Quartet-based species tree methods enable fast and consistent tree of blobs reconstruction under the network multispecies coalescent

Junyan Dai[1,*[0009−0009−5699−2196]], Yunheng Han[1,*[0000−0003−0200−5924]], and Erin K. Molloy[1,2,⋆[0000−0001−5553−3312]]

[1] Department of Computer Science, University of Maryland, College Park, MD 20742, USA
[2] University of Maryland Institute for Advanced Computer Studies, College Park, MD 20742, USA
ekmolloy@umd.edu

Abstract. Gene flow between species or populations is an important force in evolution, modeled by the network multispecies coalescent. Reconstructing evolutionary histories, called species networks, under this model is notoriously challenging, with the leading methods scaling to just tens of species. Divide-and-conquer is a promising path forward; however, methods with statistical consistency guarantees require the *tree of blobs* (TOB), which displays only the tree-like parts of the network, to perform subset decomposition. TOB reconstruction under the NMSC is challenging in its own right, with the only available method TINNiK having time complexity $O(n^5 + n^4k)$, where k is the number of input gene trees and n is the number of species. Here, we present a framework for TOB reconstruction that operates by (1) seeking a refinement of the TOB and then (2) contracting edges in it. For step (1), we show that an optimal solution to Weighted Quartet Consensus is a TOB refinement almost surely, as the number of gene trees increases, motivating the use of fast quartet-based methods for species tree estimation such as ASTRAL or TREE-QMC. For step (2), we contract edges in the refinement tree based on the same hypothesis tests as TINNiK, which are applicable to subsets of four taxa. We show that sampling just $O(n)$ four-taxon subsets around each edge enables statistically consistent TOB estimation, with asymptotic runtime dominated by tree reconstruction. Leveraging TREE-QMC for this step gives our method a time complexity of $O(n^3k)$ and its name TOB-QMC. On simulated data sets, TOB-QMC is at least as accurate and often more accurate than TINNiK. Moreover, TOB-QMC scales to larger data sets and enables fast and interpretable exploration of hyperparameters used in hypothesis testing. We demonstrate the importance of this feature on phylogenomic data sets. Lastly, our framework is related to ad hoc analyses performed by biologists, as network methods do not scale. Our theoretical results provide justification for such approaches as well as context for interpreting species trees estimated with quartet-based methods in the presence of gene flow; this is critical given the recent result that tree-based network inference with ASTRAL can be positively misleading.

TOB-QMC is implemented within TREE-QMC, available on Github: `https://github.com/molloy-lab/TREE-QMC`.

Keywords: Phylogenetic Networks · Multispecies Coalescent · Quartets · Tree of Blobs

⋆ Corresponding author
*J. Dai and Y. Han—These authors contributed equally to this work.

1 Introduction

Gene flow between species or populations is an important force in evolution; however, it complicates the reconstruction of evolutionary histories, requiring them to be represented as phylogenetic networks rather than trees. In this context, phylogenetic networks (called species or population networks) govern ancestry, specifically the genealogies of genomic regions (called gene trees), as modeled by the network multi-species coalescent (NMSC) [34]. Despite significant progress over the last decade, phylogenetic network reconstruction under the NMSC continues to be a major scientific challenge. Perhaps the most statistically rigorous methods developed to date (e.g., [35]) are those that employ likelihood functions accounting for (1) gene trees evolving within a species network under the NMSC [34, 37] and (2) molecular sequences evolving down a gene tree [11], integrating over all possible gene trees. Although the later step can be avoided by estimating gene trees from molecular sequences prior to network reconstruction, the simplified likelihood is still prohibitively expensive, motivating the development of pseudo-likelihood methods [36, 32]. However, even the most popular pseudo-likelihood method SNaQ [32] is limited to around 30 species [20].

Given scalability limitations, researchers often rely on tree-based network reconstruction [8], which adds edges representing gene flow to a fixed tree, typically estimated with the popular species tree method ASTRAL [22] (e.g. Fig. 2 in [9]). ASTRAL is a heuristic for the NP-hard Weighted Quartet Consensus (WQC) problem [21], which can be framed as weighting quartets (i.e., unrooted trees on four species) according to their frequencies in the input gene trees and then seeking a species tree to maximize the total weight of the quartets it displays. Unfortunately, Dinh and Baños (2025) recently showed that tree-based network reconstruction with ASTRAL can be positively misleading because an optimal solution to WQC is not guaranteed to coverage in probability to a tree displayed by the network, as the number of (error-free) gene trees tends to infinity [10]. This negative result is important given the numerous prior studies using ASTRAL in the presence of gene flow and suggests scalable phylogenetic network methods are urgently needed.

A promising path forward is divide-and-conquer [16, 38, 20]. As an example, the recently developed InPhyNet method [20] partitions species into subsets, estimates a network on each subset with SNaQ, and then combines the subset networks together, similar to disjoint tree mergers [24, 25]. Of the existing divide-and-conquer methods only InPhyNet [20] claims statistical consistency; however, the proof assumes that species are partitioned into subsets according to the tree of blobs (TOB) [13], which displays only the tree-like parts of the network. To our knowledge, the first and only available method for TOB reconstruction under the NMSC is TINNiK [2]. TINNiK is statistically consistent and has time complexity $O(n^4k + n^5)$, where k is the number of input gene trees and n is the number of species. The first term comes from computing quartet weights, which are needed to perform hypothesis tests on all subsets of four species, and the second term comes from updating the results to ensure any remaining signals of network-like evolution are removed. The final results are encoded as a distance matrix so that applying Neighbor Joining [30] and contracting zero-length edges yields the tree of blobs. However, TINNiK cannot scale to large numbers of species, as its asymptotic runtime would suggest, so InPhyNet implements an alternative approach to subset decomposition on large data sets, without statistical guarantees.

To address this issue, we introduce a new framework for TOB reconstruction that operates by (1) seeking a refinement of the TOB and then (2) contracting edges in it. For step (1), we show that an optimal solution to WQC is a TOB refinement almost surely, as the number of gene trees increases, motivating the use of fast quartet-based methods for species tree estimation such as ASTRAL or TREE-QMC [14]. For step (2), we show that applying the same hypothesis test as TINNiK [1] to just $O(n)$ four-taxon subsets around each edge is sufficient for statistically consistent TOB reconstruction. We implement our framework within the quartet-based method TREE-QMC, giving our method a time complexity of $O(n^3k)$ as well as its name TOB-QMC. On simulated data sets, TOB-QMC is at least as accurate and often more accurate than TINNiK. Moreover, TOB-QMC scales to larger data sets and enables fast and interpretable exploration of hyperparameters used in hypothesis testing. We demonstrate the importance of this feature on phylogenomic data sets for seed plants, the bee subfamily *Nomiinae*, and the butterfly subfamily *Heliconiinae*.

Overall, our study advances theory and practice for scalable evolutionary analyses in the presence of gene flow. On the theoretical side, our results explain why ASTRAL, and related quartet-based methods, have great utility in the presence of gene flow, provided the results are interpreted properly, as TOBs, rather than network display trees; this context is critical given ASTRAL's widespread use and the negative result from Dinh and Baños [10] (also see [23]). On the empirical side, our framework enables researchers to reconstruct

TOBs on large data sets and evaluate the impact of network-like evolution on their analyses more broadly. Long term, we expect TOB-QMC to have applications in phylogenetic network reconstruction.

2 Methods

To present TOB-QMC, we briefly review key background and terminology; see the Supplement for details.

2.1 Background and Terminology

Phylogenetic networks. A *directed* phylogenetic network N^+ is a simple directed acyclic graph with the property that there exists a directed path from the root vertex to all other vertices. Leaf vertices are bijectively labeled by a set S of species. Non-leaf (i.e., internal) vertices are either speciation events or reticulation events. Speciation nodes are vertices with in-degree 1 and out-degree > 1; *reticulation nodes* are vertices with in-degree > 1 and out-degree 1. A tree edge is an arc whose child is a tree node; all other arcs are called *reticulation edges*. Henceforth, networks are assumed to be *binary*, meaning all internal vertices have degree 3, except the root, which has degree 2. A phylogenetic tree is simply a phylogenetic network with no reticulation vertices. We say that a tree is **displayed** by a directed network if it can be obtained from the network by deleting only incoming edges to reticulation nodes but not their end points (Fig. 1A,D). The **class** of the network is the number of topologically distinct display trees it induces [3].

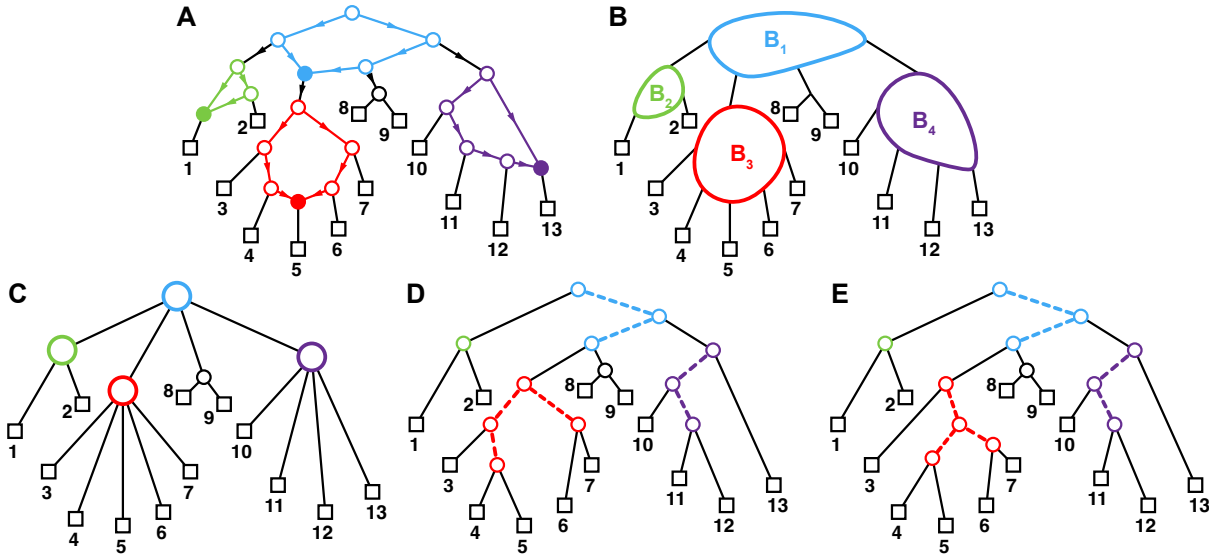

Fig. 1: **A)** A level-1 directed phylogenetic network N^+ with reticulation vertices filled in different colors. **B)** The unrooted version of N^+ with blob edges colored based on the single reticulation node in the blob. **C)** Tree of blobs for N^+. **D)** A TOB refinement that is induced by a display tree of N^+. **E)** A TOB refinement that is NOT induced by any display trees of N^+. Dashed edges are contracted to obtain the TOB.

A directed network N^+ can be converted into an *undirected* network N by suppressing the root and undirecting its edges. A bridge is an edge whose deletion increases the number of connected components of the graph. The maximal bridgeless subgraphs of an undirected network are called **blobs**. The degree of a blob is given by the number of bridges adjacent to it (an m-blob is a blob with degree m). Edges in blobs are **blob edges**, and all other edges (i.e., bridges) are **cut edges**. The **tree of blobs (TOB)** [13] can be computed from a directed network obtained by (1) un-directing the network, (2) contracting all blob edges, and then (3) suppressing all vertices with degree two (Fig. 1A–C). The contraction step results in each blob being collapsed into a single vertex, which may have degree greater than 3 (referred to as a polytomy). We say that a tree T refines the TOB if T can be obtained from the TOB by contracting edges. It is easy to see that any tree displayed by a network is a refinement of the TOB (Fig. 1D); however, a refinement of the TOB need not be a display tree of the network (Fig. 1E).

The deletion of a vertex in a phylogenetic tree produces an m**-partition** of its species set S, denoted $M = P^1|P^2|...|P^m$; the same holds for an m-blob in a phylogenetic network. Likewise, the deletion of edge

e in a tree (or a cut edge e in a network) induces a **bipartition**, denoted $Bip(e) = A|B$. If the tree is binary, edge $e = (s,t)$ also induces **quadrapartition** and denoted $Quad(e) = A_1|A_2|B_2|B_2$, formed by deleting e along with its incident vertices s and t. Later we consider sampling 4-taxon subsets **around an edge**, meaning that two leaves are drawn from each block of its bipartition (sometimes this is restricted to drawing one leaf from each block of a quadrapartition). Lastly, recall that a **quartet** q is an unrooted tree four species. There are three possible quartet topologies denoted by their non-trivial bipartition: $a,b|c,d$ and $a,c|b,d$ and $a,d|b,c$, where $S = \{a,b,c,d\}$.

Network Multispecies Coalescent (NMSC). The NMSC is parameterized by a species network $\mathcal{N} = (N, \Theta)$, where N is a phylogenetic network and Θ is a set of numerical parameters (branch lengths and inheritance proportions). A gene genealogy can be generated under the NMSC by sampling one gene per species and tracing their ancestry backward in time until all gene lineages coalesce (i.e., trace back to a common ancestor). The network topology N governs the pool of potential ancestors. We let $\mathbb{P}_{\mathcal{N}}(T)$ denote the probability of a gene genealogy T under the NMSC given $\mathcal{N}$. Although gene trees are not observed, they can be estimated from molecular sequence data. Our goal is to reconstruct the TOB of a species network from gene trees. A major challenge is that gene trees can differ from the display trees of the network due to **incomplete lineage sorting (ILS)**, which can occur when lineages enter a branch and then fail to coalesce. Henceforth, we assume that all species networks are binary, metric, have strictly positive lengths on their tree edges, and are **quartet-nonanamolous**.

Definition 1. A 4-taxon binary species network $\mathcal{N}$ on species set S is *quartet-nonanomalous* if it displays more than one quartet trees <u>or</u> if it displays exactly one quartet tree q_1^S and the following inequalities hold: $\mathbb{P}_{\mathcal{N}}(q_1^S) > \mathbb{P}_{\mathcal{N}}(q_2^S)$ and $\mathbb{P}_{\mathcal{N}}(q_1^S) > \mathbb{P}_{\mathcal{N}}(q_3^S)$ where q_2^S and q_3^S are the other two quartet trees on S. If $|S| > 4$, the above must hold for the induced subnetwork on every subset Y four taxa, denoted $\mathcal{N}|_Y$.

The related Lemma is useful for proving our results.

Lemma 1 (Lemma 1 in [3])**.** A 4-taxon binary network N displays exactly one quartet (i.e., is of class 1) if, and only if, it does not contain a 4-blob (see Fig. S2 in the Supplement).

It follows that if $\mathcal{N}|_Y$ is a 4-blob, it is <u>non</u>anomalous. Importantly, our model assumptions are the same as the T3 test employed by TINNiK [2]. An important consideration is whether quartet-nonanomalous networks are biologically realistic. Ané *et al.* (2024) recently studied this question [3]. They proved that all binary metric species networks are quartet-nonanomalous under the NMSC with Common Inheritance. Although such strong results do not exist for more general models of inheritance, they found that quartet-anomalous networks are rare when simulating networks under a birth-death-hybridization model [3]. Collectively, these results suggest that our assumptions are reasonable. Although one might think that quartet-based species tree methods would be guaranteed to produce a display tree of the network if it is quartet-nonanonmalous; Dinh and Baños recently showed that this is not the case [10].

Weighted Quartet Consensus (WQC). The popular species tree method ASTRAL [22] is a constrained dynamic programming for the NP-hard WQC problem [21]. The input is a set $\mathcal{T}$ of k unrooted gene trees, whose leaves are labeled by a set S of n taxa. The output is an unrooted binary tree T on S that maximizes the quartet score: $QS_{\mathcal{T}}(T) = \sum_{Y \in f_4(S)} w_{\mathcal{T}}(T|_Y)$, where $w_{\mathcal{T}}(q)$ is the number of trees in $\mathcal{T}$ that induce quartet q and $f_4(S)$ denotes all 4-taxon subsets of S, that is $f_4(S) = \{\{a,b,c,d\} : a,b,c,d \in S, a \neq b \neq c \neq d\}$. Note that ASTRAL does not explicitly compute quartets weight, which are referred to as **quartet concordance factors (qCFs)** when grouped by their associated 4-taxon subset.

TINNiK. The TINNiK algorithm is motivated by the following Lemma.

Lemma 2. Let N be a network and T_N be its tree of blobs. For any subset Y of four taxa, $T_N|_Y$ is a quartet tree, if and only if, $N|_Y$ is not a 4-blob.

See proof in Supplement. The main claim in TINNiK is that consistent estimation of 4-blobs can be achieved via hypothesis testing given qCFs [1, 2]. The **star-test** evaluates if $\mathcal{N}|_Y$ is not a star (this null hypothesis is accepted for p-values greater than β). The **T3-test** evaluates if $\mathcal{N}|_Y$ is not a 4-blob (this null hypothesis is rejected for p-values less than α). The T3-test does not make any assumptions about the structure of 4-blobs but assumes that $\mathcal{N}$ is quartet-nonanomalous; otherwise, it may incorrectly reject the null, saying that $\mathcal{N}|_Y$ is a 4-blob when it is not. However, anomalous quartets are thought to be rare in practice, as previously discussed, so the T3 test is recommended by TINNiK. The default hyperparameter

settings used by TINNiK are $\alpha = 0.05$ and $\beta = 0.95$; however, the authors caution that these settings of α, β may be inappropriate when qCFs deviate from their expected values due to noise, etc.

2.2 WQC is a TOB refinement almost surely as number of gene trees increases

Recall that our method for TOB reconstruction begins by seeking a refinement of the TOB, using a quartet-based method. To motivate this approach, we show that WQC is a TOB refinement almost surely, as the number of gene trees increases. Our proof relies on the following Lemma.

Lemma 3. Let $\mathcal{N} = (N, \Theta)$ be a binary metric species network on S that has positive lengths on its tree edges and is quartet-nonanomalous. Let T_N denote the tree of blobs for N. Then, for any quartet $a, b|c, d$ induced by T_N, $\mathbb{P}_{\mathcal{N}}(a, b|c, d) > \mathbb{P}_{\mathcal{N}}(a, c|b, d)$ and $\mathbb{P}_{\mathcal{N}}(a, b|c, d) > \mathbb{P}_{\mathcal{N}}(a, d|b, c)$.

Proof. Let $Y = \{a, b, c, d\}$, and let $\mathcal{Q}$ be the set of quartets displayed by $\mathcal{N}|_Y$. If $|\mathcal{Q}| = 1$, the inequalities hold by the definition of quartet-nonanomalous networks. If $|\mathcal{Q}| > 1$, then by Lemma 1, then the taxa in $\mathcal{N}|_Y$ is a 4-blob, and by Lemma 2, $T_N|_Y$ is not a quartet tree (i.e., it is a star). This is a contradiction. □

Now we give our main result.

Theorem 1. Let $\mathcal{N} = (N, \Theta)$ be a binary metric species network on S that has positive lengths on its tree edges and is relaxed quartet-nonanomalous. Let $\mathcal{T}$ be a set of k gene trees generated under $NMSC$ given $\mathcal{N}$. Let B_N be an unrooted binary tree on S that is a refinement of T_N, and let T be an unrooted binary tree on S that is NOT. Then, for any T and any $\epsilon > 0$, there exists a $K > 0$ and a B_N such that, given $k \geq K$ gene trees, $QS_{\mathcal{T}}(B) > QS_{\mathcal{T}}(T)$ with probability at least $1 - \epsilon$. Put simply, WQC is a refinement of the tree of blobs for N almost surely, as the number of gene trees increases.

Theorem 1 follows from the two claims. The first claim is that the observed qCFs $w_{\mathcal{T}}(q_i^Y)/k$ converge to their expected values $F^*_{\mathcal{N}}(q_i^Y) = \mathbb{P}_{\mathcal{N}}(q_i^Y)$ for all 4-taxon subsets $Y \in f_4(S)$ and for all quartet topologies $i \in \{1, 2, 3\}$. This claim, which is standard in the phylogenetics literature, means there for any $\epsilon > 0$ and any $\delta > 0$, there exists a $K > 0$ such that, given $k \geq K$ gene trees, $|w_{\mathcal{T}}(q_i^Y)/k - F^*_{\mathcal{N}}(q_i^Y)| < \delta$ for all $i \in \{1, 2, 3\}$ and $Y \in f_4(S)$ with probability at least $1 - \epsilon$. The second claim is that when the absolute error in qCFs is bounded by δ^*, there exists a TOB refinement B_N^* such that $QS_{\mathcal{T}}(B_N^*) > QS_{\mathcal{T}}(T)$ for any non-refinement T. The proof of claim 2 (and setting of δ^*) is in the Supplement (Lemma 4); it relies on the next result.

Lemma 5. Assume the same conditions and notation as Theorem 1. There exists TOB refinement that maximizes the quartet score when the error in all qCFs is zero (i.e., they equal their expected values under the NMSC given $\mathcal{N}$).

Proof. As previously mentioned, restricting the TOB, denoted T_N, to any 4-tuple of species $Y \in f_4(S)$, yields either a quartet or a star. Let $\mathcal{S}_{tree}(T_N)$ and $\mathcal{S}_{star}(T_N)$ denote the 4-taxon subsets in $f_4(S)$ partitioned based on those two cases, respectively. Now the quartet score of an unrooted binary tree t can be written as

$$\sum_{Y \in \mathcal{S}_{tree}(T_N)} F^*_{\mathcal{N}}(t|_Y) + \sum_{Y \in \mathcal{S}_{star}(T_N)} F^*_{\mathcal{N}}(t|_Y) \tag{1}$$

provided the input qCFs are error-free.

Tree part of quartet score. We first consider maximizing the "tree part" of the quartet score. Observe that for a binary tree T on S that is NOT a refinement of T_N, we have $\sum_{Y \in \mathcal{S}_{tree}(T_N)} F^*_{\mathcal{N}}(T_N|_Y) > \sum_{Y \in \mathcal{S}_{tree}(T_N)} F^*_{\mathcal{N}}(T_N|_Y)$ by Lemma 3. Thus, the tree part of the quartet score is maximized by T_N.

Star part of quartet score. We now consider maximizing the "star part" of the quartet score. First, we partition $\mathcal{S}_{star}(T_N)$ based on the observation that $Y = \{a, b, c, d\} \in \mathcal{S}_{star}(T_N)$ means there exists a unique vertex v in T_N with degree $m > 3$ whose deletion induces an m-partition $Part(v) = P_1|P_2|\ldots|P_m$ such that $a \in P_i, b \in P_j, c \in P_s, d \in P_t$ with $P_i \neq P_j \neq P_s \neq P_t$ (as an induced quartet has exactly two internal vertices of degree three and an induced star has exactly one internal vertex degree). This gives us

$$\sum_{Y \in \mathcal{S}_{star}(T_N)} F^*_{\mathcal{N}}(t|_Y) = \sum_{v \in V_{>3}(T_N)} \sum_{P \in f_4(Part(v))} \sum_{Y \in f_{cover}(P)} F^*_{\mathcal{N}}(t|_Y) \tag{2}$$

where $V_{>3}(T_N)$ contain all vertices in T_N with degree greater than 3, $f_4(Part(v))$ contains all ways of selecting 4 blocks from $Part(v)$, that is $f_4(Part(v)) = \{P_i, P_j, P_s, P_t : P_i, P_j, P_s, P_t \in Part(v), P_i, \neq P_j \neq$

$P_s \neq P_t\}$, and $f_{cover}(\{P_i, P_j, P_s, P_t\})$ contains all ways of selecting one taxa from each of the blocks, that is, $f_{cover}(\{P_i, P_j, P_s, P_t\}) = \{\{a, b, c, d\} : a \in P_i, b \in P_j, c \in P_s, d \in P_t\}$. Now it suffices to show that there exists a refinement of T_N, denoted B_N^*, such that

$$\sum_{P \in f_4(Part(v))} \sum_{Y \in f_{cover}(P)} F_{\mathcal{N}}^*(B_N^*|_Y) \geq \sum_{P \in f_4(Part(v))} \sum_{Y \in f_{cover}(P)} F_{\mathcal{N}}^*(T|_Y) \tag{3}$$

for all $v \in V_{>3}(T_N)$. Let $T_N(v)$ denote the tree obtained by contracting all edges in the tree of blobs T_N except those that are incident to v, that is $E(T_N(V)) = \{e : Bip(e) = P_i|S \setminus P_i, P_i \in Part(v)\}$. We ***claim*** that for every $v \in V_{>3}(T_N)$ there exists a refinement of $T_N(v)$, denoted B_v^*, that (1) maximizes Equation 3 and (2) does not break apart blocks of $Part(v)$, meaning that each of its non-trivial bipartitions $Z|W$ satisfies $P_i \cap Z = P_i$ and $P_i \cap W = \emptyset$ (or vice versa) for all $P_i \in Part(v)$.

It is well-established that any 4-tuple of species covering $P = \{P_i, P_j, P_s, P_t\}$ has the same expected qCFs, as a lineage sampled from any species in P_i must enter the blob in $\mathcal{N}$, associated with v in T_N, from the same cut edge (and similarly for P_j, P_s, P_t), and only after the four sampled lineages enter this blob can they coalesce; thus, the expected qCFs are fully dictated by the blob and do not depend on which particular taxon was selected; see [4] for details. This allows us to rewrite the summations in Equation 3

$$\sum_{P \in f_4(Part(v))} \sum_{Y \in f_{cover}(P)} F_{\mathcal{N}}^*(T|_Y) = \sum_{P \in f_4(Part(v))} \sum_{i=1}^{3} \alpha_i^P(T) F_{\mathcal{N}}^*(q_i^P). \tag{4}$$

where $\alpha_1^P(T)$ is the number of quartets in T that display topology $q_1^P = a, b|c, d$ for all ways of selecting $a \in P_i, b \in P_j, c \in P_s, d \in P_t$ (and similarly for α_2 and $q_2^P = a, c|b, d$ as well as α_3 and $q_3^P = a, c|b, d$). Observe that $\alpha_1^P(T) + \alpha_2^P(T) + \alpha_3^P(T) = |P_i| \cdot |P_j| \cdot |P_s| \cdot |P_t| = \beta^P$ for any binary tree T on S and that if T is a refinement of $T_N(v)$, then exactly one of $\alpha_1^P(T), \alpha_2^P(T), \alpha_3^P(T)$ equals β^P and the other two equal 0 because $T_N(v)$ forces the topology on any four species covering P to be resolved the same way. Moreover, we can maximize the inner sum for any arbitrary $P \in f_4(Part(v))$, if we select q_*^P such that $F_{\mathcal{N}}^*(q_*^P) \geq F_{\mathcal{N}}^*(q_i^P)$ for $i \in \{1, 2, 3\}$. Therefore, we just need to achieve the single best quartet resolution for P, which can then be applied all subsets of four species covering P. In other words, a (binary) solution t^* to WQC with the following inputs

- the species set contains just one taxa from each block of $Part(v)$ and
- the weight of a quartet $q = a, b|c, d$ is $\beta^P F_{\mathcal{N}}^*(q)$ where $P = \{g(a), g(b), g(c), g(d)\}$ where g is a function that maps species in S to their partition $Part(v)$ (note that β^P returns 0 if $|P| < 4$)

achieves the maximum score possible for Equation 4. It follows that turning each taxa a in t^* into a vertex incident to all taxa from block $g(a)$ maximizes Equation 3 (and 4). The resulting tree, B_v^*, is a refinement of $T_N(v)$ and contains no edges that split taxa in the same block of $Part(v)$, by construction.

Putting the tree and star parts together. Although we showed how to maximize the different parts of the quartet score separately, by bipartition compatibility, there exists a compatibility supertree for $\{B_v^* : v \in V_{>3}(T_N)\} \cup \{T_N\}$ (see Chapters 3 and 7 in [33]). This gives us our result. □

A corollary of Theorem 1 is that ASTRAL (version 3 or earlier) returns a TOB refinement almost surely, as the number of gene trees increases. This is because ASTRAL solves WQC within a constrained search space, defined by the bipartitions induced by the input gene trees, and all bipartitions have probability strictly greater than zero under the NMSC and thus will appear as the number of gene trees increases, as noted by [22]. This motivates the use of ASTRAL, and similar WQC heuristics, in seeking a TOB refinement.

2.3 False positives can be identified with a linear number of hypothesis tests

A TOB refinement B_N contains every edge in the TOB T_N plus additional edges that are not in T_N; these edges are false positives (FPs). To obtain a consistent estimator of the TOB, we need to detect (and then contract) all FP branches in a TOB refinement almost surely, as the number of gene trees increases. The hypothesis tests from TINNiK can be used for this purpose based on the following observation.

Observation 1. Let N be a network on taxon set S, T_N be its tree of blobs, and B_N be a refinement of T_N. If an edge e is in B_N but not in T_N (i.e., e is a FP edge), there exists at least one subset $Y \subseteq S$ of four taxa <u>around</u> e such that $N|_Y$ is a 4-blob.

This is easy to see otherwise e is a cut edge and must be in T_N. Critically, whether the induced subnetwork is a 4-blob depends on the choice of Y. Taking Fig. 1E as an example, the edge inducing bipartition $4,5,6,7|S\setminus\{4,5,6,7\}$ is an FP. If we select sample $Y_1 = \{1,3,4,6\}$, then $N|_Y$ is not a 4-blob; $N|_Y$ is a 4-blob if we sample $Y_2 = \{1,3,4,5\}$ (also see Fig. S3 in the Supplement). TINNiK builds the TOB from tree quartets and must apply a second inference rule to determine that $N|_{Y_1}$ should not be used for tree building. Our approach, in contrast, just needs to test one 4-taxon subset Y around e such that the $N|_Y$ is a 4-blob, if such a Y exists, because as the number of gene trees increases, the test will fail almost surely and then edge will be contracted. A naive approach that tests all 4-taxon subsets around each edge yields a consistent estimator of the TOB; however, it has time complexity $O(n^4k+n^5)$, which is the same as TINNiK. Thus, the challenge is detect FP branches efficiently, without **exhaustive search**. The following lemma gives conditions that when satisfied ensure successful detection of FP edges.

Lemma 6. Let $N^+ = (N^+, \Theta)$ be a directed binary species network on taxon set S. Suppose that $\mathcal{B}$ is a blob in N that is contracted into vertex $v_\mathcal{B}$ in its TOB T_N. Now suppose that e is an FP edge in a TOB refinement B_N that refines $v_\mathcal{B}$. Then, edge e will be successfully detected as an FP during hypothesis testing, if we select 4 species $\{a_1, a_2, a_3, a_*\}$ around e such that the following two conditions holds:

1. Each taxon is selected from a different block of the partition induced by $v_\mathcal{B}$ in T_N, that is, for $Part(v_\mathcal{B}) = P_1|P_2|\ldots|P_m$, we have $a_1 \in P_i, a_2 \in P_j, a_3 \in P_s, a_* \in P_t$ where $i \neq j \neq s \neq t$.
2. One of a_1, a_2, a_3, a_* is from a block of the $Part(v_\mathcal{B})$ that attaches to $\mathcal{B}$ via a cut edge that is incident to a reticulation vertex in $\mathcal{B}$.

The proof of this lemma is given in the Supplement. When these two conditions hold the induced subnetwork on a_1, a_2, a_3, a_* contains a 4-blob and thus e will be successfully detected as an FP, yielding a consistent estimator. There always exists a reticulation node in $\mathcal{B}$ that is incident to a cut edge of $\mathcal{B}$ (Lemma 8 in the Supplement), so we just need to design efficient sampling algorithms that is guaranteed to find one 4-taxon subset that meets these two conditions. It is easy to see that selecting one taxon from each block of the quadrapartition induced by edge e ensures condition (1) is met and that testing every taxon in S ensures we find 4-taxon subset that satisfies condition (2). This leads us to propose an algorithm that selects one taxon from each block of the quadrapartition, fixes the selection for three of the blocks, and then tests all ways of selecting a different taxa from the fourth block (and repeating so that the "alter" step is performed for every block of the quadrapartition). We refer to this as the **3-fix, 1-alter quadrapartition search (3f1a)** algorithm; see Section 1.5 in the Supplement. The advantage of 3f1a is that only $O(n)$ 4-taxon subsets need to be tested on each edge, so $O(n^2)$ in total. The qCFs for any 4-taxon subset can be computed on the fly in $O(k)$ time using the well-known binary lifting algorithm to retrieve the LCA of any pair of leaves in a gene tree in $O(1)$ time. This requires a precomputation step with time and storage complexity of $O(n\log(n)k)$, meaning the time complexity of applying the 3f1a algorithm to every branch is $O(n^2k + n\log(n)k)$.

As we will show, the 3f1a search algorithm appears to work well for biological data but not so much for simulated data, potentially because it is difficult to set α without manual inspection when qCFs deviate from their expected values, which occurs in practice, as they must be estimated from a relatively small number of gene trees, which may have estimation error, missing taxa, etc. In this realistic setting, there can be multiple 4-taxon subsets around an edge that are 4-blobs and some may yield lower p-values than others, so more aggressively searching to find a lower p-value can increase robustness to the choice of α. Sampling around a bipartition (rather than a quadrapartition) allows greater exploration of 4-taxon subsets; however, not every subset sampled will satisfy condition (1). Indeed, there are cases where $\Omega(n^2)$ 4-taxon subsets must be sampled uniformly at random to ensure that condition (1) is met by one subset with high probability, as discussed in the Supplement. For any deterministic sampling schematics around a bipartition, at least $\Omega(n^2)$ 4-taxons subsets are needed to sample to meet condition (1); otherwise, an adaptive adversary can always make none of element in our sampled 4-taxons subsets meet condition (1) since there are at least $\Omega(n^2)$ 4-taxon in the cases do not meet condition (1). Ultimately, we settled on a 3f1a-like algorithm that searches around a bipartition until a predetermined number of tests have been performed (see Algorithm 2 in the Supplement). By default, we bound the number of tests by $O(n^2)$ so that the time complexity of FP detection $O(n^3k+n\log(n)k)$ does not exceed the time of seeking a TOB refinement with TREE-QMC. (note that TREE-QMC has time complexity $O(n^3k)$, with some assumptions about the divide-and-conquer, and ASTRAL-III has time complexity $O(n^{1.726}k^{1.726}D)$, where $D = O(nk)$ is the sum of degrees of all unique nodes in the input gene trees). Lastly, we note that combining this bipartition search with 3f1a yields a consistent method and does not increase the time complexity.

2.4 Tree of Blobs Reconstruction with TOB-QMC

We implemented TOB-QMC within TREE-QMC, invoking the code from TINNiK to perform star-tests and tree-tests. TOB-QMC first seeks a TOB refinement (called the **base tree**) with the quartet-based method TREE-QMC [14, 15]. Base trees computed with ASTRAL or other methods can alternatively be given as input. Given the base tree, TOB-QMC seeks a 4-taxon subset around each edge that minimizes the p-value computed via tree-tests. *TOB-QMC-exhaustive* tests all 4-taxons subsets around each edge (bipartition) in the base tree. *TOB-QMC-3f1a* executes the 3f1a algorithm (Algorithm 1 in Supplement). *TOB-QMC-fast* and *TOB-QMC-default* performs heuristic search around the edge (biparition), stopping when the maximum number of iterations is reached, either $2n^2$ and $n^2/4$, respectively, where n is the number of taxa, or it meets a local optimal (Algorithm 2 in Supplement). TOB-QMC also applies the star-test to rule out singularities. The minimum p-values for the tree-test is stored, which allows us to perform star-tests and tree-tests given various values of α and β in an offline fashion, contracting edges based on different settings.

2.5 Method Evaluation

Simulation study. Data sets were simulated based on the procedure in [20]. First, networks with 50, 100, and 200 taxa were simulated via a birth-death-hybridization process using SiPhyNetwork [18]. For each number of taxa, 50 level-1 and 50 level-2 networks were selected for further analyses, and the TOB for these networks was computed (see Table S1 for network/TOB summary statistics). Second, 1000 gene trees were simulated for each network under NMSC (with Independent Inheritance) using PhyloCoalSimulations [12]. To vary the amount of ILS, we multiplied the branch lengths in the network by scaling factors: 2.0, 1.0, 0.5, and 0.25; this resulted in ILS levels of $\sim$ 45%, 65%, 80%, 90%, respectively (Fig. S4 in the Supplement). Third, sequences were simulated for each gene tree under the GTR+F+G model. Fourth, gene trees were estimated from sequences using IQ-TREE [26, 19]. The mean gene tree estimation error ranged from 40% to 70% (Figs. S5–S8 in Supplement). Lastly, gene trees were given as input to TOB-QMC and TINNiK, implemented within the R package MSCQuartets [2]. Software versions and commands are provided in the Supplement; also see the TOB-QMC tutorial: `https://github.com/molloy-lab/TREE-QMC/blob/main/tutorial/tree-of-blobs/README.md`.

Evaluation metrics. Methods were evaluated based on tree error and runtime. Tree error was computed as the false positive rate (FPR), false negative rate (FNR), and arithmetic mean of FNR and FPR. FNR is the number of branches (bipartitions) in the true TOB not in the estimated TOB, divided by the total number of internal branches in the true TOB. FPR is the number of branches in the estimated TOB not in the true TOB, divided by the total number of internal branches in the estimated TOB.

Hyperparameter tuning. Both TOB-QMC and TINNiK rely on α, β hyperparameters. Lower α and higher β result in more edges being contracted by TOB-QMC. There is no approach for selecting α, β. However, the authors of $TINNiK$ evaluated their method on a species network with 23 taxa and found that setting $\beta = 1$ and $\alpha \in [10^{-7}, 10^{-3}]$ yielded the correct TOB; however, higher settings of α gave an under-resolved TOB and lower settings gave an over-resolved TOB [2]. To ensure fair settings for both methods, we tuned hyperparameters on the 50- and 100-taxon data sets with <u>true</u> gene trees. Looking at the mean of FNR and FPR across all combinations of $\alpha \in \{10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}\}$ and $\beta \in \{0.90, 0.95, 1.00\}$, we found that setting $\beta = 0.95$ and $\alpha = 10^{-7}$ gave the good results for both methods, as error increased for higher settings of α (Fig. S9–S12 in the Supplement). We also explored the impact of TOB-QMC's search algorithms, compared to the base tree computed with TREE-QMC without any edges contracted (Figs. S13–S14 in the Supplement). All algorithms reduced FPR compared to the base tree with little impact on FNR. TOB-QMC-3f1a had the highest FPR error, whereas default and fast modes achieved similar FPR to TOB-QMC-exhaustive. Thus, in the simulation study, we benchmarked TOB-QMC-default and TOB-QMC-fast against TINNiK, with $\beta = 0.95$ and $\alpha = 10^{-7}$, on <u>estimated</u> gene trees.

Phylogenomic Analyses. To evaluate the utility of TOB-QMC-3f1a we analyzed phylogenomic data sets for the bee subfamily *Nomiinae* (31 taxa, 852 gene trees) [7, 5], the butterfly subfamily *Heliconiinae* (63 taxa, 3393 gene trees) [9], and seed plants (96 taxa, 406 gene trees) [27, 20]. The bee data set was the smallest so we ran also ran TINNiK on this data set. Both TINNiK and TOB-QMC-3f1a were fast, taking less than 1 minute, as expected based on our scalability study. The other phylogenomic data sets had larger numbers of taxa but we were able to analyzed them with TREE-QMC-3f1a in just a few minutes on a laptop.

3 Results

3.1 Simulation Study

Impact of ILS level and network level. We first evaluated TOB-QMC and TINNiK varying the ILS level and the network level on data sets with estimated gene trees. For data sets with 100 taxa (Fig. 2), TOB-QMC's base tree had the worst FPR and the best FNR of all methods. Its FNR was generally low ($\leq 5\%$) for all model conditions, except for the highest ILS level, in which case FNR was $\leq 10\%$ on average. TOB-QMC-fast and TOB-default had slightly higher FNR compared to the base tree but much lower FPR, suggesting branch contraction based on the minimum p-value search was effective. Their FPR varied with model condition, ranging from $\sim$5% for the easiest (low ILS and level-1) to $\sim$18% for the hardest (high ILS and level-2). Likewise, the performance difference between TOB-QMC and TINNiK increased with model condition difficulty, with TOB-QMC outperforming TINNiK. These overall trends held on data sets with 50 and 200 taxa (Fig. S15–S16), although TINNiK could not be run on these larger data sets.

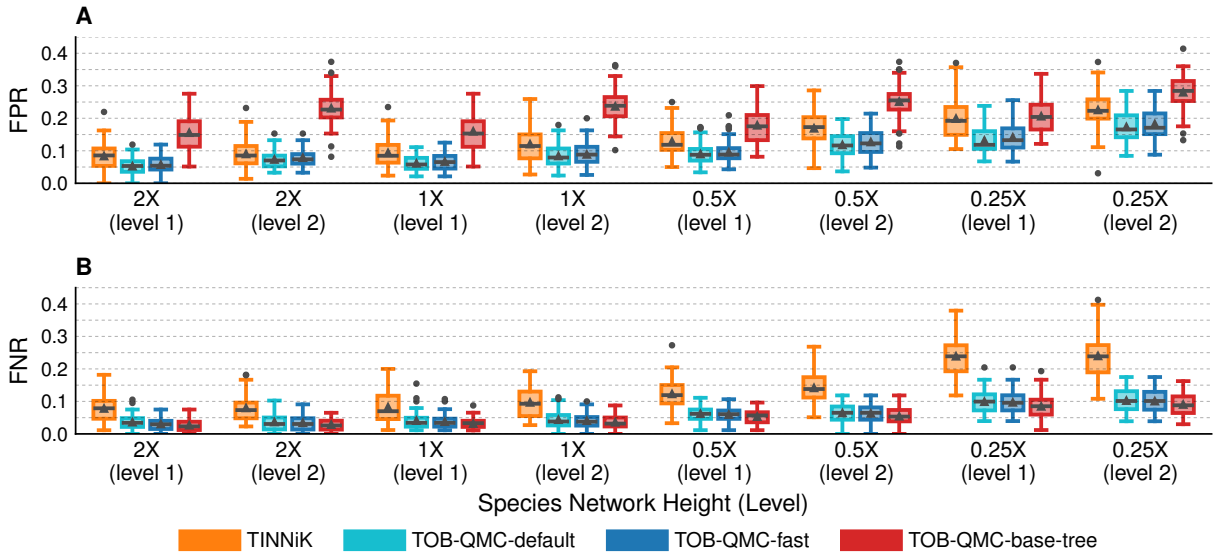

Fig. 2: FPR and FNR for TOB reconstruction methods on 100-taxon data sets with estimated gene trees. Model condition on x-axis indicates network level and branch scaling factor (lower factor is higher ILS).

Impact of varying the number of taxa. Next we evaluated TOB-QMC and TINNiK for increasing numbers of taxa (Figs. 3). All methods were fast for 50 taxa. At 100 taxa, TINNiK took over 2 hours on average, while both TOB-QMC version took less than 0.5 hours. At 200 taxa, TINNiK could not run within our wall clock limit (48 hours) but TOB-QMC-default took less than 2.5 hours on average; TOB-QMC-fast took less than 1 hour on average. For level-2 networks (no scaling), tree error (arithmetic mean of FNR and FPR) was relatively stable for TOB-QMC ($\sim$ 5%), which is reasonable given the summary statistics for networks (Table S1); however, tree error increased for TINNiK going from $\sim$ 5% for 50 taxa to $\sim$ 10% for 100 taxa, on average. This may be due to TINNiK building the tree after applying tests to

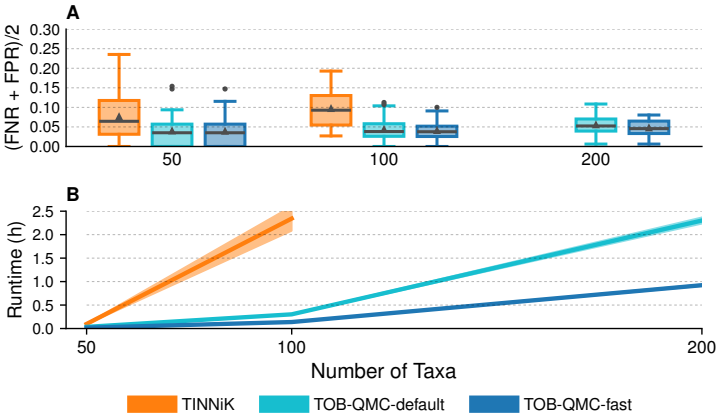

Fig. 3: Tree error and runtime for TOB reconstruction methods for level-2 networks (no branch length scaling).

all 4-taxon subsets and the greater potential to make mistakes as the number of taxa grows. Similar trends were observed for the other model conditions (Figs. S17–S20).

3.2 Phylogenomic Analyses

Nomiinae bees. On the bee data set, both TINNiK and TOB-QMC-3f1a returned the same binary TOB when α was set to 10^{-5} and 10^{-7}, respectively. Increasing α to 10^{-4} and 10^{-6}, respectively, resulted in a blob for the genus *Stictonomia* (Fig. S21; Table S2). A prior study by [7] reported conflicting species tree topologies on this genus, and gene flow was detected by [5] using several different statistical tests for hybridization, including MSCquartets [29], which is similar to TINNiK, as well as HyDe [6] and Patterson's D statistics [28]. As α increased, the TOBs returned by both methods became less refined. Notably, TINNiK needed to be re-run for each setting of α, β, whereas TOB–3f1a did not because edges in its annotated "base tree" were simply contracted based on a different significance thresholds. This saved time and enabled us to interpret why edges were being contracted. For example, TOB-QMC-3f1a contracted the edge associated with *Stictonomia* because the T3-test applied to these three taxa plus *Nomiapis bispinosa* was $p = 1.4 \times 10^{-07}$. The associated qCFs were 80/69/21 (0.47/0.41/0.12). Although these qCFs are suggestive of gene flow (because two qCFs are equal and lesser than the third for tree-like evolution), there are only 170 gene trees so there is limited data to support this conclusion.

Heliconiinae butterflies. The TOB reconstructed by TOB-QMC-3f1a for butterflies contracted 14 edges associated with introgression, when setting $\alpha = 10^{-15}$ and $\beta = 0.95$ (Fig. S22; Table S3). The contracted edges were all congruent with the original study, which constructed a species tree with ASTRAL and then annotated it with introgression events (Fig. 2B in [9]); see the Supplement for details. The original study also detected introgression impacting 3 species in *Wallacei* clade (branch 43) and in *Eueides* clade (e.g., branches 48 and 50). Branch 43 had $p = 1.5 \times 10^{-8}$ and qCFs of 2710/205/106 (0.90/0.07/0.04), and branch 48 had $p = 0.0001$ and qCFs of 1533/103/55 (0.91/0.06/0.03). Thus, these events were significant in the TOB-QMC-3f1 analysis but did not survive the high significance threshold of $\alpha = 10^{-15}$. Both are events are characterized by lower amounts of introgression. We exhaustively tested all taxa around the quadrapartition induced by branch 43 (i.e., we performed T3 tests on qCFs for *H. burneyi*, *H. wallacei*, *H. egeria*, plus every all other taxa, so 63 tests total). All tests had similar qCFs and p-values (Table S4), so the strength of intro-

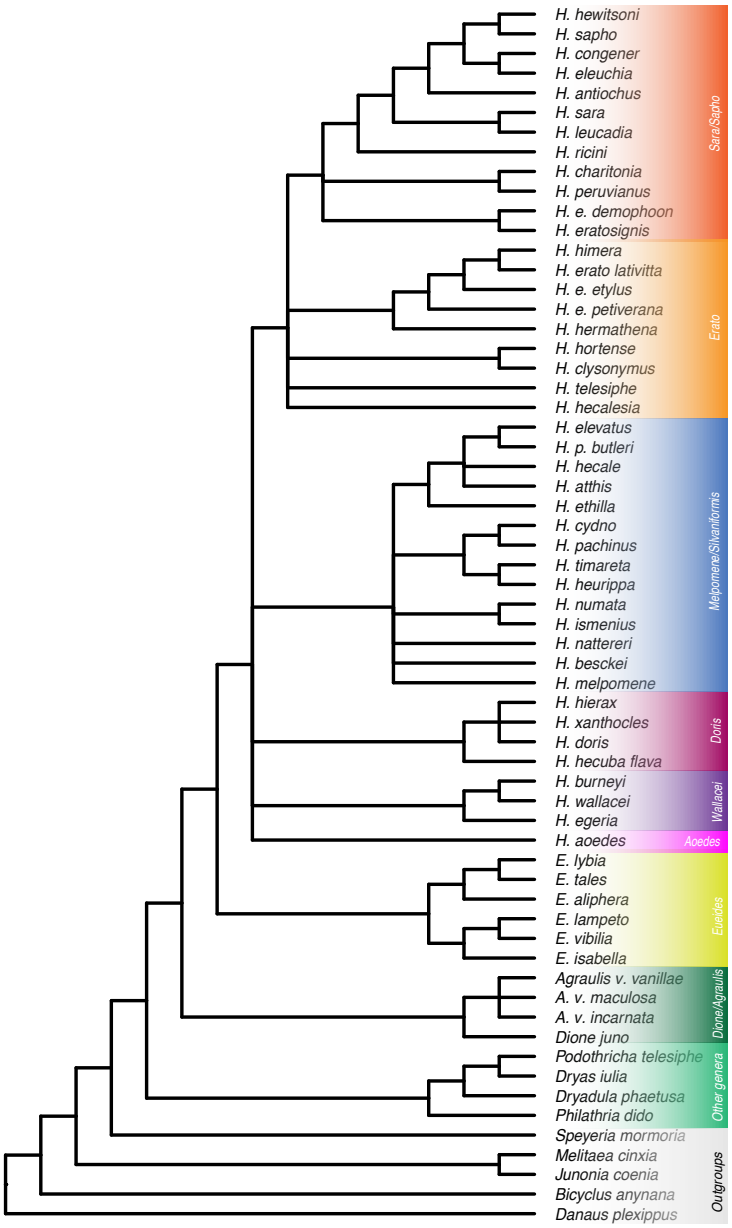

Fig. 4: TOB for butterflies setting $\alpha = 10^{-15}$ and $\beta = 0.95$.

gression is accurately recovered by the 3f1a algorithm. Overall, this suggests that the α hyperparameter and amount of introgression (inheritance proportion) are important factors to consider in TOB reconstruction. It also highlights the utility of branch annotations when constructing TOBs.

Seed Plants. Lastly, we applied TOB-QMC to a seed plant data set, previously analyzed to evaluate the divide-and-conquer network method InPhyNet [20]. In this study, TINNiK was used to build a TOB for subset decomposition, with $\alpha = 0.001$. The resulting TOB was fairly unresolved, which may impact on the network reconstructed by InPhyNet. The TOB produced by TOB-QMC-3f1a also fairly unresolved for $\alpha = 0.001$, contracting 25 branches (Fig. S23; Tables S5–S6). However, the amount of gene flow (inheritance proportion) was highly variable across the tree. Some branches towards the root had weaker signal (e.g., the qCFs were 220/27/3 (0.88/0.11/0.01) and 179/26/2 (0.87/0.13/0.01) for branches 6 and 7, respectively), whereas some branches close to the tips has stronger signal (e.g., the qCFs were 149/134/25 (0.48/0.44/0.08) for branch 25). Stronger signals are less likely to be due to spurious noise, although problems in identification of single copy orthologs can lead to biases that appear like introgression. Thus, this branch information has the potential to inform network reconstruction based on estimated TOBs.

4 Conclusions

We have presented a new framework for reconstructing tree of blobs (TOBs) under the network multispecies coalescent (NMSC) that first seeks a TOB refinement and then contracts edges from it to get a TOB. This approach is simple and related to how researchers often analyze biological data, reconstructing a species tree with ASTRAL and annotating it with introgression events. This approach was shown to positively misleading by [10] when the ASTRAL tree is intrepreted as a display tree of the network. We provide theoretical justification and context for interpretting such analyses, by showing that an optimal solution to WQC is a TOB refinement almost surely, as the number of gene trees increases. Moreover, we show that only a linear number of 4-taxon subsets around each branch need to be tested to detect false positive edges almost surely, as the number of gene trees increases. Together, these results yield a consistent and fast approach for TOB reconstruction. We also present the first large-scale study of TOB reconstruction methods, finding that our method (TOB-QMC) generally outperformed the only other existing method TINNiK. This type of benchmarking is difficult because of the impact of the α, β hyperparameters. In real analyses, TOB-QMC enabled interpretable exploration of these hyperparameters as well as the strength of the introgression (inheritance proportions) across different parts of the tree.

Our study is not without limitations. Perhaps the biggest one is that we assume species networks are quartet-nonanomalous, as does TINNiK. A prior study by [3] suggests quartet-anomalous networks are rare; however, future work should attempt to relax this assumption and investigate the impact of anomalous networks on TOB reconstruction. Our study also does evaluate methods for hypothesis testing, instead relying on the same tests as TINNiK. Related tests have been developed, for evaluating hard polytomies [31] and hybridization (e.g.[28]). Future work should explore the utility of other tests and develop procedures for automatically setting significance thresholds and other hyperparameters, as they can be difficult to configure in practice. Lastly, it is also worth noting that during the time frame of our study, a new network reconstruction method, called Squirrel, was released [17]. Squirrel creates candidate TOBs, through edge contractions, which is similar to TOB-QMC, and then "refines" the blobs to produce a network (the MSCquartet package, which supports TINNiK, also includes tools for blob refinement). Unlike TINNiK and TOB-QMC, Squirrel is not designed for coalescent-based reconstruction and has no statistical guarantees under the NMSC (it uses quarnets estimated from molecular sequences and contracts branches based on split support). We anticipate that TOB-QMC will greatly enable network reconstruction under the NMSC, either within a blob refinement framework or a divide-and-conquer framework; future work should explore this further.

Acknowledgments. This material is based upon work supported by the U. S. National Science Foundation (NSF) under Grant No. 2441458 (to EKM) as well as the State of Maryland. The authors thank Francesco Cicconardi for sharing the gene trees for *Heliconiinae* butterflies. The authors thank Claudia Solís-Lemus and Nathan Kolbow for discussions regarding phylogenetic network simulations and applications of tree of blobs, which were initiated at the NSF-funded Institute for Computational and Experimental Research in Mathematics.

Code and Data Availability. TOB-QMC is implemented within TREE-QMC. The source code is publicly available on Github at `https://github.com/molloy-lab/TREE-QMC`.

References

1. Allman, E.S., Baños, H., Mitchell, J.D., Rhodes, J.A.: The tree of blobs of a species network: identifiability under the coalescent. Journal of Mathematical Biology **86**(1), 10 (2023)
2. Allman, E.S., Baños, H., Mitchell, J.D., Rhodes, J.A.: Tinnik: inference of the tree of blobs of a species network under the coalescent model. Algorithms for Molecular Biology **19**(1), 23 (2024)
3. Ané, C., Fogg, J., Allman, E.S., Baños, H., Rhodes, J.A.: Anomalous networks under the multispecies coalescent: theory and prevalence. Journal of Mathematical Biology **88**(3), 29 (2024). https://doi.org/10.1007/s00285-024-02050-7
4. Baños, H.: Identifying species network features from gene tree quartets under the coalescent model. Bulletin of Mathematical Biology **81**(2), 494–534 (2019). https://doi.org/10.1007/s11538-018-0485-4
5. Bjorner, M.B., Molloy, E.K., Dewey, C.N., Solís-Lemus, C.: Detectability of varied hybridization scenarios using genome-scale hybrid detection methods. Bulletin of the Society of Systematic Biologists **3**(1) (2024). https://doi.org/10.18061/bssb.v3i1.9284
6. Blischak, P.D., Chifman, J., Wolfe, A.D., Kubatko, L.S.: HyDe: A python package for genome-scale hybridization detection. Systematic Biology **67**(5), 821–829 (03 2018). https://doi.org/10.1093/sysbio/syy023
7. Bossert, S., Murray, E.A., Pauly, A., Chernyshov, K., Brady, S.G., Danforth, B.N.: Gene tree estimation error with ultraconserved elements: An empirical study on Pseudapis bees. Systematic Biology **70**(4), 803–821 (2020). https://doi.org/10.1093/sysbio/syaa097
8. Cao, Z., Zhu, J., Nakhleh, L.: Empirical Performance of Tree-Based Inference of Phylogenetic Networks. In: Huber, K.T., Gusfield, D. (eds.) 19th International Workshop on Algorithms in Bioinformatics (WABI 2019). Leibniz International Proceedings in Informatics (LIPIcs), vol. 143, pp. 21:1–21:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany (2019). https://doi.org/10.4230/LIPIcs.WABI.2019.21
9. Cicconardi, F., Milanetti, E., Pinheiro de Castro, E.C., Mazo-Vargas, A., Van Belleghem, S.M., Ruggieri, A.A., Rastas, P., Hanly, J., Evans, E., Jiggins, C.D., Owen McMillan, W., Papa, R., Di Marino, D., Martin, A., Montgomery, S.H.: Evolutionary dynamics of genome size and content during the adaptive radiation of heliconiini butterflies. Nature Communications **14**(1), 5620 (Sep 2023). https://doi.org/10.1038/s41467-023-41412-5
10. Dinh, V., Baños, H.: Misspecification strikes: Astral can mislead in the presence of hybridization, even for nonanomalous scenarios. Molecular Biology and Evolution **42**(3), msaf049 (2025). https://doi.org/10.1093/molbev/msaf049
11. Felsenstein, J.: Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution **17**(6), 368–376 (1981). https://doi.org/10.1007/BF01734359
12. Fogg, J., Allman, E.S., Ané, C.: PhyloCoalSimulations: A simulator for network multispecies coalescent models, including a new extension for the inheritance of gene flow. Systematic Biology **72**(5), 1171–1179 (2023). https://doi.org/10.1093/sysbio/syad030
13. Gusfield, D., Bansal, V., Bafna, V., Song, Y.S.: A decomposition theory for phylogenetic networks and incompatible characters. Journal of Computational Biology **14**(10), 1247–1272 (2007). https://doi.org/10.1089/cmb.2006.0137
14. Han, Y., Molloy, E.K.: Improving quartet graph construction for scalable and accurate species tree estimation from gene trees. Genome Research **33**(7), 1042–1105 (2023). https://doi.org/10.1101/gr.277629.122
15. Han, Y., Molloy, E.K.: Improved robustness to gene tree incompleteness, estimation errors, and systematic homology errors with weighted tree-qmc. Systematic Biology p. syaf009 (2025). https://doi.org/10.1093/sysbio/syaf009
16. Hejase, H.A., VandePol, N., Bonito, G.M., Liu, K.J.: Fastnet: Fast and accurate statistical inference of phylogenetic networks using large-scale genomic sequence data. In: Blanchette, M., Ouangraoua, A. (eds.) Comparative Genomics. pp. 242–259. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00834-5_14
17. Holtgrefe, N., Huber, K.T., van Iersel, L., Jones, M., Martin, S., Moulton, V.: Squirrel: Reconstructing semi-directed phylogenetic level-1 networks from four-leaved networks or sequence alignments. Molecular Biology and Evolution **42**(4), msaf067 (2025). https://doi.org/10.1093/molbev/msaf067
18. Justison, J.A., Solis-Lemus, C., Heath, T.A.: SiPhyNetwork: An R package for simulating phylogenetic networks. Methods in Ecology and Evolution **14**(7), 1687–1698 (2023). https://doi.org/10.1111/2041-210X.14116
19. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A., Jermiin, L.S.: Modelfinder: fast model selection for accurate phylogenetic estimates. Nature methods **14**(6), 587–589 (2017)
20. Kolbow, N., Kong, S., Solís-Lemus, C.: A method for massively scalable phylogenetic network inference. bioRxiv (2025). https://doi.org/10.1101/2025.05.05.652278
21. Lafond, M., Scornavacca, C.: On the weighted quartet consensus problem. Theoretical Computer Science **769**, 1–17 (2019). https://doi.org/10.1016/j.tcs.2018.10.005
22. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T.: ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics **30**(17), i541–i548 (2014). https://doi.org/10.1093/bioinformatics/btu462

23. Molloy, E.K., Durvasula, A., Sankararaman, S.: Advancing admixture graph estimation via maximum likelihood network orientation. Bioinformatics **37**(Supplement_1), i142–i150 (2021). https://doi.org/10.1093/bioinformatics/btab267
24. Molloy, E.K., Warnow, T.: Statistically consistent divide-and-conquer pipelines for phylogeny estimation using NJMerge. Algorithms for Molecular Biology **14**(1), 14 (2019). https://doi.org/10.1186/s13015-019-0151-x
25. Molloy, E.K., Warnow, T.: TreeMerge: a new method for improving the scalability of species tree estimation methods. Bioinformatics **35**(14), i417–i426 (2019). https://doi.org/10.1093/bioinformatics/btz344
26. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q.: Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution **32**(1), 268–274 (2015)
27. One Thousand Plant Transcriptomes Initiative: One thousand plant transcriptomes and the phylogenomics of green plants. Nature **574**(7780), 679–685 (2019). https://doi.org/10.1038/s41586-019-1693-2
28. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D.: Ancient admixture in human history. Genetics **192**(3), 1065–1093 (2012). https://doi.org/10.1534/genetics.112.145037
29. Rhodes, J.A., Baños, H., Mitchell, J.D., Allman, E.S.: MSCquartets 1.0: quartet methods for species trees and networks under the multispecies coalescent model in R. Bioinformatics **37**(12), 1766–1768 (2020). https://doi.org/10.1093/bioinformatics/btaa868
30. Saitou, N., Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution **4**(4), 406–425 (1987)
31. Sayyari, E., Mirarab, S.: Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes (Basel) **9**(3), 132 (2018). https://doi.org/10.3390/genes9030132
32. Solís-Lemus, C., Ané, C.: Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLOS Genetics **12**(3), 1–21 (2016). https://doi.org/10.1371/journal.pgen.1005896
33. Warnow, T.: Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation. Cambridge University Press, Cambridge, United Kingdom (2017)
34. Yu, Y., Degnan, J.H., Nakhleh, L.: The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. PLOS Genetics **8**(4), 1–10 (04 2012). https://doi.org/10.1371/journal.pgen.1002660
35. Yu, Y., Dong, J., Liu, K.J., Nakhleh, L.: Maximum likelihood inference of reticulate evolutionary histories. Proceedings of the National Academy of Sciences **111**(46), 16448–16453 (2014). https://doi.org/10.1073/pnas.1407950111
36. Yu, Y., Nakhleh, L.: A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics **16 Suppl 10**(S10), S10 (2015). https://doi.org/10.1186/1471-2164-16-S10-S10
37. Yu, Y., Ristic, N., Nakhleh, L.: Fast algorithms and heuristics for phylogenomics under ils and hybridization. BMC Bioinformatics **14**(15), S6 (2013). https://doi.org/10.1186/1471-2105-14-S15-S6
38. Zhu, J., Liu, X., Ogilvie, H.A., Nakhleh, L.K.: A divide-and-conquer method for scalable phylogenetic network inference from multilocus data. Bioinformatics **35**(14), i370–i378 (2019). https://doi.org/10.1093/bioinformatics/btz359