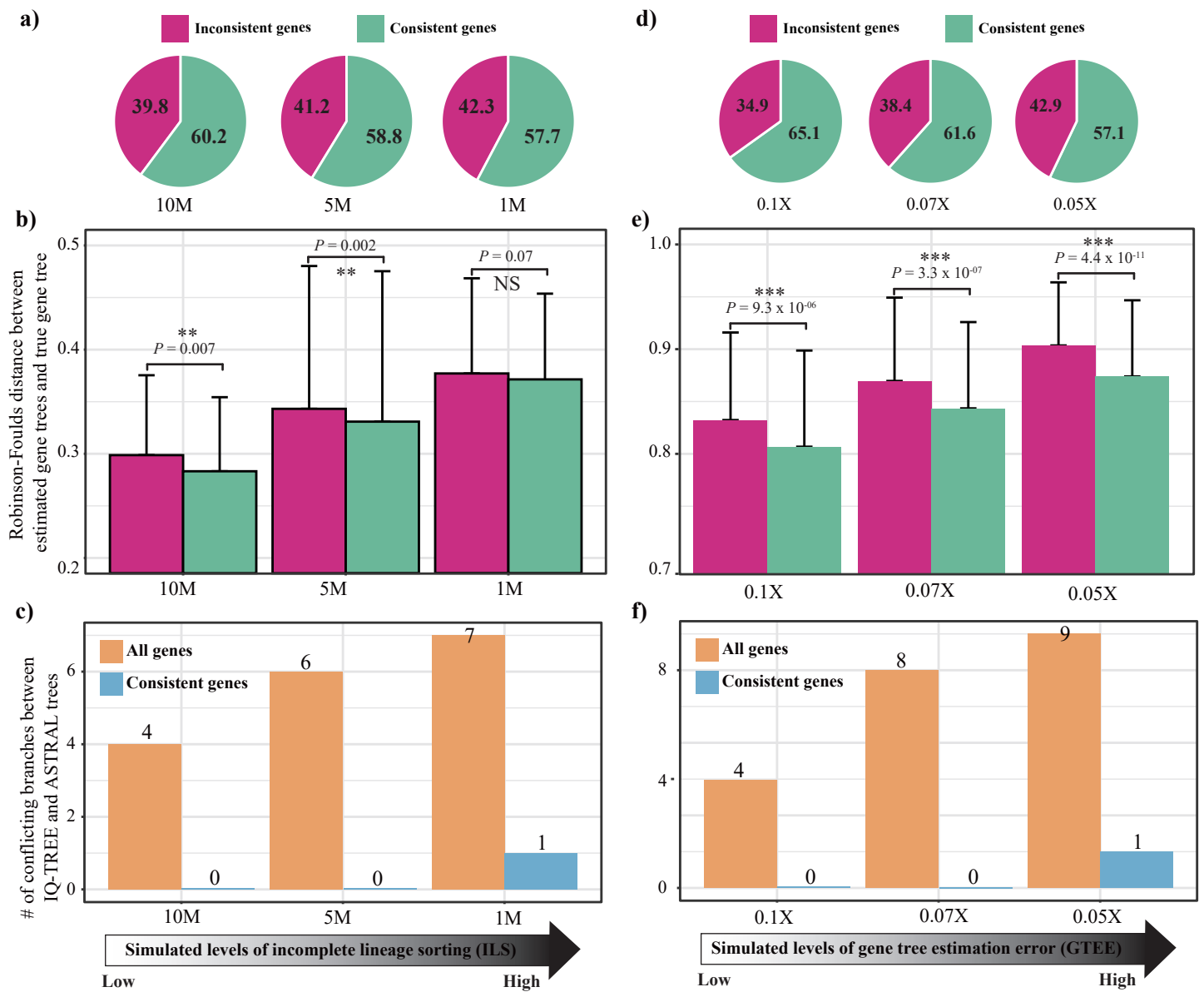
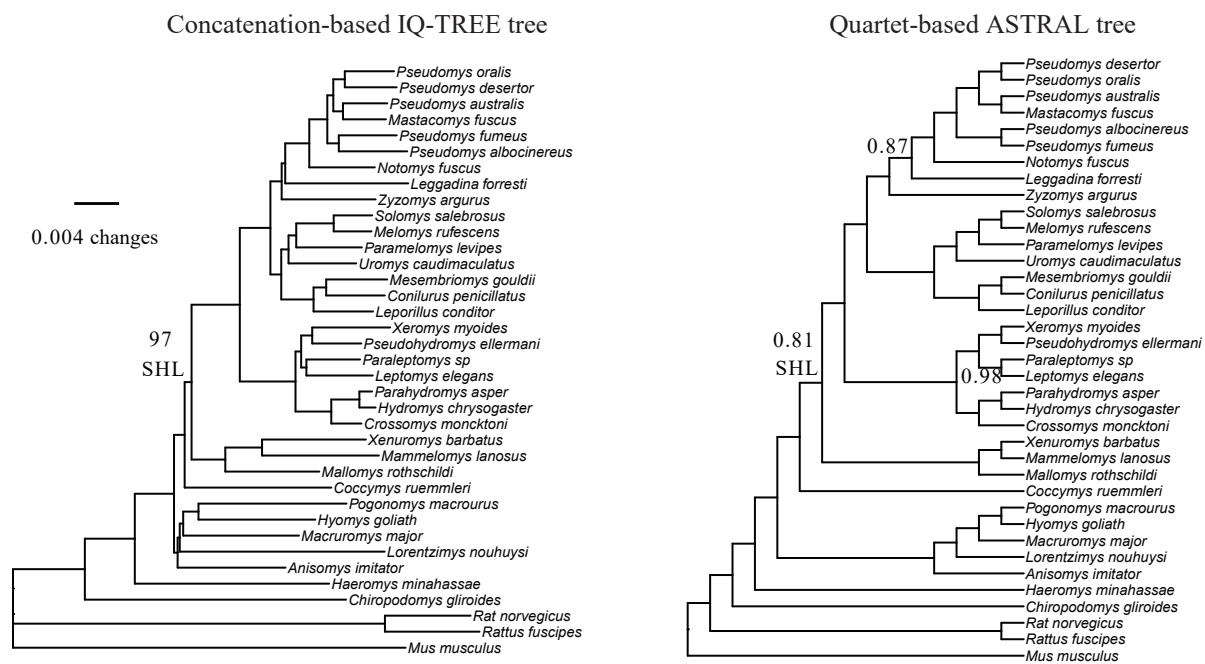


Figure 7



The data associated with this paper are available for review via Dryad.

The following is a temporary direct download link. Please copy and paste it directly into a web browser to download the data files to your computer (unfortunately this may not work as a link to click on)

The temporary link: <https://datadryad.org/stash/share/w-CrkUMQ-OTdSrzvIRxpmuUY3kTS2pVB6-Da30U1SH8>.