QT-WEAVER: Correcting quartet distribution improves phylogenomic analyses despite gene tree estimation error

Navid Bin Hasan[1,†], Sohaib[1,†], and Md. Shamsuzzoha Bayzid[1,*]

[1]Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
[†]These authors contributed equally to this work
[*]Corresponding author: shams_bayzid@cse.buet.ac.bd

Abstract

Summarizing individual gene trees into species phylogenies using coalescent-based methods has become a standard approach in phylogenomics. However, gene tree estimation error (GTEE) arising from a combination of reasons (ranging from analytical factors to more biological causes, as in short gene sequences) can potentially impact the accuracy of phylogenomic inference. We, for the first time, introduce the problem of correcting the quartet distribution induced by a set of estimated gene trees, which involves updating the weights of the quartets to better reflect their relative importance within the gene tree distribution. We present QT-WEAVER, the first method of its kind, which learns the conflicts within the quartet distribution induced by a given set of gene trees and generates an updated quartet distribution by adjusting the weights accordingly. QT-WEAVER is a general-purpose technique needing no explicit modeling of the subject system or reasons for GTEE or gene tree heterogeneity. Experimental studies on a collection of simulated and empirical data sets suggest that QT-WEAVER can effectively account for GTEE, which results in a substantial improvement in the species tree accuracy. Additionally, the concept of quartet conflicts and related algorithmic and combinatorial innovations introduced in this study will benefit various quartet-based computations. Therefore, QT-WEAVER advances the state-of-the-art in species tree estimation from gene trees in the face of GTEE. QT-WEAVER is freely available in open-source form at `https://github.com/navidh86/QT-WEAVER`.

1 Introduction

Species tree estimation from genes sampled across the whole genome has become routine with the advent of high-throughput sequencing technologies, generating genome-wide datasets that include hundreds or even thousands of loci. Species tree estimation from multiple genes is commonly done through concatenation (also known as "combined analysis"), which combines sequence alignments from different loci into a single supermatrix and then computes a tree on the supermatrix. While concatenation can produce accurate species trees when gene trees are concordant, it can be statistically inconsistent [1, 2], and produce incorrect trees with high support [3] when gene trees differ from the species tree due to various biological processes such as incomplete lineage sorting (ILS), gene duplication and loss (GDL), horizontal gene transfer (HGT), etc. As a result, summary methods [4–14] that combine estimated gene trees while explicitly accounting for gene tree discordance have drawn substantial interest among systematists. However, summary methods are susceptible to gene tree estimation errors (GTEE), which can arise from various reasons including short gene sequences, inaccurate alignments, and the limitations of the models and algorithms used for inferring gene trees from sequence data.

Due to the growing awareness that GTEE is a major contributor to inaccuracies of summary methods, there has been great interest in developing tools [15–24] to account for GTEE for improved species tree estimations. Most of these methods are species tree aware as they use a reference species tree in addition to the input gene trees. These species tree aware methods essentially reconcile/modify gene trees to make them closer in distance to the species tree by minimizing a species tree aware cost function. However, gene trees could be discordant and are not always expected to closely match the species tree Additionally, "integrative" methods such as ProfileNJ and TreeFix use available sequence data as well as the estimated gene trees and reference species trees. Obtaining a reasonably accurate reference species tree despite substantial amounts of GTEE is difficult and the inaccuracy in the reference tree may have cascading effects on gene tree corrections. More importantly, species tree aware methods, specifically TreeFix and TRACTION, have been criticized for their potential to increase GTEE where gene trees are discordant with species trees [25]. Bayesian techniques for co-estimating both gene trees and species trees such as BEST [26], *BEAST [27], and PHYLDOG [28] can produce substantially more accurate gene and species trees than other methods, but are not scalable to genome-level analyses [29–33]. Therefore, despite significant attempts to account for GTEE, substantial challenges remain.

In this study, we address the problem of GTEE in the context of species tree estimation, by formulating the *Quartet Distribution Correction* (QDC) problem for the first time, where we seek to "correct/adjust" the distribution of quartets induced by a given set of estimated gene trees. QDC attempts to account for GTEE without resorting to any reference species tree. Quartet-based summary methods have drawn considerable attention because quartets (unrooted gene trees with four taxa) can avoid the "anomaly zone" [34–36], where the most likely gene tree topology may differ from the true species tree topology and can produce highly accurate species trees. ASTRAL, the most widely used summary method, infers a species tree by maximizing the number of quartets in the gene trees that are consistent with the species tree. Another class of methods (known as quartet amalgamation techniques [12, 37–40]), such as wQFM [12] and wQMC [40], involves inferring weighted quartets for every group of four taxa (with weights correspond to the relative importance of the quartets) and then combining them into a single species tree. In this study, we present QT-WEAVER (**Q**uar**t**et **We**ight **A**djustment by **V**erifying, **E**stimating, and **R**ectifying quartet weights), a novel method that learns the weighted quartet distribution induced by a set of gene trees to identify and verify certain patterns of quartet conflicts and subsequently update the weights accordingly. We introduced the novel concept of "quartet conflicts" and proved key analytical results that underpin QT-WEAVER's ability to learn and adjust quartet distributions without relying on any reference species tree or sequence data, offering a robust approach to improving the accuracy of species tree inference.

The quartet distributions corrected by QT-WEAVER often align more closely with the true gene tree quartets than with those from the estimated gene trees, thereby effectively accounting for GTEE. Our experimental results, based on a diverse set of simulated and real biological datasets covering a wide range of challenging model conditions, demonstrate that amalgamating QT-WEAVER's corrected quartets significantly improves species tree inference accuracy. Moreover, QT-WEAVER is a general-purpose approach that does not require explicit modeling of the reasons for gene tree heterogeneity or GTEE, making it more resilient to model misspecification. QT-WEAVER is the first of its kind which advances the state-of-the-art in species tree estimation in the presence of GTEE.

2 Quartet Distribution Correction Problem

2.1 Problem Definition

Let $\mathcal{G} = \{g_1, g_2, \ldots, g_k\}$ be a set of k gene trees, where each g_i is a tree on taxon set $S_i \subseteq S$ (i.e., any gene tree g_i can be on the full set S of n taxa or can be on a subset S_i of taxa, making the gene tree incomplete). For a set of four taxa $a, b, c, d \in S$, the quartet tree $ab|cd$ denotes the unrooted quartet tree in which the pair a, b is separated from the pair c, d by an edge. Note that there are three alternative quartet topologies $(ab|cd, ac|bd, ad|bc)$ for four taxa. Let $\mathcal{Q}_i$ be the set of quartets in g_i. Therefore, $\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2 \cup \ldots \cup \mathcal{Q}_k$ is the multi-set of quartets present in $\mathcal{G}$. Note that there are $\binom{n_i}{4}$ quartets in $\mathcal{Q}_i$, and $3\binom{n}{4}$ quartets in $\mathcal{Q}$ are unique, as there are three alternative quartet topologies for a set of four taxa. The gene tree frequency (GTF) based weighted quartet distribution $\mathcal{QD}$ of $\mathcal{G}$ contains all possible quartets on n taxa along with their frequencies in the gene trees. That means, this is defined as the set $\{(q_1, w_1), (q_2, w_2), \ldots, (q_{\binom{3n}{4}}, w_{\binom{3n}{4}})\}$ of $3\binom{n}{4}$ tuples.

Let $\mathcal{G}_\mathcal{T}$ and $\mathcal{G}_\mathcal{E}$ be the set of true and estimated gene trees, respectively. Consequently, the definitions of $\mathcal{Q}_\mathcal{T}$, $\mathcal{Q}_\mathcal{E}$, $\mathcal{QD}_\mathcal{T}$, $\mathcal{QD}_\mathcal{E}$ extend in an obvious way. We now define the quartet distribution correction (QDC) problem as follows.

Problem	**Quartet Distribution Correction (QDC)**
INPUT	A weighted quartet distribution $\mathcal{QD}_\mathcal{E} = (q_1, w_1), (q_2, w_2), \ldots, (q_{\binom{3n}{4}}, w_{\binom{3n}{4}})$ induced by a set $\mathcal{G}_\mathcal{E}$ of estimated gene trees.
OUTPUT	A weighted quartet distribution $\mathcal{QD}'_\mathcal{E} = (q_1, w'_1), (q_2, w'_2), \ldots, (q_{\binom{3n}{4}}, w'_{\binom{3n}{4}})$ of updated weights w'_is for q_is so that the divergence/difference between $\mathcal{QD}'_\mathcal{E}$ and the true distribution $\mathcal{QD}_\mathcal{T}$ is minimized.

While we have defined the GTF-based weighted quartet distribution here, this concept extends to other weighting schemes as well.

2.2 Proposed Methodology: Identifying Quartet Conflicts

Our proposed approach is based on detecting patterns of "conflicts" among quartets, where a set of quartets is considered *conflicting* if they cannot be simultaneously satisfied by a single tree. Therefore, we try to learn the inherent patterns of conflicts within the quartets induced by the estimated gene trees and update the weights accordingly. We call the three alternative topologies ($ab|cd, ac|bd$, and $ad|bc$) for a set $\{a, b, c, d\}$ of four taxa the *topological variants* of each other, and clearly, they are conflicting. We now formalize the concept of conflicts between quartets in Theorem 1, which generalizes how the number of common taxa between two quartets affects the number of trees that can simultaneously satisfy both quartets.

The leaves in a tree T are denoted by $L(T)$. Every edge e in an unrooted leaf-labeled tree T defines a bipartition (or split) $\pi(e)$ on the leaves $L(T)$ (induced by the deletion of e).

Theorem 1. *Let q_1 and q_2 be two quartets drawn from a set of taxa. The number of binary trees on $L(q_1) \cup L(q_2)$ that simultaneously satisfy both q_1 and q_2 is determined by the number of taxa shared between them. Specifically, let ct denote the number of taxa common to both quartets. Then the possible values of ct are 0, 1, 2, 3, or 4, and the number of binary trees that satisfy both q_1 and q_2 is as follows:*

Case 1: *$ct = 0, 1, 2$ When the number of common taxa between q_1 and q_2 is less than 3, i.e., $ct = 0, 1, 2$, there will always be multiple binary trees on $L(q_1) \cup L(q_2)$ that satisfy both quartets.*

Case 2: *$ct = 3$ When q_1 and q_2 share exactly three common taxa ($ct = 3$), meaning each quartet has one unique taxon relative to the other, two sub-cases arise:*

1. *If the unique taxon in q_1 replaces the unique taxon in q_2 (meaning that when the unique taxon in quartet q_1 is replaced by the unique taxon in the other quartet q_2, the resulting topology will match that of q_2), there are exactly three distinct binary trees on $L(q_1) \cup L(q_2)$ that satisfy both quartets q_1 and q_2.*

2. *In all other configurations where the quartets share 3 common taxa, there is exactly one binary tree on $L(q_1) \cup L(q_2)$ that satisfies both quartets.*

Case 3: *$ct = 4$ When q_1 and q_2 share all 4 taxa ($ct = 4$), no binary tree can satisfy both quartets.*

Proof. We examine each case based on the number of common taxa between q_1 and q_2. Due to space constraints, we present the complete proof of this theorem in Appendix A. However, we present the proof for Case 2 ($ct = 3$) here as our proposed algorithm for QT-WEAVER is based on this particular case.

- **Case 2: $ct = 3$ (three common taxa)** When q_1 and q_2 share exactly three taxa, the number of trees that can satisfy both quartets depends on the relationship between the unique taxon in each quartet. Two distinct sub-cases arise:

  - Sub-case 1 (replacement of unique taxa): If the unique taxon in q_1 simply replaces the unique taxon in q_2 (preserving the relationship between the shared taxa) and vice versa, then two of the common taxa will be sister species in both q_1 and q_2. In this case, exactly three distinct binary trees on the set $L(q_1) \cup L(q_2)$ of five leaves satisfy both quartets. These three trees correspond to the possible ways to arrange the unique taxa from q_1 and q_2 with respect to the shared taxa. For example, $q_1 = ab|cd$ and $q_2 = ab|ce$ satisfy this condition where d and e are the unique taxa in q_1 and q_2 respectively and replacing d with e in q_1 (or e with d in q_2) makes q_1 and q_2 identical. Moreover, a and b are closer to each other than they are to other species in both q_1 and q_2. In this case, we can place e on three branches in q_1 (as shown in Figure 1a), resulting in three different trees on five taxa $\{a, b, c, d, e\}$ that satisfy both q_1 and q_2. Similarly, we can place d on three branches in q_2 (as shown in Figure 1b), producing three different trees that satisfy both q_1 and q_2. Note that the set of three trees on five taxa resulting from different placedments of e in Figure 1a are identical to the set of of three trees resulting from different placements of d in Figure 1b. We now consider two other branches (incidents on a and b) in $ab|cd$. First, if taxon e is placed as the sister to a, one of the resulting induced quartets is $ae|bc$, which does not satisfy $ab|ce$ as e is now the sister to a. Similarly, if e is placed as the sister to b, the quartet $ac|be$ is formed, which also is not consistent with $ab|ce$. Thus, there are exactly three trees on $\{a, b, c, d, e\}$ that satisfy both q_1 and q_2.

(a) Placements of e on the quartet $ab|cd$ **(b)** Placements of d on the quartet $ab|ce$

Figure 1: Placements for e and d on q_1 and q_2 respectively.

  - Sub-case 2 (other configurations): In all other scenarios where the unique taxon in q_1 does not simply replace the unique taxon in q_2 (and vice versa), there is exactly one binary tree on the set $L(q_1) \cup L(q_2)$ of five taxa that satisfies both quartets. This is because the three common taxa, along with their relationships to the unique taxa, fully constrain the tree structure. Here, unlike sub-case 1, no pair of taxa among the three common taxa are sisters in both quartets. Consider the two quartets, $q_1 = ab|cd$ and $q_2 = ae|bd$ that satisfy this condition. In order to find a tree that satisfies both q_1 and q_2, we can insert e to q_1 by making it a sister of a, as shown in Figure 2a. Similarly, we can insert c into q_2 by making it the sister of d as shown in Figure 2b. Note that these two trees are identical (shown in shown in Figure 3a). Furthermore, the placements of e on four other edges in q_1 (the edges incident on b, c, d, and the internal edge) result in trees that do not support q_2 ($ae|bd$) as b becomes closer to a than e is to a. Similarly, the placements of c on four other edges in q_2 do not produce any trees that satisfy q_1 ($ab|cd$) as b becomes closer than c to d. Thus, there is exactly one tree (as shown in Figure 3a) that satisfies both q_1 and q_2.

□

We define *canonical quartet pair* as a pair of quartets (q_1, q_2) that can be satisfied by exactly one tree T, where $L(T) = L(q_1) \cup L(q_2)$, as in Sub-case 2 of Case 2 in Theorem 1. Thus, a canonical quartet pair

(a) Placement of e over the quartet $ab|cd$ **(b)** Placement of c over the quartet $ae|bd$

Figure 2: Unrooted trees satisfying both quartets $ab|cd$ and $ae|bd$

(a) Only unrooted tree satisfying $ab|cd$ and $ae|bd$. **(b)** Only unrooted tree satisfying $ab|cd$ and $ac|be$

Figure 3: Unrooted trees satisfying canonical quartet pairs.

(q_1, q_2) uniquely represent a tree on $L(q_1) \cup L(q_2)$ that satisfy both q_1 and q_2. For example, $q_1 = ab|cd$ and $q_2 = ae|bd$ is one such canonical quartet pair, which is satisfied by exactly one tree T on $\{a, b, c, d, e\}$ as shown in Figure 3a. There are $\binom{5}{4} = 5$ quartets ($ae|bd$, $ae|bc$, $ae|cd$, $ab|cd$, and $be|cd$) in T including the canonical pair corresponding to this tree ($ab|cd$ and $ae|bd$). The canonical pair of quartets will clearly be in conflict with the topological variants of the three other quartets, $ae|bc$, $ae|cd$, and $be|cd$. Note that each quartet has two other topologically distinct variants. Thus, we have the following Corollary 1.1.

Corollary 1.1. *Every canonical pair of quartets (q_1, q_2) is in conflict with six quartets on $L(q_1) \cup L(q_2)$.*

We now have the Corollaries 1.2-1.5. The proofs and related discussions are presented in Appendix A.

Corollary 1.2. *Every binary unrooted tree T with five taxa contains a total of four canonical pairs of quartets (q_1, q_2), (q_2, q_3), (q_3, q_4) and (q_4, q_1) with each of the four quartets q_1, q_2, q_3, q_4 being present in two canonical pairs.*

Corollary 1.3. *Every quartet is a part of eight canonical pairs of quartets, with the other members of these pairs having a particular unique taxon with respect to q_1.*

Corollary 1.4. *Every quartet q is part of 28 unique conflicting sets as a member of canonical pairs for each unique taxon $x \notin L(q)$.*

Corollary 1.5. *For a set S of n taxa, every quartet q, where $L(q) \in S$, is in $28 \times (n-4)$ unique conflicting sets as a member of canonical pairs.*

Let q_i, q_j, q_k be a conflicting set, where q_i and q_j form the canonical pair in conflict with q_k. Thus, while q_i, q_j, and q_k cannot all coexist in the same tree, any two of them can. Consequently, the presence of q_i in a species tree becomes less likely if the weights of both q_j and q_k increase (since higher weights for q_j and q_k indicate the possibility that both of them are present in the species tree). We now define the *conflict score* of a quartet as the sum of the products of the weights of pairs of quartets that are in conflict with it. Formally, for a given quartet q_i and a set Z_i containing the pairs of quartets that are in conflict with q_i, the conflict score $C(q_i)$ is calculated as follows:

$$C(q_i) = \sum_{\{(q_j, q_k) \mid (q_j, q_k) \in Z_i\}} w_j \cdot w_k \tag{1}$$

There are $28 \cdot (n-4)$ unique conflicting sets in each of which q_i is present (Corollary 1.5). By removing q_i from all these sets, we obtain a set of pairs Z_i, where $|Z_i| = 28 \cdot (n-4)$.

$$Z_i = \{(q_j, q_k) \mid q_j, q_k \in \mathcal{Q} \text{ and } q_j, q_k \text{ are in conflict with } q_i\}$$

Note that in addition to the product of weights (PoW) based definition of the conflict score (Equation 1), we can define the conflict score using a minimum of weights (MoW) based approach as in Equation 2.

$$C(q_i) = \sum_{\{(q_j,q_k)\,|\,(q_j,q_k)\in Z_i\}} min(w_j, w_k) \tag{2}$$

The rationale behind this MoW-based definition is that even if one of w_j or w_k is very large, it does not necessarily challenge the presence of q_i as long as the other is quite low. Therefore, considering the minimum of the two values (w_j and w_k) as a contributor to the conflict score of q_i is a natural choice.

2.3 Algorithm: QT-WEAVER

Let $\mathcal{QD}_{\mathcal{E}} = \{(q_1, w_1), (q_2, w_2), \ldots, (q_{\binom{3n}{4}}, w_{\binom{3n}{4}})\}$ be the set of weighted quartets of a set $\mathcal{G}_{\mathcal{E}}$ of estimated gene trees, S be the set of all taxa present in $\mathcal{G}_{\mathcal{E}}$ and $\mathcal{Q} = \{q_1, q_2, \ldots, q_{\binom{3n}{4}}\}$ be the set of all quartets in $\mathcal{QD}_{\mathcal{E}}$. The algorithm takes the set $\mathcal{QD}_{\mathcal{E}}$ as input and for each quartet $q_i \in \mathcal{QD}_{\mathcal{E}}$, the algorithm adjusts the weight w_i to w'_i using the concept of quartet conflict we introduced in Section 2.2. We maintain a set $\mathcal{A} \subseteq \mathcal{Q}$, which represents the set of quartets whose weights have already been adjusted. Initially, $\mathcal{A} = \emptyset$, indicating that no quartet weights have been adjusted at the start of the algorithm.

We iterate over the set of quartets, $q \in \mathcal{Q}$. If a quartet q is already adjusted, i.e., $q \in \mathcal{A}$, we skip the quartet. Otherwise, let $\mathcal{T} = \{q_1, q_2, q_3\}$ be the set of three possible quartets on $L(q)$. We then adjust the weights of the quartets in $\mathcal{T}$ and insert them to $\mathcal{A}$, i.e., $\mathcal{A} \leftarrow \mathcal{A} \cup \mathcal{T}$. To adjust the weight for each quartet q_i in $\mathcal{T}$ ($i \in \{1, 2, 3\}$), we construct Z_i, a set of all quartet pairs conflicting with q_i.

We do this by iterating over the set of taxa not present in q_i, i.e. $S \setminus L(q_i)$. For each $t \in S \setminus L(q_i)$, we obtain a set of 28 quartet pairs in conflict with the quartet q_i (Corollary 1.4). We add all such pairs to Z_i. Thus, Z_i contains $28 \times |S \setminus L(q_i)| = 28 \times (n-4)$ pairs of quartets. Next, we compute the *conflict score* for each of the quartets in $\mathcal{T}$ using the PoW or MoW-based formula in Equation (1) or Equation (2) (this is a user defined configuration in our algorithm).

Note that the conflict score of a quartet q_i indicates a level of inconsistency of q_i with respect to other quartets in the distribution. Therefore, a quartet with a higher conflict score should have a lower readjusted weight (i.e., we want to decrease its relative importance as this quartet has a higher degree of conflict with others). Therefore, we obtain $\tilde{w_i}$ by dividing w_i by its conflict score $C(q_i)$.

$$\tilde{w_i} = \frac{w_i}{C(q_i)}$$

This operation on three quartets in $\mathcal{T}$ denormalizes the weights of the quartets, which must be restored to their original sum $s = w_1 + w_2 + w_3$. To normalize these weights, we multiply each by $\frac{s}{\tilde{w_1}+\tilde{w_1}+\tilde{w_3}}$ so that their sum equals their original sum s. The final normalized weight w'_i is given by:

$$w'_{i\in\{1,2,3\}} = \frac{s}{\tilde{w_1} + \tilde{w_2} + \tilde{w_3}} \cdot \tilde{w_i}$$

The pseudo-code for the QT-WEAVER algorithm is presented in Appendix B

3 Experimental Study

3.1 Datasets

We evaluated the performance of QT-WEAVER using a collection of previously studied simulated and biological datasets. We studied two simulated datasets: a 37-taxon mammalian dataset based on biological data [41] and a 15-taxon dataset, both generated in prior studies [42, 43]. These datasets were created through a multi-stage simulation process, starting with a species tree, followed by the simulation of gene trees under the multi-species coalescent model (which can result in gene trees topologically distinct from the species tree), and finally, the simulation of gene sequence alignments down the gene trees. The datasets exhibit varying degrees of incomplete lineage sorting (ILS), ranging from moderate to high levels, and differ in the number of genes and the extent of gene tree estimation errors (controlled by sequence lengths). Thus, these simulated datasets provide a wide array of conditions under

which we assessed the performance of QT-WEAVER. Table A3 in Appendix D presents a summary of these datasets. We also evaluated QT-WEAVER on a challenging biological avian dataset from Jarvis *et al.* [44] comprising 14,446 genes sampled from 48 birds.

3.2 Species tree estimation methods

We used wQFM [12], the best existing weighted quartet amalgamation technique, to estimate species trees from weighted quartets. We also used wQMC, which is another well-known weighted quartet amalgamation technique. We ran GTF based wQFM and wQMC on uncorrected embedded quartets in the input gene trees with weights reflecting the frequencies of the quartets. wQFM (and wQMC) was also run on the adjusted/corrected weighted quartets, generated by QT-WEAVER, to demonstrate the impact of quartet weight correction on species tree estimation. We refer to this variant as *wQFM-corrected* or *wQFM+QT-WEAVER* interchangeably. We compared wQFM with ASTRAL-III [5, 45] (v. 5.7.8), which is the leading quartet based species tree method. Note that ASTRAL cannot take a set of weighted quartets as input, and as such, we cannot evaluate ASTRAL on corrected quartet distributions.

3.3 QT-WEAVER configurations

As discussed in Sections 2.2, 2.3, we can define the conflict score of a quartet q using PoW or MoW-based approaches (Equations 1 and 2). Additionally, beyond using all 28 conflicting sets that a quartet q belongs to for each taxon $e \notin L(q)$, we explore subsets of these conflicting sets, analyzing subsets of sizes four, six, and eight. Thus, we have explored eight configurations (two weighting schemes combined with four conflicting set subsets). Our empirical exploration using the 15-taxon datasets (results presented in Appendix C) suggests that the MoW-based approach is preferable over the PoW-based scheme. Moreover, utilizing six conflicting sets is both computationally faster and yields better performance across all conditions. Therefore, for the remaining experiments, we run QT-WEAVER with the configuration that uses 6 (out of 28) conflicting sets and the MoW-based approach to update quartet weights.

3.4 Measurements

We evaluated the accuracy of the estimated trees on simulated datasets by comparing them to the model species tree using the normalized Robinson-Foulds (RF) distance [46].Additionally, we assessed quartet scores, reflecting the number of quartets from the gene trees that are consistent with the estimated species tree. For the biological dataset, we compared the inferred species trees with those reported in the scientific literature. Multiple replicates under different model conditions were analyzed, and statistical significance between methods was determined using a two-sided Wilcoxon signed-rank test (with $\alpha = 0.05$). To assess the accuracy of quartet distributions, we compared both corrected and uncorrected distributions to the true quartet distributions derived from the true gene trees using Jansen-Shannon divergence [47].

4 Results and Discussion

4.1 Results on 37-taxon dataset

The average RF rates of wQFM using both uncorrected and corrected (by QT-WEAVER) quartet distributions and ASTRAL on various model conditions in the 37-taxon dataset are shown in Figure 4(a). We vary the gene tree estimation error (by varying the sequence length from 50 to 1000 bp), the amount of ILS (by multiplying or dividing all internal branch lengths in the model species tree by two – producing three model conditions that are referred to as 1X (moderate ILS), 0.5X (high ILS) and 2X (low ILS)), and the number of genes (from 100 to 500). In general, wQFM+QT-WEAVER is more accurate than wQFM across all model conditions – clearly demonstrating the benefit and positive impact of correcting quartet distributions for GTEE by QT-WEAVER. wQFM is, in general, more accurate than ASTRAL (which was also reported by prior studies [12, 48]). However, the improvements of wQFM over ASTRAL are often not statistically significant. Remarkably, when wQFM is run on weighted quartet distributions corrected by QT-WEAVER, wQFM+QT-WEAVER becomes notably better than ASTRAL and in most cases (six out of nine), the improvements are statistically significant ($p \ll 0.05$). Moreover, the accuracy achieved by ASTRAL and wQFM on true gene trees (i.e., without any estimation error)

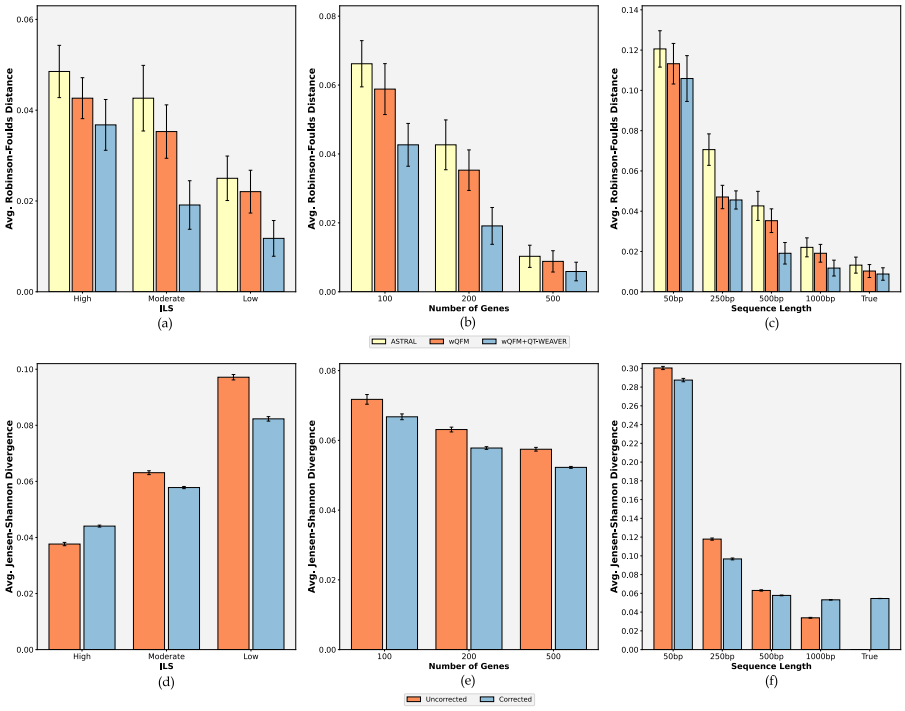


Figure 4: Results on 37-taxon dataset. (a)-(c) We show the average RF rates with standard errors over 20 replicates for the methods ASTRAL, wQFM (wQFM-uncorrected), and wQFM+QT-WEAVER (wQFM-corrected). (a) The amount of ILS was varied from high to low, keeping the number of genes fixed at 200 and sequence length at 500bp. (b) The number of genes was varied keeping ILS fixed at moderate (1X) and the sequence length at 500bp. (c) Varying sequence lengths with moderate ILS and 200 genes. We also include true gene trees with no GTEE. (d)-(f) Average Jensen-Shannon divergence of the quartet distributions before and after correction with respect to the true quartet distributions.

was matched by wQFM+QT-WEAVER even when using estimated gene trees derived from sequences as short as 500 bp. In fact, wQFM+QT-WEAVER with 1000 bp sequences is slightly more accurate than ASTRAL on true gene trees. These clearly show the power and efficacy of QT-WEAVER in accounting for gene tree estimation error. Even though there is no GTEE in true gene trees, the limited number of genes can result in a weighted quartet distribution that may fail to represent the true species trees. Interestingly, indeed, wQFM+QT-WEAVER outperformed both ASTRAL and wQFM on true gene trees, indicating that adjusting the weighted quartet distribution of true gene trees can still lead to more accurate species trees. We also evaluated the performance of wQMC, another well-known weighted quartet amalgamation technique, on quartets corrected by QT-WEAVER. Although wQMC is generally less accurate than wQFM [12, 48], we assessed wQMC+QT-WEAVER to demonstrate the usability of the corrected quartets generated by QT-WEAVER across different amalgamation methods. As shown in Figure A5 in Appendix E.1, wQFM is better than wQMC (both for corrected and uncorrected data) and wQMC+QT-WEAVER consistently outperforms wQMC – further demonstrating the effectiveness of QT-WEAVER in adjusting quartet distributions and the superiority of wQFM over wQMC.

ASTRAL has been evolved and improved over successive versions for both accuracy and scalability, offering the theoretical guarantee of identifying the species tree that maximizes quartet support within the search space defined by the bipartitions of the input gene trees and thereby leaving limited (or no) scope for further enhancements in quartet score optimization. Thus, addressing GTEE becomes a critical avenue for pushing the accuracy frontier of quartet-based methods. In this context, the substantial improvements achieved by wQFM+QT-WEAVER over ASTRAL is particularly noteworthy.

Next, we performed a series of experiments to further investigate the impact and quality of the quartet distributions produced by QT-WEAVER. First, to assess the similarity between the estimated quartet distributions (before and after correction) and the true quartet distributions (inferred from the true gene trees), we compare the Jensen-Shannon divergence between estimated and true quartet distributions in Figure 4(d)-(f) and Table A4 in Appendix E.1. The corrected quartet distributions almost always

have lower divergence than the uncorrected distributions, except for model conditions with very low divergence to begin with (less than 4 percent; 0.5x-200gt-500bp and 1X-200gt-1000bp conditions). The model conditions with relatively long sequence length (1000bp) have low amounts of GTEE and have less divergence to begin with (less than 4 percent), and correcting it does not seem to further reduce the divergence (rather making the distribution slightly more diverged). However, the corrected distributions overall result in better species trees as reflected by the RF rates (Figure 4(c)). Overall, these results suggest that QT-WEAVER effectively bridges the gap with the overall nature of the true distribution.

Next, we computed the quartet scores of different estimated species trees and the true species tree with respect to both estimated and true gene trees (see Tables A5 and A6). When the quartet scores are computed based on original (uncorrected) estimated quartet distribution (Table A5), even though ASTRAL-estimated trees are less accurate than wQFM and wQFM+QT-WEAVER, ASTRAL achieves higher quartet scores than wQFM since ASTRAL is guaranteed to maximize the quartet score within a constrained search space. However, it "overshoots" the quartet score as it returns trees with higher quartet scores than the quartet score of the true tree. This is mostly because the statistical consistency of quartet score maximization criterion may not hold in the presence of gene tree estimation errors and limited number of genes. The true tree having the lower quartet scores across all model conditions supports this claim (see also [49–51] for more related discussions). The quartet scores of wQFM, especially wQFM+QT-WEAVER, with respect to the uncorrected estimated gene trees, are closer to the true quartet score than the scores of ASTRAL-estimated trees are to the true score. These further explain the superior performance of wQFM+QT-WEAVER over wQFM and ASTRAL. We also report the quartet score of wQFM+QT-WEAVER with respect to the quartet distribution corrected by QT-WEAVER. Notably, wQFM+QT-WEAVER has the highest quartet score when the score is computed based on the corrected quartet distribution. We note that these scores are not directly comparable to other methods' quartet scores as other reported scores are computed with respect to a different quartet distribution (i.e., original uncorrected distribution). Therefore, we cannot consider these scores as "over-estimation" compared to true quartet scores. Rather, it suggests that the weight adjustment/correction by QT-WEAVER results in quartet distributions that may lead to species trees with higher quartet consistency scores (w.r.t corrected distributions), which were not attainable with the uncorrected distributions.

Finally, we examined the quartet scores with respect to true gene trees with no estimation errors (Table A6). The superiority of wQFM as a quartet-based tree estimation method and the efficacy of QT-WEAVER as a quartet correction method are even more evident from these quartet scores. In most cases, wQFM and wQFM+QT-WEAVER achieve higher and closer (to true quartet score) quartet scores than ASTRAL. This implies that the corrected distributions result in species trees that better correspond to the true gene tree distribution – supporting the trends observed in RF rates (Figure 4(a)).

4.2 Results on 15-taxon dataset

In the simulated 15-taxon datasets, we evaluated the performance on varying gene tree estimation error using 100bp and 1000bp sequence lengths and on varying numbers of gene trees (100 and 1000). Similar to the 37-taxon dataset, wQFM+QT-WEAVER consistently outperforms ASTRAL and wQFM across all four model conditions (Figure 5). In particular, wQFM+QT-WEAVER is significantly better ($p \ll 0.05$) than ASTRAL on challenging model conditions with short sequences (100bp), i.e., high GTEE.

Similar to 37-taxon dataset, we investigated the divergence of uncorrected and corrected quartet distributions from true quartet distributions (Figure 5(c)-(d) and Table A7) and the quartet scores of different estimated species trees with respect to both estimated and true gene trees (see Appendix E.2). The Jensen-Shannon divergence clearly improves after correction when the gene tree estimation error is high (100bp). Thus, the quartet distributions inferred from gene trees with high estimation error realign with the true distribution after correction using QT-WEAVER. On the other hand, when the distribution has low amounts of GTEE with relatively long (1000bp) sequences and thus very low divergence to begin with ($\leq 5\%$), correcting it does not seem to reduce the difference. Interestingly, even though the divergence increases slightly for the more accurate distributions (1000bp), the corrected distribution as a whole seems to be more representative of the true distribution, as supported by the better RF rates.

We again assess the quartet scores like we did for the 37-taxon dataset in Tables A8 and A9, and observed similar trends. wQFM+QT-WEAVER achieves higher and closer (to true quartet score) quartet scores than ASTRAL and wQFM across all four model conditions when the quartet scores are computed based on true gene trees – further demonstrating the efficacy of QT-WEAVER in accounting for GTEE.

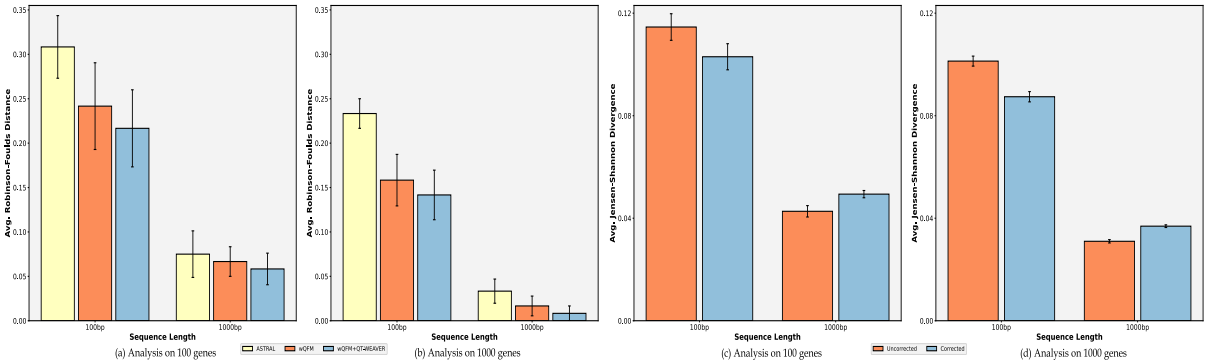


Figure 5: Results on 15-taxon dataset. (a)-(b) Average RF rates of ASTRAL, wQFM, and wQFM+QT-WEAVER on varying model conditions with varying sequence lengths and numbers of genes. (c)-(d) Average Jensen-Shannon divergence for the uncorrected and corrected quartet distributions.

5 Iterative corrections using QT-WEAVER

The demonstrated effectiveness of using QT-WEAVER in adjusting quartet distribution begs the question: what happens if we iteratively correct the adjusted distribution, using the output of QT-WEAVER as input to QT-WEAVER for the next iteration? To investigate this, we performed 50 iterations and analyzed how different evaluation metrics evolve in Figure 6. Notably, in model conditions with high gene tree estimation errors (100bp), RF rates steadily and dramatically improve (Figure 6a). For the 100bp-100gene condition, the RF rate drops from 25% to 11% after 15 iterations of weight corrections, and for the 100bp-1000gene condition, the RF rate decreases from 16% to as low as 2% after 10 iterations. Note that the RF rates of ASTRAL on these two model conditions are 31% and 23%, respectively, as indicated by the horizontal lines in Figure 6a. However, in model conditions with well-estimated gene trees (1000bp sequences) where the initial RF rates are already low ($\sim$ 5%), repeated iterations may initially improve accuracy (e.g., a drop from 7% to 1% in the 1000bp-100gene condition) but eventually lead to performance degradation. This occurs because continuously adjusting well-estimated and corrected quartet distributions (as in cases with long sequences and large numbers of genes) can distort the weights to an extent that can mislead the tree search algorithm toward less accurate trees.

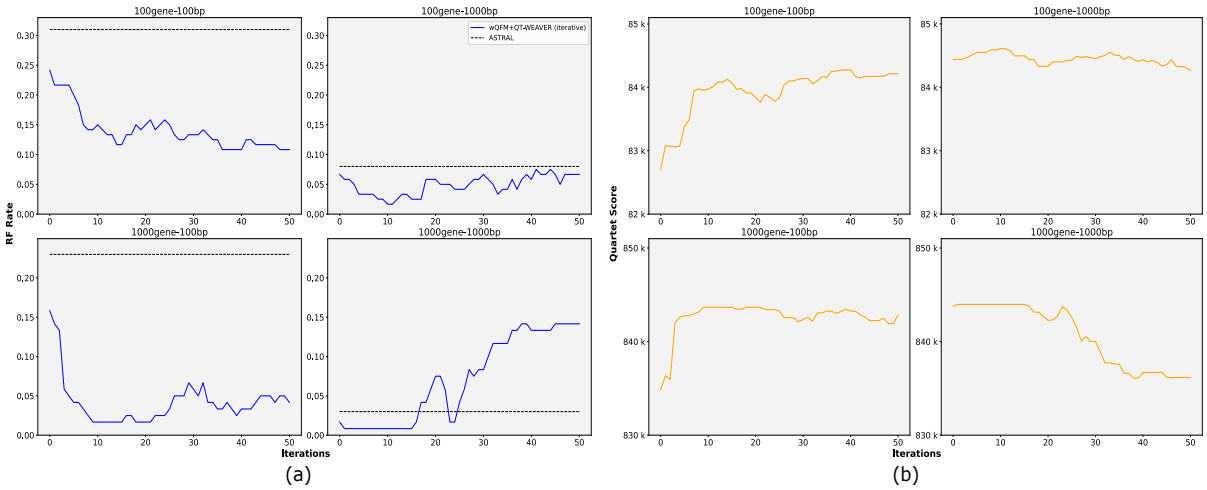


Figure 6: Iterative corrections on the 15-taxon dataset. We show the changes in RF rates and quartet scores (with respect to true gene trees) of wQFM+QT-WEVER over 50 iterations of weight corrections using QT-WEAVER. ASTRAL's RF rates are indicated by dashed horizontal lines in (a).

In Figure 6(b), we show the changes in quartet scores (w.r.t the true gene tree distributions) for wQFM+QT-WEAVER over multiple iterations. As we can see, the increase and decrease in the quartet scores over multiple iterations are perfectly aligned with the decrease and increase of the RF rates,

respectively. When the GTEE is high, the scores maintain a rising trend. On the other hand, on the low GTEE model conditions, the quartet scores start to degrade after a few iterations in a pattern identical to the RF rate degradation. Additionally, we investigated the Jensen-Shanon divergence and quartet scores (relative to corrected distributions) across iterations (Figure A6 and related discussion in Appendix E.3).

Thus, performing multiple iterations of weight adjustment shows potential, though improvements may not be consistent across all datasets. For instance, in the 37-taxon dataset, no further improvement was seen beyond the first iteration, as in most cases, the RF rates dropped below 5% after one iteration, and successive iterations distorted the distributions to an extent that resulted in worse trees. Nonetheless, QT-WEAVER's iterative approach shows promise and warrants further investigation to develop a robust stopping criterion and scoring scheme for selecting the best tree from successive iterations.

6 Results on biological avian dataset

We have reanalyzed the avian biological dataset from [44], which has 14,446 genes (including exons, introns, and UCEs) across 48 taxa. This dataset contains high levels of gene tree discordance likely driven by rapid radiation events in the evolutionary history of these species. Mahbub et al. [12] compared ASTRAL, wQFM, and wQMC with the binned MP-EST tree which was presented in Jarvis *et al.* [44]. We include wQFM+QT-WEAVER in this comparison (Figure A7 in Appendix F).

All three estimated trees are highly congruent with the reference binned tree, with wQFM+QT-WEAVER being the most congruent and ASTRAL the least. The wQFM trees (on uncorrected and corrected distributions) are almost identical, with mainly one noticeable difference: wQFM+QT-WEAVER was able to correctly resolve the Cursores clade (crane and killdeer), which both wQFM and ASTRAL failed to recover. ASTRAL failed to recover Otidimorphae (bustard, turaco, and cuckoo), whereas both wQFM and wQFM+QT-WEAVER reconstructed this clade.

All three methods successfully reconstructed the well-established Australaves clade (passeriformes, parrots, falcons, and seriemas). They also recovered the Afroaves, Core Waterbird, and Caprimulgimorphae clades successfully. ASTRAL failed to recover Otidimorphae, unlike both the wQFM methods. All three failed to recover Columbea but were able to recover the constituent clades Columbimorphae (mesite, sandgrouse, and pigeon) and Phoenicopterimorphae (flamingo and grebe).

7 Running Time

The running time of QT-WEAVER solely depends on the number of taxa n and not on the number of genes, which determines the number of unique quartets ($3\binom{n}{4}$) in the weighted quartet distribution. All analyses were run on the same machine with AMD Ryzen 7 5800H CPU (8 cores), 16GB RAM, and NVIDIA GeForce RTX 3060 GPU (6GB memory). The simulated 15-taxon and 37-taxon datasets took 1.65 and 242 seconds on average per replicate. The biological avian dataset with 48 taxa took 983 seconds to run.

8 Conclusions

This study, for the first time, introduces the quartet distribution correction problem and shows the impact and clear benefit of using quartet distributions corrected by QT-WEAVER for improved species tree estimations. QT-WEAVER learns the overall quartet distribution based on the pattern of quartet conflicts (a concept that we have introduced in this study), and seeks to update the weights to better reflect their relative importance. The concept of quartet conflict and related theoretical results have broad applicability and will be valuable for a range of quartet-based computational methods. Our experimental study shows that QT-WEAVER may result in substantial improvements over the leading method ASTRAL. Therefore, the idea of estimating species trees by correcting quartet distributions has merit and should be pursued and used in future phylogenomic studies. As a future study, we plan to evaluate QT-WEAVER on a diverse set of real biological datasets as the pattern of quartet conflicts are sufficiently complex and heterogeneous across various datasets. Another important research direction is the automatic selection of QT-WEAVER configurations. Additionally, considering the dramatic improvements achieved from multiple iterations of QT-WEAVER on certain datasets and model conditions, future studies need to investigate how to automatically identify an appropriate QT-WEAVER configuration for a given input, based on the input gene tree topologies, to ensure iterative enhancement.

References

[1] Sebastien Roch and Mike Steel. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theoretical Population Biology, 100:56–62, 2015.

[2] J H Degnan, M DeGiorgio, D Bryant, and N A Rosenberg. Properties of consensus methods for inferring species trees from gene trees. Systematic Biology, 58:35–54, 2009.

[3] L S Kubatko and J H Degnan. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology, 56:17, 2007.

[4] Liang Liu, Lili Yu, and Scott V Edwards. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology, 10:302, 2010.

[5] Siavash Mirarab, Rezwana Reaz, Md S Bayzid, Théo Zimmermann, M Shel Swenson, and Tandy Warnow. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics, 30(17):i541–i548, 2014.

[6] B Larget, S K Kotha, C N Dewey, and C Ané. BUCKy: Gene tree/species tree reconciliation with the Bayesian concordance analysis. Bioinformatics, 26(22):2910–2911, 2010.

[7] E Mossel and S Roch. Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 7(1):166–171, 2011.

[8] L S Kubatko, B C Carstens, and L L Knowles. Stem: Species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics, 25:971–973, 2009.

[9] Julia Chifman and Laura Kubatko. Quartet from snp data under the coalescent model. Bioinformatics, 30(23):3317–3324, 2014.

[10] Liang Liu, Lili Yu, Dennis K Pearl, and Scott V Edwards. Estimating species phylogenies using coalescence times among sequences. Systematic Biology, 58(5):468–477, 2009.

[11] Liang Liu and Lili Yu. Estimating species trees from unrooted gene trees. Systematic Biology, 60(5):661–667, 2011.

[12] Mahim Mahbub, Zahin Wahab, Rezwana Reaz, M Saifur Rahman, and Md Shamsuzzoha Bayzid. wQFM: highly accurate genome-scale species tree estimation from weighted quartets. Bioinformatics, 06 2021. btab428.

[13] Pranjal Vachaspati and Tandy Warnow. Astrid: accurate species trees from internode distances. BMC Genomics, 16(10):S3, 2015.

[14] Mazharul Islam, Kowshika Sarker, Trisha Das, Rezwana Reaz, and Md Shamsuzzoha Bayzid. Stelar: A statistically consistent coalescent-based species tree estimation method by maximizing triplet consistency. BMC Genomics, 21(1):1–13, 2020.

[15] Dannie Durand, Bjarni V. Halldórsson, and Benjamin Vernot. A hybrid micro–macroevolutionary approach to gene tree reconstruction. Journal of Computational Biology, 13(2):320–335, 2006.

[16] Yi-Chieh Wu, Matthew D. Rasmussen, Mukul S. Bansal, and Manolis Kellis. Treefix: Statistically informed gene tree error correction using species trees. Systematic Biology, 62(1):110–120, 2012.

[17] Matthew D Rasmussen and Manolis Kellis. A bayesian approach for fast and accurate gene tree reconstruction. Molecular Biology and Evolution, 28(1):273–290, 2011.

[18] Sarah Christensen, Erin K. Molloy, Pranjal Vachaspati, and Tandy Warnow. Traction: Fast non-parametric improvement of estimated gene trees. Leibniz International Proceedings in Informatics (LIPIcs), 143(1):4:1–4:16, 2019.

[19] Lafond M Seguin J Boussau B Guéguen L El-Mabrouk N Tannier E. Noutahi E, Semeria M. Efficient gene tree correction guided by genome evolution. PLoS ONE, page e0159559, 2016.

[20] Paweł Górecki and Oliver Eulenstein. A linear time algorithm for error-corrected reconciliation of unrooted gene trees. Bioinformatics Research and Applications, pages 148–159, 2011.

[21] Mukul S Bansal, Yi-Chieh Wu, Eric J Alm, and Manolis Kellis. Improved gene tree error correction in the presence of horizontal gene transfer. Bioinformatics, 31(8):1211–1218, 2015.

[22] Thi Hau Nguyen, Jean-Philippe Doyon, Stéphanie Pointet, Anne-Muriel Arigon Chifolleau, Vincent Ranwez, and Vincent Berry. Accounting for gene tree uncertainties improves gene trees and reconciliation inference. In Algorithms in Bioinformatics: 12th International Workshop, WABI 2012, Ljubljana, Slovenia, September 10-12, 2012. Proceedings 12, pages 123–134. Springer, 2012.

[23] Edwin Jacox, Mathias Weller, Eric Tannier, and Celine Scornavacca. Resolution and reconciliation of non-binary gene trees with transfers, duplications and losses. Bioinformatics, 33(7):980–987, 2017.

[24] Manuel Lafond, Cedric Chauve, Nadia El-Mabrouk, and Aida Ouangraoua. Gene tree construction and correction using supertree and reconciliation. IEEE/ACM transactions on computational biology and bioinformatics, 15(5):1560–1570, 2017.

[25] Zhi Yan, Huw A Ogilvie, and Luay Nakhleh. "correcting" gene trees to be more like species trees frequently increases topological error. Genome Biology and Evolution, 15(6):evad094, 2023.

[26] L Liu. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics, 24:2542–2543, 2008.

[27] J Heled and A J Drummond. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution, 27:570–580, 2010.

[28] B. Boussau, G. J. Szöllősi, L. Duret, M. Gouy, E. Tannier, and V. Daubin. Genome-scale coestimation of species and gene trees. Genome Research, 23(2):323–330, 2013.

[29] Md Shamsuzzoha Bayzid and Tandy Warnow. Naive binning improves phylogenomic analyses. Bioinformatics, 29(18):2277–2284, 2013.

[30] Brian Tilston Smith, Michael G Harvey, Brant C Faircloth, Travis C Glenn, and Robb T Brumfield. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology, 63(1):83–95, 2013.

[31] Md Shamsuzzoha Bayzid and Tandy Warnow. Estimating optimal species trees from incomplete gene trees under deep coalescence. Journal of Computational Biology, 19(6):591–605, 2012.

[32] A D Leaché and B Rannala. The accuracy of species tree estimation under simulation: a comparison of methods. Systematic Biology, 60(2):126–137, 2011.

[33] Gergely J Szöllősi, Eric Tannier, Vincent Daubin, and Bastien Boussau. The inference of gene trees with species trees. Systematic biology, 64(1):e42–e62, 2015.

[34] J H Degnan and N A Rosenberg. Discordance of species trees with their most likely gene trees. PLoS Genetics, 2:762 – 768, 2006.

[35] J H Degnan and N A Rosenberg. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology and Evolution, 26(6), 2009.

[36] James H Degnan. Anomalous unrooted gene trees. Systematic Biology, 62(4):574–590, 2013.

[37] S. Snir and S. Rao. Quartets MaxCut: a divide and conquer quartets algorithm. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 7(4):704–718, 2010.

[38] Rezwana Reaz, Md Shamsuzzoha Bayzid, and M Sohel Rahman. Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. PLoS ONE, 9(8):e104008, 2014.

[39] Sharmin Akter Mim, Md Zarif-Ul-Alam, Rezwana Reaz, Md Shamsuzzoha Bayzid, and Mohammad Saifur Rahman. Quartet fiduccia–mattheyses revisited for larger phylogenetic studies. Bioinformatics, 39(6):btad332, 2023.

[40] Eliran Avni, Reuven Cohen, and Sagi Snir. Weighted quartets phylogenetics. Systematic Biology, 64(2):233–242, 2015.

[41] Sen Song, Liang Liu, Scott V Edwards, and Shaoyuan Wu. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proceedings of the National Academy of Sciences, 109(37):14942–14947, 2012.

[42] Siavash Mirarab, Md Shamsuzzoha Bayzid, Bastien Boussau, and Tandy Warnow. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science, 346(6215):1250463, 2014.

[43] Md Shamsuzzoha Bayzid, Siavash Mirarab, Bastien Boussau, and Tandy Warnow. Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. PLoS ONE, 10(6), 2015.

[44] Erich D Jarvis, Siavash Mirarab, Andre J Aberer, Bo Li, Peter Houde, Cai Li, Simon YW Ho, Brant C Faircloth, Benoit Nabholz, Jason T Howard, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science, 346(6215):1320–1331, 2014.

[45] Chao Zhang, Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab. Astral-iii: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19(6):153, 2018.

[46] D.F. Robinson and L.R. Foulds. Comparison of phylogenetic trees. Mathematical Biosciences, 53:131–147, 1981.

[47] Bent Fuglede and Flemming Topsoe. Jensen-shannon divergence and hilbert space embedding. In International symposium onInformation theory, 2004. ISIT 2004. Proceedings., page 31. IEEE, 2004.

[48] Sazan Mahbub, Shashata Sawmya, Arpita Saha, Rezwana Reaz, M Sohel Rahman, and Md Shamsuzzoha Bayzid. Qt-gild: Quartet based gene tree imputation using deep learning improves phylogenomic analyses despite missing data. In International Conference on Research in Computational Molecular Biology, pages 159–176. Springer, 2022.

[49] Ishrat Tanzila Farah, Muktadirul Islam, Kazi Tasnim Zinat, Atif Hasan Rahman, and Shamsuzzoha Bayzid. Species Tree Estimation from Gene Trees by Minimizing Deep Coalescence and Maximizing Quartet Consistency: A Comparative Study and the Presence of Pseudo Species Tree Terraces. Systematic Biology, 70(6):1213–1231, 04 2021.

[50] Ishrat Tanzila Farah, Md Muktadirul Islam, Kazi Tasnim Zinat, Atif Hasan Rahman, and Md Shamsuzzoha Bayzid. Phylogenomic terraces: presence and implication in species tree estimation from gene trees. bioRxiv, 2020.

[51] Navid Bin Hasan, Avijit Biswas, Zahin Wahab, Mahim Mahbub, Rezwana Reaz, and Md Shamsuzzoha Bayzid. Leveraging weighted quartet distributions for enhanced species tree inference from genome-wide data. bioRxiv, pages 2024–09, 2024.

[52] Erfan Sayyari and Siavash Mirarab. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution, 33(7):1654–1668, 2016.

Appendix A Identifying Quartet Conflicts

Theorem 1. *Let q_1 and q_2 be two quartets drawn from a set of taxa. The number of binary trees on $L(q_1) \cup L(q_2)$ that simultaneously satisfy both q_1 and q_2 is determined by the number of taxa shared between them. Specifically, let ct denote the number of taxa common to both quartets. Then the possible values of ct are 0, 1, 2, 3, or 4, and the number of binary trees that satisfy both q_1 and q_2 is as follows:*

Case 1: *$ct = 0, 1, 2$ When the number of common taxa between q_1 and q_2 is less than 3, i.e., $ct = 0, 1, 2$, there will always be multiple binary trees on $L(q_1) \cup L(q_2)$ that satisfy both quartets.*

Case 2: *$ct = 3$ When q_1 and q_2 share exactly three common taxa ($ct = 3$), meaning each quartet has one unique taxon relative to the other, two sub-cases arise:*

1. *If the unique taxon in q_1 replaces the unique taxon in q_2 (meaning that when the unique taxon in quartet q_1 is replaced by the unique taxon in the other quartet q_2, the resulting topology will match that of q_2), there are exactly three distinct binary trees on $L(q_1) \cup L(q_2)$ that satisfy both quartets q_1 and q_2.*

2. *In all other configurations where the quartets share 3 common taxa, there is exactly one binary tree on $L(q_1) \cup L(q_2)$ that satisfies both quartets.*

Case 3: *$ct = 4$ When q_1 and q_2 share all 4 taxa ($ct = 4$), no binary tree can satisfy both quartets.*

Proof. We examine each case based on the number of common taxa between q_1 and q_2.

- **Case 1:** $ct = 0, 1, 2$ When q_1 and q_2 share fewer than 3 taxa, there are always multiple binary trees that satisfy both quartets.

  Their sets of taxa are mostly disjoint. In this situation, there are multiple binary trees that can satisfy both quartets because the taxa in q_1 and q_2 do not impose sufficient constraints on the tree structure to uniquely define a single tree. For example, $ct = 0$ means q_1 and q_2 do not have any common taxa and thus they can be placed independently within a tree, leading to multiple distinct trees that satisfy both quartets.

  - Sub-case $ct = 0$ Since there are no shared taxa, the quartets are completely independent, and the quartets can be placed independently within a tree on $L(q_1) \cap L(q_2)$, leading to multiple distinct trees that satisfy both quartets.
  - Sub-case $ct = 1$ Suppose $q_1 = ab|cd$ and $q_2 = ax|yz$ be two quartets with a as the common taxon. Here, {a,b} and {a,x} are separated from {c,d} and {y,z} in q_1 and q_2, respectively. Then it is easy to see that any binary tree on $L(ab|cd) \cup L(ax|yz) = \{a, b, c, d, x, y, z\}$ containing an edge e that induces a bipartition $\pi(e) = abx|cdyz$ separating $\{a, b\} \cup \{a, x\} = \{a, b, x\}$ from $\{c, d\} \cup \{y, z\} = \{c, d, y, z\}$ is consistent with both $ab|cd$ and $ax|yz$. Since there are 3 possible ways to arrange $\{a, b, x\}$ at one partition of e and 15 possible ways to arrange $\{c, d, y, z\}$ in the other partition, there are at least $3 \times 15 = 45$ trees that satisfy both quartets.
  - Sub-case $ct = 2$ let the two common taxa in q_1 and q_2 be a and b, and the other four taxa (two from each quartet) be c, d, x, and y, respectively. Now, depending on the relative placements of a and b in q_1 and q_2, three cases may arise.
    - Sub-sub-case 1 (a and b are sister taxa in both quartets): Let q_1 and q_2 be $ab|cd$ and $ab|xy$ respectively. It is easy to see that any binary tree drawn on $L(q_1) \cup L(q_2)$ containing the bipartition $ab|cdxy$ satisfies both quartets.
    - Sub-sub-case 2 (a and b are sisters on exactly one quartet): Without the loss of generality, let q_1 and q_2 be $ab|cd$ and $ax|by$, a and b being sisters in q_1. Then, any binary tree that contains both the bipartitions $ax|bcdy$ and $abx|cdy$ satisfies $ab|cd$ and $ax|by$. Similarly, any binary tree that contains both the bipartitions $by|acdx$ and $aby|cdx$ satisfies $ab|cd$ and $ax|by$.
    - Sub-sub-case 3 (a and b are not sisters on any quartet): Let q_1 and q_2 be $ac|bd$ and $ax|by$ respectively. Any binary tree on $L(q_1) \cup L(q_2)$ containing an edge e that induces a bipartition $\pi(e) = acx|bdy$ separating $\{a, c\} \cup \{a, x\} = \{a, c, x\}$ from $\{b, d\} \cup \{b, y\} = \{b, d, y\}$ is consistent with both $ac|bd$ and $ax|by$. Note that there are 9 such trees.

Case 2: $ct = 3$ **(three common taxa)** When q_1 and q_2 share exactly three taxa, the number of trees that can satisfy both quartets depends on the relationship between the unique taxon in each quartet. Two distinct sub-cases arise:

- Sub-case 1 (replacement of unique taxa): If the unique taxon in q_1 simply replaces the unique taxon in q_2 (preserving the relationship between the shared taxa) and vice versa, then two of the common taxa will be sister species in both q_1 and q_2. In this case, exactly three distinct binary trees on the set $L(q_1) \cup L(q_2)$ of five leaves satisfy both quartets. These three trees correspond to the possible ways to arrange the unique taxa from q_1 and q_2 with respect to the shared taxa. For example, $q_1 : ab|cd$ and $q_2 : ab|ce$ satisfy this condition where d and e are the unique taxa in q_1 and q_2 respectively and replacing d with e in q_1 (or e with d in q_2) makes q_1 and q_2 identical. Moreover, a and b are closer to each other than they are to other species in both q_1 and q_2. In this case, we can place e on three branches in q_1 (as shown in Figure A1a), resulting in three different trees on five taxa $\{a, b, c, d, e\}$ that satisfy both of them. Similarly, we can place d on three branches in q_2 (as shown in Figure A1b), producing three different trees that satisfy both q_1 and q_2. Note that the three trees in Figure A1a are identical to the trees depicted in Figure A1b. For instance, placing taxon e on the middle branch in Figure A1a yields the same tree as placing taxon d as the sister to taxon c in Figure A1b.

  Let us now consider the two additional cases where taxon e is placed on the other two branches of the quartet $ab|cd$. First, if taxon e is placed as the sister taxon to a, one of the resulting induced quartets is $ae|bc$, which does not satisfy $ab|ce$ as e is now the sister to a. Similarly, if e is placed as the sister to b, the quartet $ac|be$ is formed, which also is not consistent with $ab|ce$. Thus, there are exactly three trees on $\{a, b, c, d, e\}$ that satisfy both q_1 and q_2.

(a) Placement of e over the quartet $ab|cd$ **(b)** Placement of d over the quartet $ab|ce$

Figure A1: Comparison of placements for e and d

- Sub-case 2 (other configurations): In all other scenarios where the unique taxon in q_1 does not simply replace the unique taxon in q_2 (and vice versa), there is exactly one binary tree on the set $L(q_1) \cup L(q_2)$ of five taxa that satisfies both quartets. This is because the three common taxa, along with their relationships to the unique taxa, fully constrain the tree structure. Here, unlike sub-case 1, no pair of taxa among the three common taxa are sisters in both quartets. Consider the two quartets, $q_1 = ab|cd$ and $q_2 = ae|bd$ that satisfy this condition. In this context, unique taxon c in q_1 and e in q_2 do not replace each other. In order to find a tree that satisfies both q_1 and q_2, we can insert e in q_1 by making it a sister of a, as shown in Figure A2a. Similarly, we can insert c into q_2 by making it the sister of d as shown in Figure A2b. Note that these two trees are identical. Furthermore, the placements of e on four other edges in q_1 (the edges incident on b, c, d, and the internal edge) result in trees that do not support q_2 ($ae|bd$) as b becomes closer to a than e is to a. Similarly, the placements of c on four other edges in q_2 do not produce any trees that satisfy q_1 ($ab|cd$) as b becomes closer than c to d. Thus, there is exactly one tree (as shown in Figure A3a) that satisfies both q_1 and q_2.

Case 3: $ct = 4$ **(all taxa are shared)** When q_1 and q_2 share all four taxa, the two quartets represent topologically distinct arrangements of the same taxa – making them topologically distinct variants of each other. Since a single binary tree cannot accommodate two conflicting topologies for the same set of taxa, no tree can simultaneously satisfy both quartets.

□

(a) Placement of e over the quartet $ab|cd$ **(b)** Placement of c over the quartet $ae|bd$

Figure A2: Unrooted trees satisfying both quartets $ab|cd$ and $ae|bd$

(a) Only unrooted tree satisfying $ab|cd$ and $ae|bd$ **(b)** Only unrooted tree satisfying $ab|cd$ and $ac|be$

Figure A3: Unrooted trees satisfying two different canonical pairs

We define *canonical quartet pair* as a pair of quartets (q_1, q_2) that can be satisfied by exactly one tree T, where $L(T) = L(q_1) \cup L(q_2)$, as in Sub-case 2 of Case 2 in Theorem 1. Thus, a canonical quartet pair (q_1, q_2) uniquely represent a tree on $L(q_1) \cup L(q_2)$ that satisfy both q_1 and q_2. For example, $q_1 = ab|cd$ and $q_2 = ae|bd$ is one such canonical quartet pair, which is satisfied by exactly one tree T on $\{a, b, c, d, e\}$ as shown in Figure A3a. There are $\binom{5}{4} = 5$ quartets ($ae|bd$, $ae|bc$, $ae|cd$, $ab|cd$, and $be|cd$) in T including the canonical pair corresponding to this tree ($ab|cd$ and $ae|bd$). The canonical pair of quartets will clearly be in conflict with the topological variants of the three other quartets ($ae|bc$, $ae|cd$, and $be|cd$). Note that each quartet has two topologically distinct variants. Thus, we have the following Corollary 1.1.

Corollary 1.1. *Every canonical pair of quartets (q_1, q_2) is in conflict with six quartets on $L(q_1) \cup L(q_2)$.*

Note that the tree in Figure A3a is constructed using $q_1 = ab|cd$ as the backbone, with the unique taxon e in $q_2 = ae|bd$ placed as a sister to a. Consequently, both $q_2 = ae|bd$ and $ae|bc$ are induced by T. Therefore, if we instead consider $q_2 = ae|bc$ (rather than $ae|bd$), $q_1 = ab|cd$ and $q_2 = ae|bc$ forms another canonical pair, which is also satisfied by the same tree in Figure A3a. Thus, two canonical pairs, $(ab|cd, ae|bd)$ and $(ab|cd, ae|bc)$, that contain $ab|cd$ as one member and the other member having e as the unique taxon with respect to $ab|cd$ correspond to the tree shown in Figure A3a. Additionally, two other canonical pairs, $(ae|bc, be|cd)$ and $(ae|bd, be|cd)$, both sharing $be|cd$ as a common member and excluding $ab|cd$, also represent the same tree in Figure A3a. Consequently, the tree shown in Figure A3a contains a total of four canonical pairs of quartets (q_1, q_2), (q_2, q_3), (q_3, q_4) and (q_4, q_1), where $q_1 = ab|cd, q_2 = ae|bd, q_3 = be|cd, q_4 = ae|bc$. Notably, the quartet $ae|cd$, induced by the tree in Figure A3a, does not appear in any canonical pair, as there is no other quartet that can pair with $ae|cd$ to place taxon b on the internal branch of $ae|cd$. This is because, when two quartets form a canonical pair, the unique taxon from one quartet is positioned as a sibling of one of the taxa in the other quartet. All other quartets in the tree, except $ae|cd$, are part of two canonical pairs.

Additionally, using $q_1 = ab|cd$ as the backbone and placing e as the sister to b, we obtain a tree T (shown in Figure A3b) that satisfies two other canonical quartet pairs, $(ab|cd, ac|be)$ and $(ab|cd, ad|be)$, that contain $ab|cd$ as one member. Similarly, e can be placed as a sister to the other taxa in q_1 (i.e, c and d). Thus, for a given quartet $q_1 = ab|cd$, we can form eight canonical pairs of quartets with $q_1 = ab|cd$ being one member and the other member q_2 having e as the unique taxon to q_1. These lead to the following Corollaries 1.2 and 1.3 (these are Corollaries 1.2 and 1.3 in the main text respecetively).

Corollary 1.2. *Every binary unrooted tree T with five taxa contains a total of four canonical pairs of quartets (q_1, q_2), (q_2, q_3), (q_3, q_4) and (q_4, q_1) with each of the four quartets q_1, q_2, q_3, q_4 being present in two canonical pairs.*

Corollary 1.3. *Every quartet is a part of eight canonical pairs of quartets with the other members of these pairs having a particular unique taxon with respect to q_1.*

As stated in Corollary 1.1, a quartet $q_1 = ab|cd$ is in conflict with six other quartets having a unique taxon with respect to q_1. Additionally, according to Corollary 1.3, q_1 is a member of eight pairs of canonical quartets. Table A1 lists the six conflicting quartets for each of these eight canonical pairs of quartets, where $q_1 = ab|cd$ is one member and the other member contains a unique taxon e. Thus, every canonical pair and its six conflicting quartets form a set of three quartets that cannot coexist in a tree, resulting in 48 such sets. We call these conflicting trios of quartets as a *conflicting set*. Ignoring the duplicates, as marked in Table A1, 28 unique conflicting sets remain.

Canonical Pair	**Conflicting Quartets**
$ab\|cd$ & $ae\|bd$	$ab\|ce$ || $ac\|be$ || $ac\|de$ || $ad\|ce$ || $bc\|de$ || $bd\|ce$
$ab\|cd$ & $ae\|bc$	$ab\|de$ || $ad\|be$ || $ac\|de$ || $ad\|ce$ || $bc\|de$ || $bd\|ce$
$ab\|cd$ & $ac\|be$	$ab\|de$ || $ae\|bd$ || $ac\|de$ || $ad\|ce$ || $bc\|de$ || $bd\|ce$
$ab\|cd$ & $ad\|be$	$ab\|ce$ || $ae\|bc$ || $ac\|de$ || $ad\|ce$ || $bc\|de$ || $bd\|ce$
$ab\|cd$ & $ad\|ce$	$ac\|be$ || $ae\|bc$ || $ad\|be$ || $ae\|bd$ || $bc\|de$ || $be\|cd$
$ab\|cd$ & $bd\|ce$	$ac\|be$ || $ae\|bc$ || $ad\|be$ || $ae\|bd$ || $ac\|de$ || $ae\|cd$
$ab\|cd$ & $ac\|de$	$ac\|be$ || $ae\|bc$ || $ad\|be$ || $ae\|bd$ || $bd\|ce$ || $be\|cd$
$ab\|cd$ & $bc\|de$	$ac\|be$ || $ae\|bc$ || $ad\|be$ || $ae\|bd$ || $ad\|ce$ || $ae\|cd$

Table A1: List of conflicting quartets for eight canonical quartet pairs containing $ab|cd$. The items marked in blue are part of sets that were already encountered in a previous row. For example, the second quartet on the third row, $ae|bd$, constitutes the conflicting set $\{ab|cd, ac|be, ae|bd\}$. This set has already been encountered before. In the first row, by considering the second quartet $ac|be$, we obtain the conflicting set $\{ab|cd, ae|bd, ac|be\}$, which is identical to the set in consideration.

Corollary 1.4. *Every quartet q is part of 28 unique conflicting sets as a member of canonical pairs for each unique taxon $x \notin L(q)$.*

For a given set S of n taxa, and a quartet q ($L(q) \in S$), there are $n-4$ unique taxa in S with respect to q. Thus, the following corollary follows naturally.

Corollary 1.5. *For a set S of n taxa, every quartet q, where $L(q) \in S$, is in $28 \times (n-4)$ unique conflicting sets as a member of canonical pairs.*

Appendix B QT-WEAVER: pseudo-code

Algorithm 1: QT-WEAVER

```
Input  : QD_E, set of weighted quartets
         S, set of all taxa
Output: QD'_E, set of weighted quartets with adjusted weights
/* A is the set of adjusted quartets                                  */
A ← ∅;
foreach q ∈ QD_E do
    if q ∈ A then // already adjusted
        continue
    end
    /* Generate all 3 quartets on the leaves of q                     */
    T ← GenerateQuartets(L(q))
    s ← 0 // sum of actual weights of quartets in T
    s' ← 0 // sum of adjusted weights of quartets in T
    foreach q_i ∈ T do
        C_i ← 0 // Conflict score of q_i
        s ← s + W_i
        Z_i ← ∅ ;
        foreach t ∈ S \ L(q_i) do
            /* For each external taxa, we generate 28 conflict pairs  */
            Z_i ← Z_i ∪ GenerateConflicts(q_i, t)
        end
        foreach (q_j, q_k) ∈ Z_i do
            if heuristic is PoW then // Using Product of Weights definition
                C_i ← C_i + W_j · W_k
            end
            else if heuristic is MoW then // Using Minimum of Weights definition
                C_i ← C_i + min(W_j, W_k)
            end
        end
        W'_i ← W_i / C_i
        s' ← s' + W'_i
    end
    /* Normalize the weights                                          */
    foreach q_i ∈ T do
        W'_i ← (s / s') · W'_i
    end
    A ← A ∪ T
end
```

Appendix C QT-WEAVER configurations

The comparison of the eight configurations of QT-WEAVER (the two weighting schemes combined with four different conflicting subsets) along with wQFM on the original distribution has been presented in Figure A4. The original distribution is derived by taking the frequency of each quartet in all gene trees (GTF: Gene Tree Frequency). We vary both the number of genes (100 and 1000) and the amount of gene tree estimation errors (by varying the sequence lengths from 100bp to 1000bp). As expected, the performance of the methods improves when the number of genes increases or when the gene tree estimation errors decrease. The MoW-based weighting scheme outperforms the PoW variant. Furthermore, subsets of four or six conflicting sets tend to perform better than all conflicting sets.

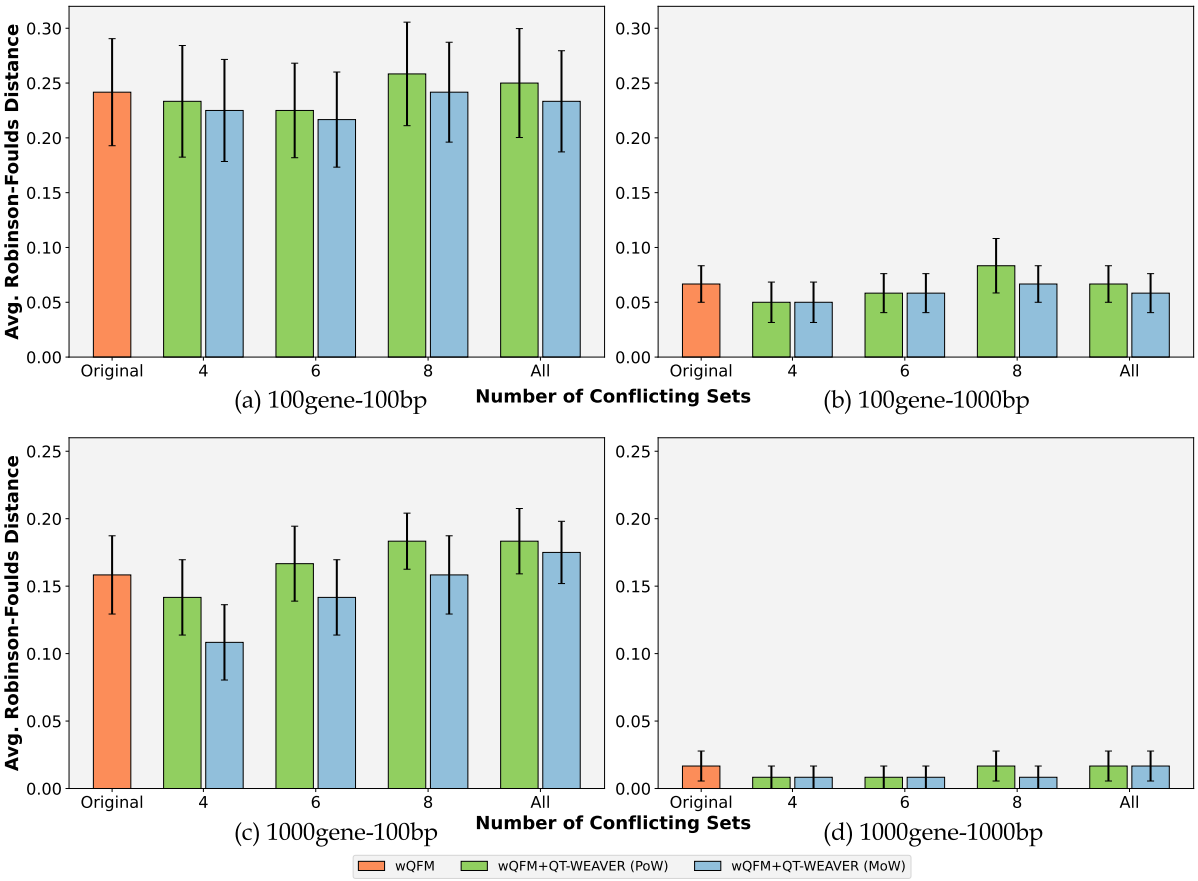


Figure A4: Results on 15-taxon dataset. Comparison of performance after correcting the weighted distributions using four, six, eight, and all 28 conflicting sets. We report the RF rates averaged over 10 replicates with standard errors. The distributions were amalgamated using wQFM. We show the effects of using both PoW and MoW weighting schemes.

These results raise the question – why does considering a small subset of all conflicting sets tend to yield better results? We hypothesize that using all 28 conflicting sets may impose excessive topological constraints on a quartet, making it challenging for QT-WEAVER to effectively distinguish among different topological variants.

In contrast, the six selected conflicting sets (as listed in Table A2) appear suitable for assessing conflict levels across quartet topologies, as indicated by our experimental results. While this specific choice of six sets lacks theoretical backing, we anticipate that the optimal subset–both in size and composition–may vary by dataset. Thus, an important research avenue involves automating the selection of these subsets, tailored to the topological features of the input gene trees, to optimize performance.

Canonical Pair	Conflicting Quartets
$ab\|cd$ & $ac\|be$	$ab\|de$ || $ae\|bd$
$ab\|cd$ & $ae\|bc$	$ad\|be$
$ab\|cd$ & $ac\|de$	$bd\|ce$ || $be\|cd$
$ab\|cd$ & $ad\|ce$	$bc\|de$

Table A2: We list the six conflicting sets chosen from all the conflicting sets from Table A1.

Appendix D Dataset

Table A3: Properties of the simulated datasets. The level of ILS is presented in terms of the average topological distance between true gene trees and true species trees.

Dataset	ILS level	# genes	# sites	# Replicates	Ref.
15-taxon	82%	100 - 1000	100 - 1000	10	[43]
37-taxon	18%, 32%, 54%	100 - 500	50 -1000	20	[42]

Appendix E Additional Results

E.1 Additional results on 37-taxon dataset

Table A4: The Jensen-Shannon divergence with respect to the true gene trees for different model conditions averaged over all replicates. The best-performing method (i.e., the least amount of divergence) for each model condition has been highlighted in bold.

Model Condition	**GTF**	**QT-WEAVER**
0.5X-200-500	**0.038**	0.044
1X-100-500	0.072	**0.067**
1X-200-25	0.354	**0.345**
1X-200-50	0.3	**0.287**
1X-200-250	0.118	**0.097**
1X-200-500	0.063	**0.058**
1X-200-1000	**0.034**	0.053
1X-500-500	0.057	**0.052**
2X-200-500	0.097	**0.082**

Table A5: Average quartet scores of estimated species trees and the true tree with respect to the estimated gene trees for the 37-taxon dataset. For wQFM+QT-WEAVER, we also report the quartet score with respect to the corrected estimated quartet distribution. The highest scores among the estimated trees have been italicized, and the scores closest to the true tree are shown in bold.

Model Condition	**ASTRAL**	**wQFM**	**wQFM+QT-WEAVER (Original Distribution)**	**wQFM+QT-WEAVER (Corrected Distribution)**	**True Tree**
0.5X-200-500	10007426	10006594	**10006474**	*10415670*	10003930
1X-100-500	5640132	5639605	**5639168**	*5806039*	5636884
1X-200-25	6491920	6489889	**6489155**	*6662792*	6481431
1X-200-50	7525465	**7523379**	7523903	*7778394*	7517956
1X-200-250	10569844	10568958	**10568350**	*10941339*	10565292
1X-200-500	11271938	11271426	**11270680**	*11603816*	11267991
1X-200-1000	11586641	11586355	**11585945**	*11886743*	11584970
1X-500-500	28171698	**28171682**	28168879	*28983587*	28171386
2X-200-500	11947266	**11947020**	11945419	*12175146*	11946371

Table A6: Average quartet scores of estimated species trees and the true tree with respect to the true gene trees for the 37-taxon dataset. The highest scores among the estimated trees have been italicized, and the scores closest to the true tree are shown in bold.

Model Condition	ASTRAL	wQFM	wQFM+QT-WEAVER	True Tree
0.5X-200-500	***10311630***	10311452	10311480	10311622
1X-100-500	5870845	5871208	***5871324***	5874698
1X-200-25	***11637683***	11613065	11627649	11744079
1X-200-50	11688779	***11707476***	11703316	11744079
1X-200-250	11733853	***11735693***	11731082	11744079
1X-200-500	11738631	11739572	***11741356***	11744079
1X-200-1000	11743809	11743721	***11744134***	11744079
1X-500-500	29362939	***29363114***	29356078	29364014
2X-200-500	12565282	***12565347***	12563879	12570343

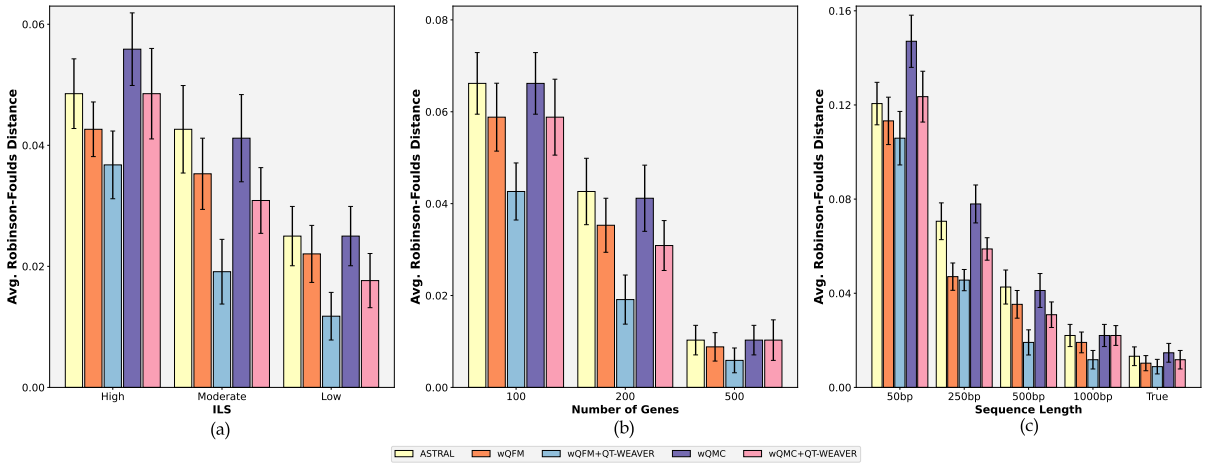


Figure A5: Results on 37-taxon dataset. We show the average RF rates with standard errors over 20 replicates for the methods ASTRAL, wQFM (wQFM-uncorrected), wQFM+QT-WEAVER (wQFM-corrected), wQMC (wQMC-uncorrected), and wQMC+QT-WEAVER (wQMC-corrected). The settings in (a)-(c) are identical to the ones in Figure 4.

E.2 Additional results on 15-taxon dataset

Table A7: The Jensen-Shannon divergence with respect to the true gene trees for different model conditions averaged over all replicates. The best-performing method (i.e., the least amount of divergence) for each model condition has been highlighted in bold.

Model Condition	GTF	QT-WEAVER
100gene-100bp	0.115	**0.103**
100gene-1000bp	**0.043**	0.049
1000gene-100bp	0.101	**0.087**
1000gene-1000bp	**0.031**	0.037

Table A8: Average quartet scores of estimated species trees and the true tree with respect to the estimated gene trees for the 15-taxon dataset. For wQFM+QT-WEAVER, we also report the quartet score with respect to the corrected estimated quartet distribution. The highest scores among the estimated trees have been italicized, and the scores closest to the true tree are shown in bold.

Model Condition	**ASTRAL**	**wQFM**	**wQFM+QT-WEAVER (Original Distribution)**	**wQFM+QT-WEAVER (Corrected Distribution)**	**True Tree**
100gene-100bp	69934	**69777**	69792	*73598*	69307
100gene-1000bp	82167	82129	**82125**	*87562*	82099
1000gene-100bp	693656	692949	**692664**	*727712*	690268
1000gene-1000bp	818022	817994	**817953**	*869904*	817937

Table A9: Average quartet scores of estimated species trees and the true tree with respect to the true gene trees for the 15-taxon dataset. The highest scores among the estimated trees have been italicized, and the scores closest to the true tree are shown in bold.

Model Condition	**ASTRAL**	**wQFM**	**wQFM+QT-WEAVER**	**True Tree**
100gene-100bp	82159	82708	***83078***	84635
100gene-1000bp	84345	84437	***84444***	84635
1000gene-100bp	827970	834890	***836349***	844184
1000gene-1000bp	843363	843792	***843973***	844184

E.3 Additional results on iterative corrections using QT-WEAVER

We investigated how the Jensen-Shannon divergence evolves across iterations (see Figure A6(a) in Appendix E.3). Initially, the divergence decreases during the first one or two iterations but then begins to increase. This occurs because, for the three alternative quartet topologies ($ab|cd, ac|bd, ad|bc$) on four taxa, QT-WEAVER tends to prioritize the quartet topology it identifies as "correct" –the one with the least conflict score–by increasing its weight while reducing the weights of the alternative topologies. Thus, with successive iterations, the weight of the "correct" quartet topology continues to rise (toward 100%) while the weights of the other two alternatives keep decreasing (toward 0%). Thus, this overestimation of the weight of the "correct" quartets and the underestimation of the weights of the "incorrect" ones lead the adjusted distribution to diverge from the true weighted distribution. Despite this divergence, the iterative adjustments may still guide the tree search algorithm toward more accurate trees (up to a certain point) by emphasizing and amalgamating the quartets with higher weights, as evidenced by the gradual decrease in the RF rates up to 10-15 iterations. For the same reason, when examining the quartet scores of wQFM+QT-WEAVER with respect to the corresponding corrected quartet distribution (instead of the original quartet distribution), the scores continue to increase, approaching nearly 100% (see Figure A6(b) in Appendix E.3). This occurs because, in the corrected distribution, the weights of the quartets present in the estimated species trees steadily rise, resulting in higher quartet scores for the estimated trees.

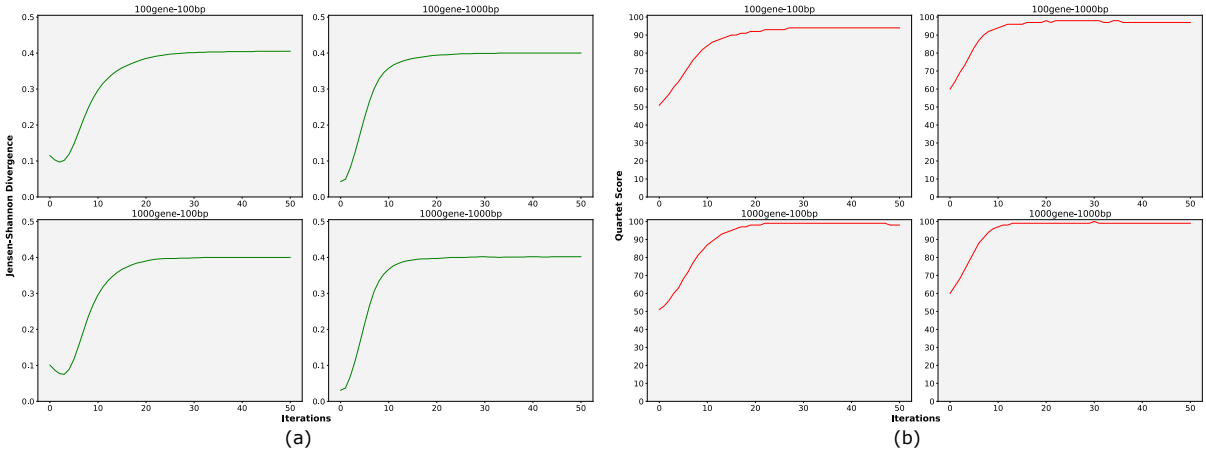
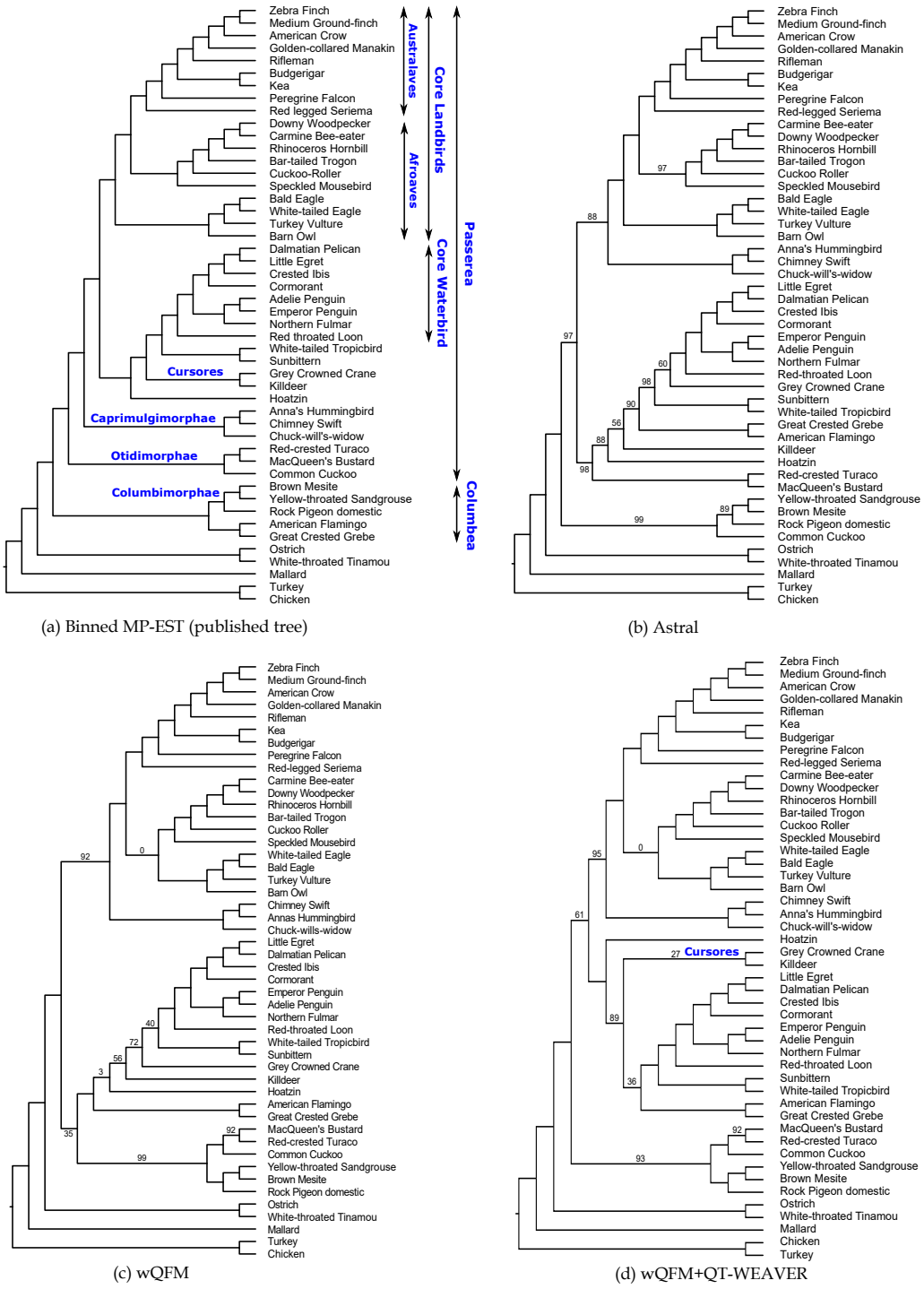


Figure A6: Results iterative corrections on 15-taxon dataset. We show the changes in (a) Jensen-Shannon divergence with respect to the true quartet distribution and (b) the quartet score of the estimated species tree with respect to its corresponding corrected quartet distribution over 50 iterations.

Appendix F Analysis of Avian Biological Dataset



(a) Binned MP-EST (published tree)

(b) Astral

(c) wQFM

(d) wQFM+QT-WEAVER

Figure A7: Analyses of the avian dataset using ASTRAL and wQFM (before and after correction). Branch supports are computed based on quartet-based local posterior probability [52] (multiplied by 100). All BS values are 100% except where noted.