Exploring the distribution of phylogenetic networks generated under a birth-death-hybridization process

Joshua A. Justison[1*] and Tracy A. Heath[1]

February 8, 2024

Author Affiliations
[1]Department of Ecology, Evolution, & Organismal Biology, Iowa State University, Ames, Iowa USA
ORCID ID
Joshua A. Justison: 0000-0002-0233-4413. Tracy A. Heath: 0000-0002-0087-2541
Email
[*]justisnojo@gmail.com

ABSTRACT
Gene-flow processes such as hybridization and introgression play important roles in shaping diversity across the tree of life. Recent studies extending birth-death models have made it possible to investigate patterns of reticulation in a macroevolutionary context. These models allow for different macroevolutionary patterns of gene flow events that can either add, maintain, or remove lineages—with the gene flow itself possibly being dependent on the relatedness between species—thus creating complex diversification scenarios. Further, many reticulate phylogenetic inference methods assume specific reticulation structures or phylogenies belonging to certain network classes. However, the distributions of phylogenetic networks under reticulate birth-death processes are poorly characterized, and it is unknown whether they violate common methodological assumptions. We use simulation techniques to explore phylogenetic network space under a birth-death-hybridization process where the hybridization rate can have a linear dependence on genetic distance. Specifically, we measured the number of lineages through time and role of hybridization in diversification along with the proportion of phylogenetic networks that belong to commonly used network classes (e.g., tree-child, tree-based, or level-1 networks). We find that the growth of phylogenetic networks and class membership are largely affected by assumptions about macroevolutionary patterns of gene flow. In accordance with previous studies, a lower proportion of networks belonged to these classes based on type and density of reticulate events. However, under a birth-death-hybridization process, these factors form an antagonistic relationship; the type of reticulation events that cause high membership proportions also lead to the highest reticulation density, consequently lowering the overall proportion of phylogenies in some classes. Further, we observed that genetic distance–dependent gene flow and incomplete sampling increase the proportion of class membership, primarily due to having fewer reticulate events. Our results can inform studies if their biological expectations of gene flow are associated with evolutionary histories that satisfy the assumptions of current methodology and aid in finding phylogenetic classes that are relevant for methods development.

Keywords: birth-death-hybridization process, phylogenetic networks, simulation

1 Introduction

Gene flow—the exchange of genetic material between species—has been observed in a number of systems (Taylor and Larson 2019) and plays an important role in shaping Earth's biodiversity (Stebbins 1959; Bock 2010; Mallet et al. 2016). Phylogenetic networks are used to directly infer reticulate histories and are vital in informing our understanding of adaptive radiations (Ottenburghs et al. 2016), mimicry patterns (Edelman et al. 2019), and even the phylodynamics of infectious diseases like COVID-19 (Wang et al. 2022). The

field has recently seen an explosion of new methods for estimating phylogenetic networks (Elworth et al. 2019). However, due to the complex topological landscape of phylogenetic networks, robust model-based techniques remain limited in their ability to scale to large datasets (Hejase and Liu 2016). The birth-death-hybridization process (Morin and Moret 2006; Woodhams et al. 2016; Zhang et al. 2018) is an extension of a model commonly used to describe lineage diversification (*i.e.,* the birth-death process, Kendall 1948; Nee et al. 1994b) and provides a statistical framework for understanding the generation of phylogenetic networks. With this model, we can understand how reticulate processes influence patterns of diversification and affect the distribution of species across the tree of life.

Birth-death processes are used to model the phylogenetic branching process on a macroevolutionary scale, thus revealing expected patterns of speciation and extinction (Kendall 1948; Raup 1985; Nee et al. 1994b; Nee 2006). A rich theoretical framework for this process allows us to describe the distribution of branch lengths (Steel and Mooers 2010; Mooers et al. 2012), growth dynamics (Nee et al. 1994a; Stadler 2008), and topological properties (Mooers and Heard 1997; Lambert and Stadler 2013) of phylogenetic trees. The birth-death family of models has proven useful for estimating diversification rates (Nee et al. 1994b,a; Höhna et al. 2015) and choosing biologically motivated priors in Bayesian analyses (Rannala and Yang 1996; Velasco 2008). However, analyses of divergence times (Pybus and Harvey 2000) and tree shape (Mooers and Heard 1997; Heath et al. 2008b; Jones 2011) have shown that empirical phylogenies differ from those expected under a constant rate birth-death process. To more accurately explain the observed patterns, these statistical models have undergone several extensions to account for a wide range of biological processes and systematic sampling challenges.

The birth-death-hybridization process (formally described below) generates phylogenetic networks under a process that allows lineages to undergo speciation, extinction, and reticulation (hybridization). Although—for eponymous consistency—events are denoted as hybridizations, the process does not make any mechanistic assumptions about gene flow; reticulate events on phylogenetic networks can model the many processes where genetic material is exchanged between lineages. Consequently, phylogenetic network inference has been used to describe histories of introgression (Thawornwattana et al. 2018; Myers et al. 2022), hybridization (Morales-Briones et al. 2018; Dolinay et al. 2021), and lateral gene transfer (Betat et al. 2015). The specific biological interpretation of a given reticulation event is a challenge that depends largely on the system under investigation and is further complicated by how reticulation events are modeled on the phylogenetic network (Hibbins and Hahn 2022).

Even though phylogenetic network methods are largely agnostic to the underlying cause of reticulation, existing implementations of the birth-death-hybridization process make different assumptions about how reticulation events are represented on the network. We denote each type of reticulation—lineage generative, lineage neutral, and lineage degenerative—based on how the events change the number of lineages on the phylogenetic network (Box 1). The structure and timing of each reticulation type can provide some biological meaning to reticulation events (Box 1; but see Hibbins and Hahn 2022). Consequently, caution is needed when trying to infer and interpret reticulate patterns; the lineage effects from the types of reticulations can have far reaching implications on species diversification dynamics.

Box 1: Description of Reticulation Types

Patterns of gene flow can take different shapes on phylogenetic networks. We differentiate types of reticulation depending on changes in the number of lineages on the phylogeny after an event (Fig. 1).

- **Lineage Generative**: A new lineage is created (denoted as m-type in Janssen and Liu 2021)
- **Lineage Neutral**: The number of lineages remains constant (n-type)
- **Lineage Degenerative**: The process has one fewer lineages after the event (y-type)

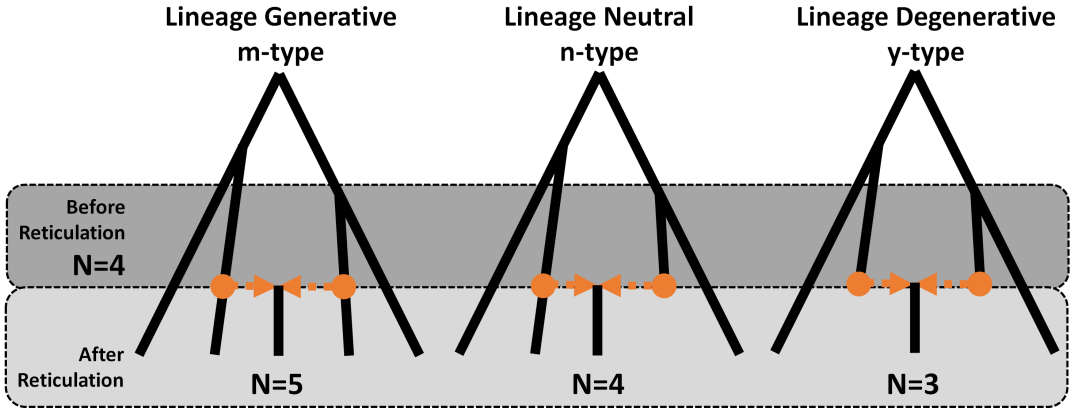

Figure 1: Types of Reticulation. Orange circles denote the parental nodes leading to the reticulate node. The dark grey box delineates the period before the reticulation event when the number of lineages is $N = 4$. The light grey box highlights the period after reticulation. Lineage generative hybridization (m-type) occurs when a reticulation event results in a gain of one lineage ($N = 5$). Lineage neutral hybridization (n-type) results in a net zero change in the number of lineages ($N = 4$), and degenerative hybridization (y-type) reduces the number of lineages by one ($N = 3$).

The table below outlines possible interpretations for each reticulate pattern. For example, lineage generative events are consistent with hybrid speciation or allopolyploidization since these events directly create new hybrid lineages. Introgression or lateral gene transfer can be seen as lineage neutral events as genetic material is exchanged, but the number of lineages remains constant. Interestingly, the formation of a new hybrid lineage also has potential to be lineage neutral or even lineage degenerative if genetic swamping occurs (Todesco et al. 2016). Genetic swamping is the loss of parental lineages after hybridization, resulting from either the hybrid having a higher fitness or a breakdown of reproductive barriers which eventually merges the hybrid lineage with its parental lineages. Extinction and incomplete sampling can further complicate the interpretation of reticulation events. Failure to sample the donor lineages of reticulation—so called ghost lineages (Ottenburghs 2020; Tricou et al. 2022)—may change its structure on the reconstructed phylogenetic network (Fig. S1). Without complete sampling, lineage generative and lineage neutral hybridization can appear as a different types of reticulation events.

Reticulate Pattern	**Possible Biological Processes**
Lineage Generative	• Hybrid speciation • Allopolyploidy
Lineage Neutral	• Genetic swamping with one parental lineage • Introgression • Lateral gene transfer • Lineage generative event with one unsampled parental lineage
Lineage Degenerative	• Genetic swamping with both parental lineages • Lineage generative event with both parental lineages unsampled • Lineage neutral event with one unsampled parental lineage

The type of reticulation also affects the topological properties of phylogenetic networks (Janssen and Liu 2021). Phylogenetic networks can be grouped into a multitude of classes based on their topological

features (see Box 2 and Zhang 2019; Kong et al. 2022). Understanding the topological properties of phylogenetic networks is exceptionally important because of the scalability challenges associated with inference (Hejase and Liu 2016; Elworth et al. 2019). Some network classes have useful identifiabilty results (Willson 2010b; Pardi and Scornavacca 2015; Solís-Lemus and Ané 2016; Francis and Moulton 2018; Erdős et al. 2019), unique encodings (Van Iersel and Moulton 2014; Linz and Semple 2020; Semple and Toft 2021), or polynomial-time algorithms for reconstruction (Erdős et al. 2019; Murakami et al. 2019; Semple and Toft 2021). Consequently, some methods will assume *a priori* that the underlying network belongs to a given class. For example, both the algorithms of Bordewich and Tokac (2016) and Wawerka et al. (2022) assume that networks are tree-child, while the software packages `SNaQ` (Solís-Lemus and Ané 2016) and `NANUQ` (Allman et al. 2019) assume that networks are level-1 (see Box 2 for definitions of network classes). Thus, to ensure methodological assumptions are met, it is important to understand the complex topological landscape of phylogenetic networks.

Some analytical results exist for network classes; however, they are currently only applicable to special cases of the birth-death-hybridization process and may be difficult to generalize (Bienvenu et al. 2022). Nevertheless, simulation has been an essential technique for characterizing the distribution of phylogenetic networks generated from various processes (Arenas et al. 2008; Janssen and Liu 2021). Through simulation, researchers found that increasing the number of hybridizations on a network of fixed size decreased the proportion of networks that belonged to certain classes (e.g., tree-based and tree-child; Janssen and Liu 2021). Additionally, for branching processes—including a special case of the birth-death-hybridization process—the type of reticulation also affected the topological properties of simulated networks. However, when comparing types of reticulation, a different simulation tool was used for each type, making it difficult to distinguish the effects of the reticulation type from artifacts of the generators themselves. Lastly, simulating to a fixed number of taxa and hybridizations obfuscates the role of hybridization in lineage diversification. The birth-death-hybridization process offers a complimentary approach for analyzing specific properties of phylogenetic networks. By having all three reticulation types as a part of the model, we can directly assess their effects on topological properties and lineage growth of phylogenetic networks.

In this study, we examine and summarize the distribution of phylogenetic networks generated by the birth-death-hybridization process. Specifically, we used a unified modeling framework to consider different assumptions about the type reticulation events. Additionally, we modeled a genetic distance dependence on hybridization by decreasing the rate of hybridization linearly based on the relatedness between two species. To understand the impact of hybridization on lineage diversification in phylogenetic networks, we simulated the average of lineages through time. We also simulated phylogenetic networks and characterized the proportion of phylogenies that belong to common network classes. We simulated phylogenies with incomplete lineage sampling and genetic distance-dependent hybridizations to consider how important systematic and biological processes might affect the properties of generated phylogenetic networks.

Our work explores the space of phylogenetic networks generated under the birth-death-hybridization process and assesses how different assumptions about the process of gene flow influence this distribution. Developing an intuition for how hybridization creates and shapes diversity will be valuable for selecting and fitting appropriate macroevolutionary models. Lastly, understanding the distribution of topological properties from a biologically relevant process can give insight into how often network class assumptions are violated in empirical systems, potentially biasing inferences and steering methods development.

2 Methods

2.1 Phylogenetic Network Properties

Phylogenetic networks are similar to phylogenetic trees but have additional reticulate nodes and edges representing genetic material crossing species boundaries. Although these nodes can represent specific biological processes such as lateral-gene transfer or introgression, the network does not distinguish which process occurs. Though, for consistency with the formal descriptions of the birth-death-hybridization process (Morin and Moret 2006; Woodhams et al. 2016; Zhang et al. 2018), we, at times, refer to reticulation events as hybridization nodes. Like phylogenetic trees, networks can be rooted or unrooted. In this study, we consider rooted phylogenetic networks. Formally, a phylogenetic network is a directed acyclic graph that contains four node types:

- *root*: A node of in-degree 0 and out-degree 1 if the process begins at the root or out-degree 2 if the root represents the most recent common ancestor of the process.
- *tree*: A node of in-degree 1 and out-degree 2. These nodes typically represent speciation events.
- *hybrid*: A node of in-degree 2 and out-degree 1. Hybrid nodes denote reticulation events.
- *leaf*: Nodes of in-degree 1 and out-degree 0 that can represent either extant lineages that survived until the present or they can denote an extinction event.

A directed edge from nodes u to w is denoted as *(u,w)*. We further describe *(u,w)* as a hybrid edge if w is a hybrid node. Hybrid edges additionally have weights (γ and $1-\gamma$) that are used to describe the proportion of genetic material that contributes to the hybrid node. Hybrid nodes are considered stacked if the child of a hybrid node is another hybrid node (*i.e.,* there exists an edge *(x,y)* where both x and y are hybrid nodes). We restrict our analyses to identifying tree-based, fold-unfold stable (FU-stable), tree-child, normal, and level-k networks (Box 2).

Box 2: Description of Network Classes

Phylogenetic network classifications describe the topological properties of a given phylogenetic network. In this study, we focus on five different classifications:

- **Level-k**: Any biconnected component of the network having at most k hybrid nodes (Jansson and Sung 2006).
- **Tree-based**: A phylogenetic network that can be constructed by starting with a phylogenetic tree and sequentially adding reticulation edges (Francis and Steel 2015). Non-tree-based networks are identified by locating a zigzag pattern that connects reticulate nodes (Zhang 2016).
- **Fold-unfold stable (FU-stable)**: A network that can be 'unfolded' into multi-labeled trees (MUL-trees) and then 'refolded' (see Huber and Moulton 2006) back into itself is FU-stable. Fold-unfold stability occurs when reticulate nodes are not stacked and when no two tree nodes have the same set of child nodes (Huber et al. 2016).
- **Tree-child**: A tree-child network is a phylogeny where all internal nodes have at least one non-reticulate child node (Cardona et al. 2008).
- **Normal**: tree-child networks that contain no directed 'shortcut' edges that directly connects two nodes (u, v) given that there exists another directed path from node u to v (Willson 2008, 2010b).

The latter four classes have the following nested hierarchy:

$$\textit{normal} \subset \textit{tree-child} \subset \textit{FU-stable} \subset \textit{tree-based}.$$

Any nested class will also have the topological properties of the all superceding classes (e.g., a tree-child network is also FU-stable and tree-based). Figure **??** gives positive and negative examples of the classifications, but also see Kong et al. (2022) and the webpage "ISIPhyNC (Information System on Inclusions of Phylogenetic Network Classes)"[a] for more information on each classification.

[a] http://phylnet.univ-mlv.fr/isiphync

2.2 Birth-Death-Hybridization Process

We use a birth-death-hybridization process (Morin and Moret 2006; Woodhams et al. 2016; Zhang et al. 2018) as a generating model for phylogenetic networks. This process has exponentially distributed waiting times between speciation, extinction, and hybridization events. We consider rates of speciation (λ) and extinction (μ) to be rates on each lineage, while the rate of hybridization (ν) is on each lineage pair. That is, when the process has N lineages, the rates of speciation, extinction, and hybridization are $N\lambda$, $N\mu$, and $\binom{N}{2}\nu$, respectively. There are three types of hybridization events: lineage generative hybridization, lineage neutral, and lineage degenerative (Box 1). We use the thinning property of a Poisson process to break the hybridization rate (ν) into separate rates for each type of hybridization: $\nu = \nu_+ + \nu_0 + \nu_-$, where the lineage generative hybridization rate is denoted as ν_+, lineage neutral hybridization as ν_0, and lineage degenerative hybridization as ν_-. We refer to the possible state space—all linear combinations of ν_+, ν_0, and ν_- for a

given hybridization rate (ν)—as the hybridization rate simplex, which we represent visually using a ternary plot (see example in Fig. 2).

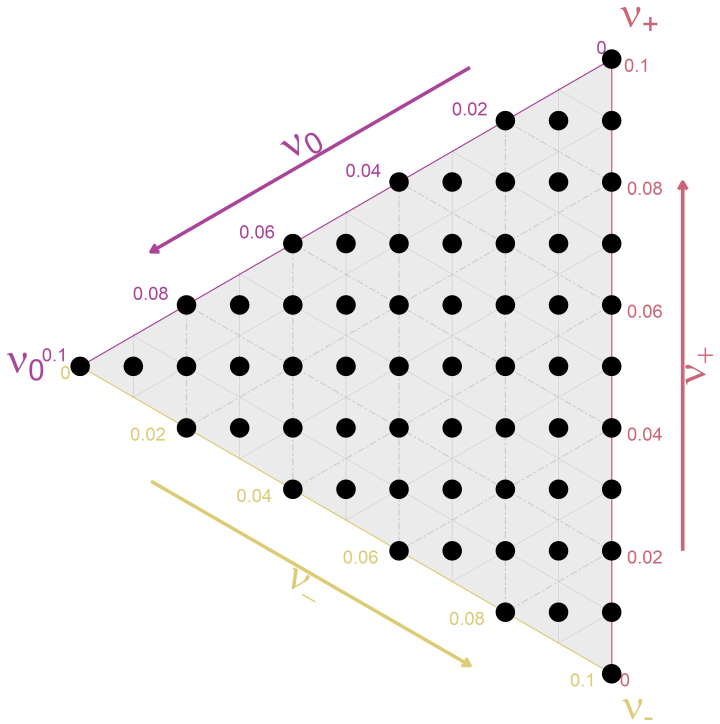

Figure 2: Ternary plot with the state space for the combination of hybridization-type rates for $\nu = 0.1$. Moving towards a corner corresponds to an increase of the rate for one of the hybridization types: the top corner (ν_+) denotes the lineage generative rate, the bottom corner (ν_-) denotes the lineage degenerative rate, and the leftmost corner (ν_0) denotes the lineage neutral rate. Each dot represents one parameter combination used when simulating across the hybrid rate simplex (Table S1). See Section 2.4.2 for details about the simulation procedure.

At a given hybridization event, we denote the genetic contribution of each parental lineage as γ and $1-\gamma$. Each hybridization event has its own associated γ that can be drawn from any distribution on $[0, 1]$. We draw the inheritance proportions γ from a $Beta(10, 10)$ distribution, creating a symmetric distribution centered on equal inheritance proportion of $\gamma = 0.5$.

In some of our simulations, hybridization events are more likely to occur if the the two parent lineages are closely related. To accomplish this, we used the approach of Woodhams et al. (2016) for genetic distance–dependent hybridization. Under this model, hybridization events are proposed with rate ν and successfully occur on the phylogeny with a probability that is proportional to the genetic distance between the two hybridizing taxa (d_{ij}):

$$Pr(Hybridization\ Success) \propto f(d_{ij}).$$

We assume a strict molecular clock where the genetic distance between two taxa is taken to be the summation of distinct branch lengths between each lineage and their most recent common ancestor. Moreover, in the case of hybrid taxa, the distance is a weighted summation across paths (see Woodhams et al. 2016).

There are many ways to model the relationship between genetic distance and the probability of hybridization (See Woodhams et al. 2016). For example, the Dobzhansky-Muller incompatibility model predicts that hybrid incompatibilities will accumulate at an accelerating rate (Orr and Turelli 2001). Here, we use a

linear decay function to relate the probability of hybridization success to genetic distance:

$$f(d_{ij}) = \begin{cases} 1 - d_{ij}\delta & d_{ij} \leq \frac{1}{\delta} \\ 0 & d_{ij} > \frac{1}{\delta} \end{cases},$$

where δ represents the strength of the genetic distance dependence. A value of $\delta = 0$ makes hybridization events independent of genetic distance, while higher values of δ make hybridization primarily occur between more closely related taxa.We can also see that, for genetic distances where d_{ij} is greater than the inverse of the dependence strength $\frac{1}{\delta}$, it is not possible for hybridization to occur.

Note that the generating process described above and our simulations differ slightly due to the stopping conditions we applied in our simulations. We focus on extant phylogenies during simulation and condition the process on survival of more than one taxon. Further, for practical purposes, we limit the run time of simulations due to the potential for some simulation replicates to never reach the specified ending age in finite time (see Section 2.4.2).

2.3 Diversification Dynamics

Measuring the rate of lineage accumulation over time allows us to compare the outcomes of different models. We examine how the diversification dynamics change both with the number of lineages and when we change our assumptions about hybridization to be either be primarily lineage generative or degenerative. Additionally, we show that, for certain parameter values of the birth-death-hybridization process, the process 'explodes,' reaching an infinite number of lineages in finite time.

2.3.1 Diversification Rate

In a non-reticulate birth-death process, the net diversification rate is simply $\lambda - \mu$ (e.g., Nee 2006; Morlon et al. 2010). However, under the birth-death-hybridization process, net diversification has a different interpretation since hybridization events also can cause the number of lineages to increase or decrease. Further, because hybridization rates scale quadratically with the number of lineages, net diversification cannot be computed as the difference between all lineage adding and lineage deleting rates, *i.e.,* $(\lambda + \nu_+) - (\mu + \nu_-)$. Instead we can compute the overall diversification rate for a process with N lineages as:

$$N\lambda - N\mu + \binom{N}{2}\nu_+ - \binom{N}{2}\nu_-, \tag{1}$$

or equivalently:

$$Nd + \binom{N}{2}s, \tag{2}$$

where $d = \lambda - \mu$ and $s = \nu_+ - \nu_-$. Thus, we interpret d as the per-lineage diversification rate and s as the lineage-pair diversification rate due to hybridization. The overall diversification rate is a function of the number of species but allows us to more directly compare the effects of speciation, extinction, and hybridization on lineage accumulation. From this perspective we can see that, as the number of lineages increases, hybridization quickly becomes the dominant force in diversification (Fig. 3a). The overall diversification rate grows rapidly when s is positive but actually becomes negative with a large number of taxa and when $s < 0$ (Fig. 3b). However, when considering a negative value of s and lineages through time, the overall diversification rate approaches zero as the negative effects from hybridization match the per-lineage diversification.

2.3.2 Explosion of the Birth-Death-Hybridization Process

Birth-death processes with sufficiently high birth rates can result in 'explosive' growth where an infinite number of events happen in finite time. We show that under certain conditions the birth-death-hybridization model exhibits 'explosive' growth. The rate of hybridization events scale with the total number of species pairs, creating a quadratic feedback term. Thus, when the rate of lineage generative hybridization exceeds the rate of lineage degenerative hybridization, new lineages tend to be added fast enough to induce explosion.

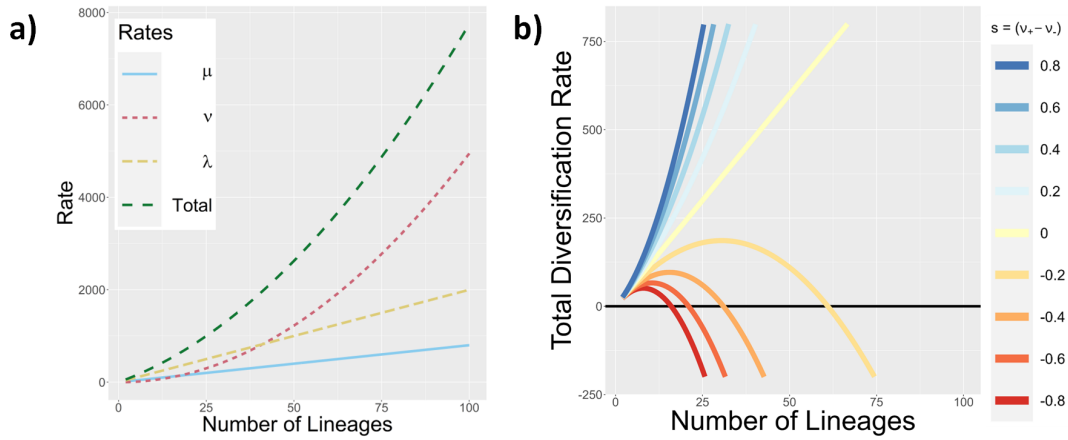

Figure 3: The effect of hybridization type on diversification dynamics. For all panels, we assume a speciation rate of $\lambda = 20$, extinction rate of $\mu = 8$, and a hybridization rate of $\nu = 1$. **a)** The rate of speciation(λ), extinction(μ), and hybridization(ν) events as the number of lineages increase. Panel **b)** shows the overall diversification rate $Nd + \binom{N}{2}s$ as a function of the number of lineages N.

Specifically, we show that the process explodes when $\lambda > 0$, $\mu \geq 0$, and $s > 0$ (*i.e.,* $\nu_+ > \nu_-$). If $\nu_+ > 0$ and $\nu_- = 0$, the birth-death-hybridization process becomes a special case of the birth-death process with combinatoric innovation and is known to explode in finite time (Steel et al. 2020). In this section we will assume that $\nu_- > 0$ and show that explosion occurs when $s > 0$. For now, we will also assume that $\lambda > \mu$. In the supplementary materials we will relax this assumption and provide a proof for explosion. Theorem 3.2.2 from Anderson (1992) (equivalently theorem 2.2 of Bansaye and Méléard 2015) states that a birth-death process almost surely has an infinite lifetime if and only if the following sum diverges

$$R = \sum_{n=1}^{\infty} \left(\frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_n \lambda_{n-1}} + \frac{\mu_n \mu_{n-1}}{\lambda_n \lambda_{n-1} \lambda_{n-2}} + ... + \frac{\mu_n ... \mu_2}{\lambda_n ... \lambda_2 \lambda_1} \right), \tag{3}$$

where λ_n and μ_n are the birth and death rates when the process has n lineages. Therefore, we can show the process almost surely has a finite lifetime (*i.e.,* explodes) if and only if the series instead converges. For the birth-death-hybridization process, both speciation and lineage generative hybridization add new lineages while extinction and lineage degenerative hybridization remove lineages, giving birth rate:

$$\lambda_n = n\lambda + \binom{n}{2}\nu_+ \tag{4}$$

and death rate:

$$\mu_n = n\mu + \binom{n}{2}\nu_-. \tag{5}$$

Defining $\sigma_n = \frac{\mu_n}{\lambda_n}$ gives:

$$R = \sum_{n=1}^{\infty} \left(\frac{1}{\lambda_n} + \frac{\sigma_n}{\lambda_{n-1}} + \frac{\sigma_n \sigma_{n-1}}{\lambda_{n-2}} + ... + \frac{\sigma_n ... \sigma_2}{\lambda_1} \right). \tag{6}$$

The sum can be written out as follows:

$$
\begin{array}{ccccccc}
\frac{1}{\lambda_1} & + & & & & & \\
\frac{1}{\lambda_2} & + & \frac{\sigma_2}{\lambda_1} & + & & & \\
\frac{1}{\lambda_3} & + & \frac{\sigma_3}{\lambda_2} & + & \frac{\sigma_3\sigma_2}{\lambda_1} & & \\
\vdots & & & & & \ddots & \\
\frac{1}{\lambda_n} & + & & ... & & + & \frac{\sigma_n...\sigma_2}{\lambda_1} \\
\vdots & & & & & & \ddots
\end{array}
\tag{7}
$$

and R can be rewritten to be the sum along each diagonal S_n, collecting each term with $\frac{1}{\lambda_n}$ into its own summation:

$$R = \sum_{n=1}^{\infty} S_n, \tag{8}$$

where:

$$S_n = \frac{1}{\lambda_n} + \frac{1}{\lambda_n} \sum_{i=n}^{\infty} \left(\prod_{j=n}^{i} \sigma_{j+1} \right). \tag{9}$$

We can show that S_n, and consequently R, converge by considering the sum $S'_n \geq S_n$:

$$S'_n = \frac{1}{\lambda_n} \sum_{i=0}^{\infty} (\sigma')^i, \tag{10}$$

where σ' is the supremum of σ_n when $n \geq 1$:

$$\sigma' = sup_{\{n \geq 1\}} \sigma_n. \tag{11}$$

Since $\lambda - \mu$ and s are positive, $\sigma_n < 1$ for all n. Consequently $\sigma' < 1$, making S'_n a geometric series that converges on $\frac{1}{\lambda_n} \frac{1}{1-\sigma'}$. We can then define R' to be the summation of all S'_n:

$$R' = \sum_{n=1}^{\infty} S'_n = \frac{1}{1-\sigma'} \sum_{n=1}^{\infty} \frac{1}{\lambda_n}. \tag{12}$$

Here $R' < +\infty$ because λ_n grows quadratically with n when $\nu_+ > 0$. By direct comparison, we get $R < R'$. We can then conclude that $R < +\infty$, meaning that the process can explode in finite time.

2.4 Birth-Death-Hybridization Process Simulations

In this study, we sought out to assess how different assumptions about hybridization affect the distribution of phylogenetic networks under a birth-death-hybridization process. We simulated both the number of lineages over time and phylogenetic networks to evaluate properties and expectations that are difficult to compute analytically under a birth-death-hybridization process. All simulations were conducted on a Dell XPS 15 with an i7 processor.

2.4.1 Simulating Lineage Diversification over Time

To understand the accumulation of lineages under the birth-death-hybridization process, we simulated the process over time using the R package GillespieSSA (avaliable on CRAN; Pineda-Krch 2008). This package

allows for efficient simulation of the number of lineages in two ways: (1) it can simulate under the birth-death-hybridization process without having to simulate and record the structure of the phylogeny and (2) each simulation records the number of lineages at multiple time points, not just at the end of the simulation. All simulations began with two lineages at $t = 0$ and were allowed to proceed until reaching the stopping time $t = 10$. We used constant rates of speciation ($\lambda = 4$), extinction ($\mu = 3$), hybridization ($\nu = 0.2$), and assumed no genetic distance dependence ($\delta = 0$). Since lineage neutral hybridization does not change the number of lineages in the process, we assumed a lineage neutral hybridization rate of zero ($\nu_0 = 0$). To change the lineage-pair diversification rate s, we modulated the values of lineage generative and degenerative hybridization between simulations to assess how different lineage-pair diversification rates s change the diversification dynamics. Since positive values of s can lead to an infinite number of lineages (see Section 2.3.2), we limited each simulation to run for 4 seconds. For each value of s, we simulated until we collected 5,000 replicates that survived until $t = 10$ or 5000 replicates that reached the timeout time of 4 seconds.

2.4.2 Simulating the Distribution of Phylogenetic Network Classes

Phylogenies were simulated using the R package SiPhyNetwork (avaliable on CRAN; Justison et al. 2023). Simulations were conditioned on survival to the present with at least two extant lineages prior to sampling and to have completed in less than 10 seconds. The latter condition is to prevent simulation replicates from spending an infinite amount of time approaching the stopping time without ever reaching it. We expect this behavior for positive lineage-pair diversification rates ($s > 0$).

All simulations began with two lineages at $t = 0$, representing the most recent common ancestor at the root, and were allowed to proceed until reaching the stopping time at $t = 1$. We used constant rates of speciation ($\lambda = 4$), extinction ($\mu = 2$), and hybridization ($\nu = 0.1$). For each specific scenario and parameter combination (Tables S2, S3, and S4) we generated 20,000 replicate phylogenetic networks.

2.4.3 Simulating Across the Hybridization-Rate Simplex

To assess how hybridization type affected the properties of simulated phylogenies, we used 61 different linear combinations of rates for the hybrid types ν_+, ν_0, and ν_-. The specific parameter combinations are visualized in Fig. 2 and listed in Table S1.

For each combination of hybridization-type rates, we used two different sampling strategies of extant-only sampling and complete sampling. First, we considered complete phylogenies with all extant and extinct taxa included in the phylogeny. For these complete phylogenies we had two scenarios where hybridization was either genetic distance–dependent ($\delta = 0.5$) or independent of genetic distance ($\delta = 0$). Next, we sampled extant-only phylogenies by pruning extinct taxa from the phylogeny. For the extant-only phylogenies, we considered both complete extant sampling ($sampling\ fraction = 1$) and incomplete sampling where a constant fraction of extant taxa were sampled ($sampling\ fraction = 0.75$).

We additionally simulated phylogenies to isolate the effects of genetic-distance dependence and incomplete sampling on the class membership of networks. We used a constant rate for each hybridization type ($\nu_+ = \nu_0 = \nu_- = \frac{1}{30}$) while separately varying the strength of genetic-distance dependence and amount of incomplete lineage sampling. For incomplete lineage sampling, we simulated extant only phylogenies and used sampling fractions that ranged from $\rho = 0.25$ to $\rho = 1.0$. For genetic distance dependence, we simulated complete phylogenies with both extinct and extant taxa while varying the strength of distance dependence from $\delta = 0$ to $\delta = 2$.

2.5 Analysis of Simulated Data

For each replicate of the lineage diversification simulations, we recorded the number of lineages through time. We then reported the average number of lineages over time for each lineage-pair diversification setting for s when conditioning on survival at each time point. We analyzed the simulated phylogenetic networks using a combination of summary statistics and network classifications to characterize the distribution of phylogenetic networks under the birth-death-hybridization process. For the replicates of each simulation scenario, we recorded the total number of taxa, total number of hybridization events, reticulation density ($\frac{Number\ of\ Hybridizations}{Number\ of\ Tips}$), observed proportion of each hybridization type, and the level of the network.

Lastly, we assessed whether phylogenies belonged to the tree-based, FU-stable, tree-child, or normal classes (Box 2).

3 Results

Our aim—through both analytical computation and simulation—was to understand and characterize the properties of phylogenetic networks generated from the birth-death-hybridization process. We also assessed how our assumptions about gene flow and sampling affected this distribution. We present and outline our key findings below.

3.1 Lineage Diversification over Time

The shape of lineage accumulation took different forms depending on the relative rates of lineage-generative and lineage-degenerative hybridization (Fig. 4). When $\nu_+ > \nu_-$, the lineage-pair diversification was positive, and the average number of lineages exhibited hyper-exponential growth. In fact, all simulation settings with $s > 0$ reached 5,000 timeouts before reaching 5,000 surviving simulations. This result was consistent with the explosive behavior that can occur when $s > 0$, where too many events—possibly infinitely many—would occur before the simulation could finish. When conditionined on survival, $\nu_+ < \nu_-$ resulted in a negative value for s, the average number of lineages approached a constant number (Fig. 4). The effects of positive diversification via speciation and extinction (*i.e.,* Nd) were neutralized by negative diversification due to hybridization (*i.e.,* $\binom{N}{2}$) and resulted in a net-zero total diversification (see Eq. 2). Interestingly, when not conditioned on survival and $s < 0$, the average number of lineages decreased over time (Fig. S2). Any time the process reached a large enough population, lineage degenerative hybridization acted a diversity dependent mechanism that inhibited growth. Then over a long enough period of time, the process would eventually run to extinction due to the stochastic nature of events, causing the decline in the average number of lineages.

3.2 Phylogenetic Network Simulation

We generated 20,000 phylogenies that met the criteria defined in Section 2.4.2 for each parameter combination (Tables S3 and S4). Across all completely sampled simulations without genetic-distance dependence, we discarded 339,054 simulation replicates (21.7% of total attempts) that went extinct or yielded only one extant taxon. We discarded 2,233 replicates (0.14% of total attempts) due to the simulation exceeding the 10 second stopping condition. A higher proportion of simulations timed out as s increased. Increasing the stopping time to 20 seconds for the settings with the highest lineage-pair diversification (s) did not have a large affect on the number timeouts (1,074 timeouts; 0.13% of simulations), indicating that we were likely able to capture the distribution of networks with little bias. Simulations rejected for timing out may have never finished or were outliers in terms of numbers of extant lineages.

3.3 Simulating Across the Hybrid Rate Simplex

Hybridization type affected the diversification dynamics of simulated phylogenies (Fig. 5). In accordance with analytical expectations, simulations with higher lineage-pair diversification rates (s) had an increased number of taxa on average, a result that is consistent across all sampling scenarios (complete, complete with genetic distance–dependent hybridization, extant-only networks, and extant-only networks with incomplete sampling). We also observed the same increasing pattern with respect to s for the number of hybridization events and reticulation density. Interestingly, however, extant-only networks had higher reticulation densities than complete networks despite having fewer reticulations (Fig. 5; Table S5). Though all extinct lineages are not observed for extant-only sampling, reticulations are only lost on the phylogeny when all the hybrid descendants fail to get sampled. This becomes increasingly difficult for highly intertwined networks with many reticulations. Further, due to incomplete sampling among some parental lineages, extant-only networks had a higher proportion of lineage-degenerative and lineage-neutral hybridization than completely sampled networks with the same hybridization rates (Fig. S3).

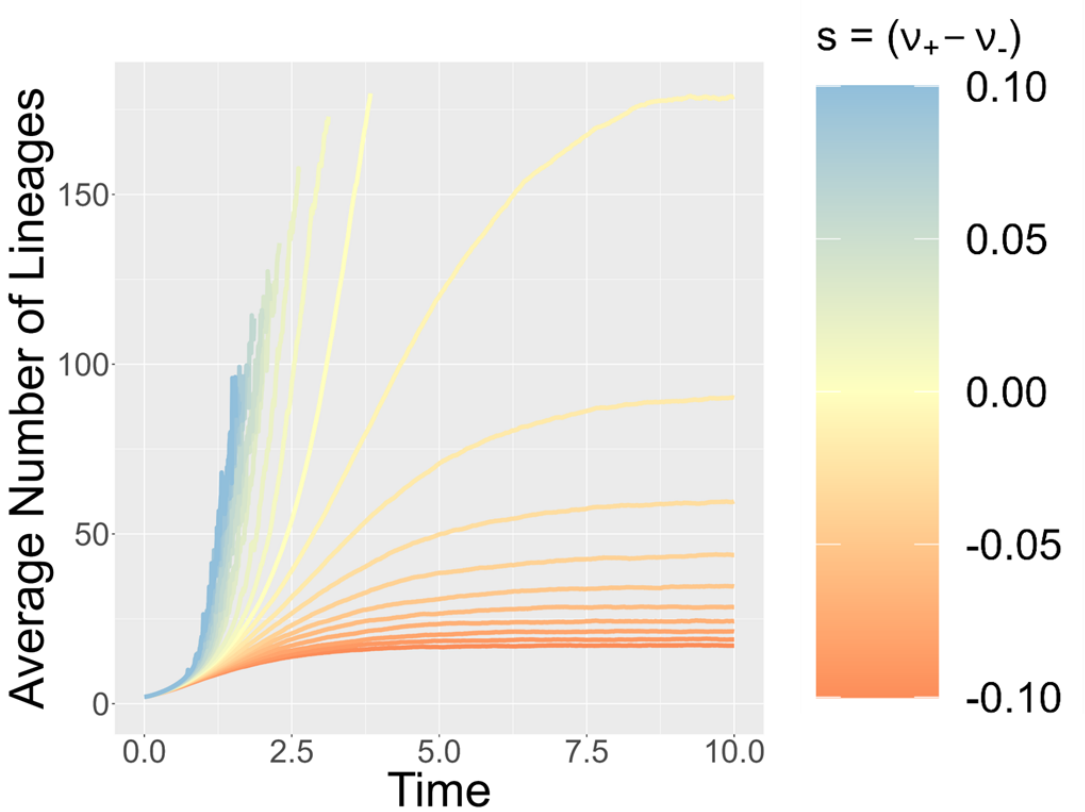

Figure 4: The average number of lineages through time with respect to s when conditioned on survival. We assume a speciation rate of $\lambda = 4$, extinction rate of $\mu = 3$, a hybridization rate of $\nu = 0.2$, and a lineage neutral hybridization rate of $\nu_0 = 0$. Higher lineage-pair diversification (s) corresponds to a darker blue while low values of s are denoted with a dark orange. Lines were only drawn for time points that had at least 4,500 replicates that had survived to that time point and had not timed out.

The proportion of networks belonging to each class was differentially affected across the hybrid rate simplex (Fig. 6; Table S6). The more deeply contained classes within the nested hierarchy of tree-based, FU-stable, tree-child, and normal classes (see Box 2) have increasingly stringent topological requirements and, consequently, lower class proportions (e.g., a smaller proportion of networks were tree-child than tree-based). More deeply nested classes also showed a greater range of proportions across the hybrid simplex (Table S6). Notably, for completely sampled, extant-only, and incompletely sampled datasets, the region with high lineage-generative rates (*i.e.,* high s) led to the lowest proportion of tree-based and FU-stable networks but resulted in relatively high proportions for tree-child and normal classes, thus indicating tree-based and FU-stable classes could be more sensitive to a high number of hybridization events than other classes. However, for genetic distance–dependent hybridization, the proportion of tree-based and FU-stable networks tended to decrease as the value of s decreased. Furthermore, for all simulation scenarios, the proportion of tree-child and normal networks decreased in regions with fewer hybridization events that consequently also had an increased proportion of lineage degenerative hybridizations (*i.e.,* lower s), suggesting hybridization type is and important factor for these classes. When comparing between sampling conditions, class proportions were lower for the incompletely sampled datasets than those that were completely sampled (despite having fewer hybridization events). The complete dataset simulated under genetic distance–dependent hybridization had relatively few reticulation events, resulting in overall higher class proportions than other sampling conditions.

Simulation replicates with only lineage-generative and neutral hybridization always produced networks with certain topological properties (dark diamonds in Fig. 6; Table S6). For completely sampled networks, all simulations with a lineage-degenerative rate of zero ($\nu_- = 0$; axis with ν_0) were tree-child. Similarly, at

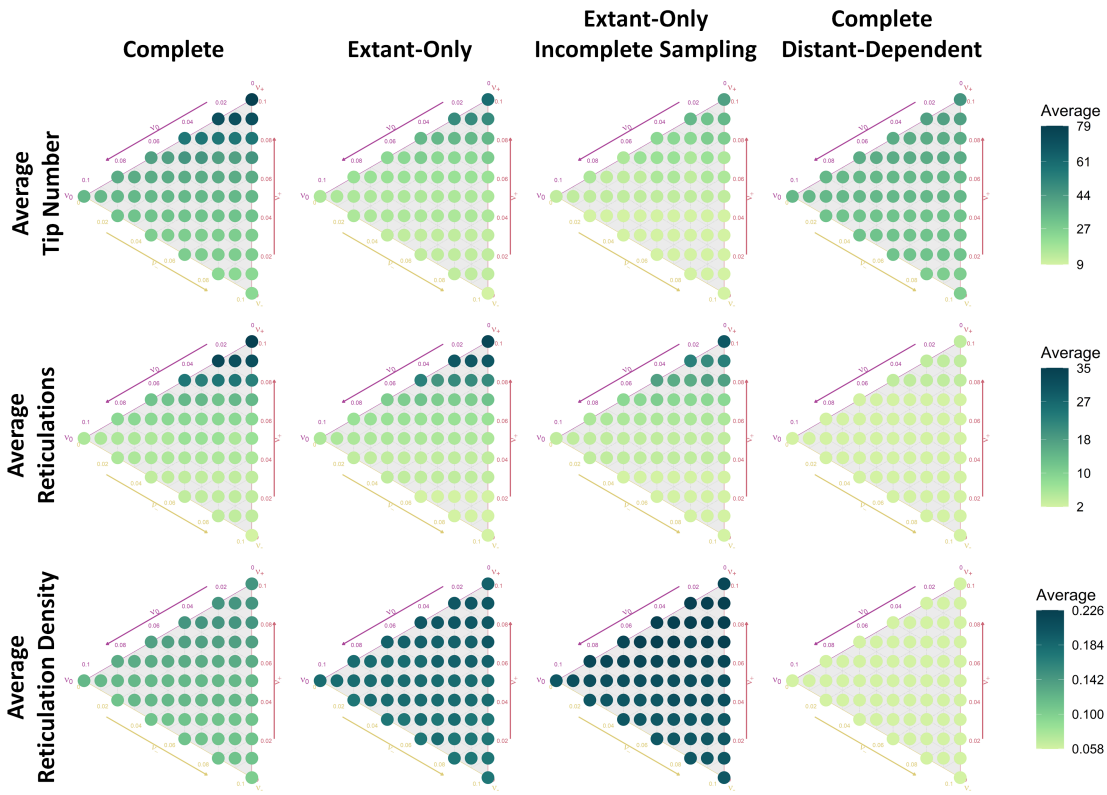

Figure 5: Diversification dynamics of phylogenetic networks simulated across the hybrid rate simplex. Each column shows one of the four simulation conditions used to simulate phylogenetic networks across the hybrid rate simplex (Section 2.4.2). The position of each point on the simplex indicates the lineage generative, lineage neutral, and lineage degenerative hybridization rates. Each row summarizes the number of tips, reticulations, and reticulation density ($\frac{reticulations}{tips}$) for each dataset. The color of each point is used to depict the average value of 20,000 replicates.

the corner of the hybrid rate simplex with only lineage generative hybridization ($\nu_+ = \nu = 0.1$), all networks were tree-based, FU-stable, tree-child, and normal. Interestingly, the same parameterizations with extant-only or incomplete sampling did not lead to unity with respect to class membership. Unsampled parental lineages caused lineage generative and lineage neutral hybridizations to be observed as another type (e.g., Figures S1 and S3).

A large portion of reticulate phylogenies were not level-1 (Fig. S4). Increasing the lineage-pair diversification rate (s) led to an increase in the average level of simulated networks and a higher proportion of level-2 (or higher) networks. Trends in the network level were strongly tied to the number of reticulations (Fig. 7). In most cases, the level of the network was the same as number of reticulations. However, there was slightly more variation in network level for those with few reticulations (Fig. 7b).

Increasing the effects of genetic-distance dependence and incomplete sampling decreased the number of reticulations on networks and consequently increased the class proportions of networks (Fig. 8). Increasing the effect of incomplete sampling (*i.e.,* decreasing the sampling fraction) only modestly affected the class proportions. However, increasing the effect of distance dependence (*i.e.,* strength of distance dependence δ) more drastically increased class proportions, making nearly all networks tree-based and FU-stable. Higher class proportions can be explained by large decreases in the overall number of hybridization events and an increase in the proportion of phylogenies with no reticulation events, particularly for distance-dependent hybridization (Figures S6 and S7). For the incomplete sampling scenario, higher reticulation densities (Fig. S6) and changes in the observed hybridization types (Fig. S5) likely offset some of the effects of having fewer reticulations, explaining a more limited increase in class proportions. Interestingly, when looking on a per-reticulation basis, the opposite trend is observed for incomplete sampling. Decreasing sampling actually decreased class proportions for a given number of reticulations (Fig. 9). When looking on

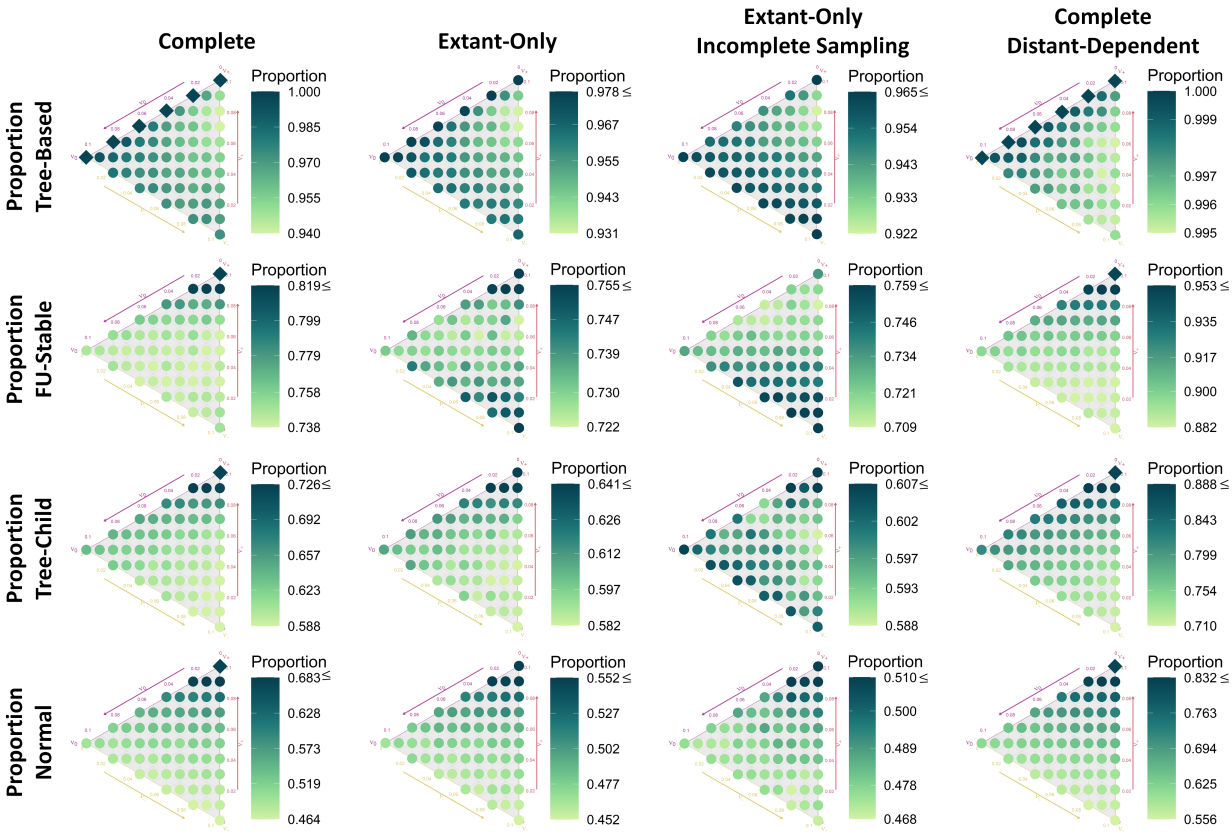

Figure 6: Class membership of simulated phylogenetic networks across the hybrid rate type simplex. Each column shows the simulation conditions used to simulate phylogenetic networks across the hybrid rate simplex (Section 2.4.2). The position of each point on the simplex indicates the lineage generative, lineage neutral, and lineage degenerative hybridization rates. Each row summarizes the proportion of phylogenies from a given class. The color of each point is used to depict the proportion of 20,000 replicates that belonged to the class, with a darker color representing a higher proportion. Colors are scaled for each class and dataset. Scales range from the minimum observed proportion to the 95th percentile, and all values above the 95th percentile are binned into the same color. Cases where all simulated networks belong to a class (proportion of 1) are indicated with a diamond.

a per-reticulation basis for genetic distance–dependent hybridization, increased dependence strength (δ) is associated with higher proportions of tree-based and FU-stable networks but lower proportions of tree-child and normal networks.

4 Discussion

4.1 Diversification under Birth-Death-Hybridization

The birth-death-hybridization process is related to the logistic branching process (Lambert 2005) in that both allow for density-dependent feedback terms that affect growth; however, the birth-death-hybridization process allows for both positive and negative feedback based on the dominant type of hybridization. With the rate scaling by each lineage pair, hybridization can eclipse rates of speciation and extinction, resulting in highly reticulate histories. However, in nature, some biological systems exhibit wide-spread gene flow (Salzburger 2018; Edelman et al. 2019; Kozak et al. 2021; Esquerré et al. 2022; Suvorov et al. 2022), while it is much more limited in others (Solís-Lemus and Ané 2016; Morales-Briones et al. 2018; Zhang et al. 2018; Karimi et al. 2020). The lack of extensive reticulation in some empirical systems may be an artifact of the computational burden associated with estimating complex phylogenetic networks, particularly when applied to genome-scale data (Hejase and Liu 2016; Elworth et al. 2019). Empirical datasets may also result in a

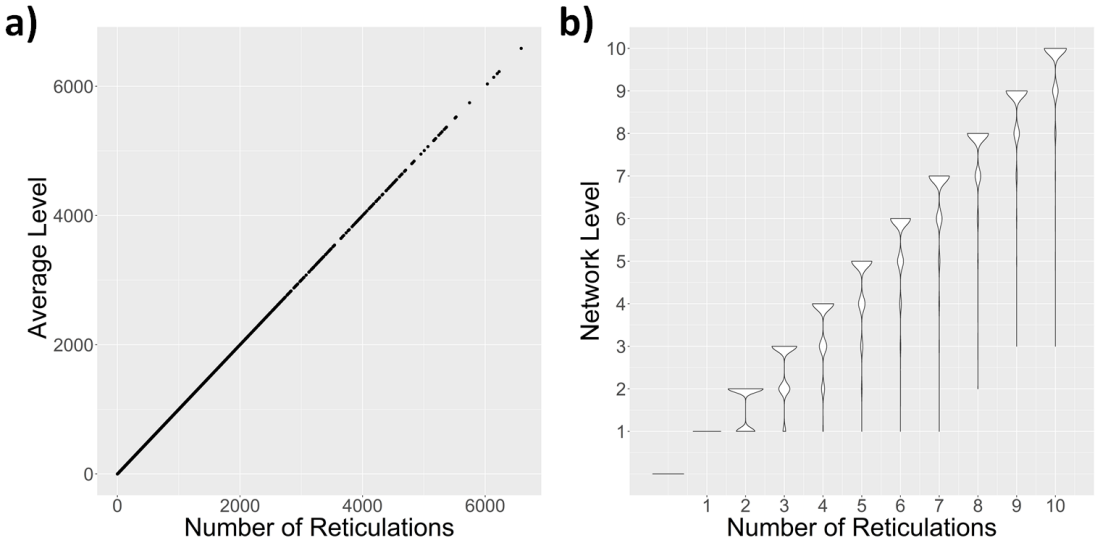

Figure 7: Completely sampled phylogenies summarized across the hybrid rate simplex by the number of reticulations. **a)** A scatterplot depicting a near one to one relationship between the number of reticulations and the average level of simulated phylogenetic networks. **b)** A violin plot showing the distribution of levels for 0 to 10 reticulations. Widths of the violin correspond to the proportion of networks with that level.

small number of inferred reticulation events due to the build-up of incompatibilities as lineages genetically diverge over time (Orr 1995; Mallet 2005; Gourbière and Mallet 2010). In fact, we showed that a linearly decreasing reticulation dependence on genetic distance had great propensity to reduce the number of events on simulated phylogenetic networks. However, there are a number of isolating mechanisms (see Mallet 2005; Soltis and Soltis 2009) that could create heterogeneity in the amount gene flow from system to system. Future work explicitly incorporating these mechanisms and modeling other genetic-distance dependencies into the birth-death-hybridization processes will be important in assessing their role in the diversification of admixed systems.

Solely changing assumptions about hybridization leads to phylogenies with considerable variation in size and pattern of diversification. The slowdown of lineage accumulation has been observed in numerous clades across the tree of life (Morlon et al. 2010). Several mechanisms have been proposed to explain slowing diversification (Moen and Morlon 2014), including density-dependent speciation rates (Rabosky and Lovette 2008) and protracted speciation (Etienne and Rosindell 2012). Under a birth-death-hybridization model where $s < 0$, hybridization acts as a diversity-dependent mechanism that reduces the overall degree of diversification and generates a pattern of slowed growth over time. In fact, estimating the likelihood of the phylogenetic network from a branching process is currently limited to a special case of the birth-death-hybridization process where only lineage-degenerative hybridization occurs ($s < 0$; Zhang et al. 2018). Given that gene flow represents a broad set of processes with heterogeneous effects on diversification, it will be important to be able to estimate rates for each type of hybridization from the birth-death-hybridization process. Although available methods do not currently include extinction or estimate lineage neutral or lineage generative rates, for systems with lineage-generative reticulation, the hyper-exponential growth when $s > 0$ might prove helpful for estimating diversification dynamics. If these systems have ages calibrated by fossil or other geological data, it would be sensible to limit the parameter search space to regions of the hybrid-rate simplex that are not expected to explode to infinitely many species over the age of the clade or to invoke other biological processes that might limit such explosive growth.

4.2 Topological properties of Phylogenetic Networks

Two key results arise from previous simulation studies of topological profiles (Arenas et al. 2008; Janssen and Liu 2021): (1) as the number of reticulations increases, the proportion of generated networks that belong to certain classes (e.g., tree-child, tree-sibling) decreases and (2) when controlling for the number of reticulations,

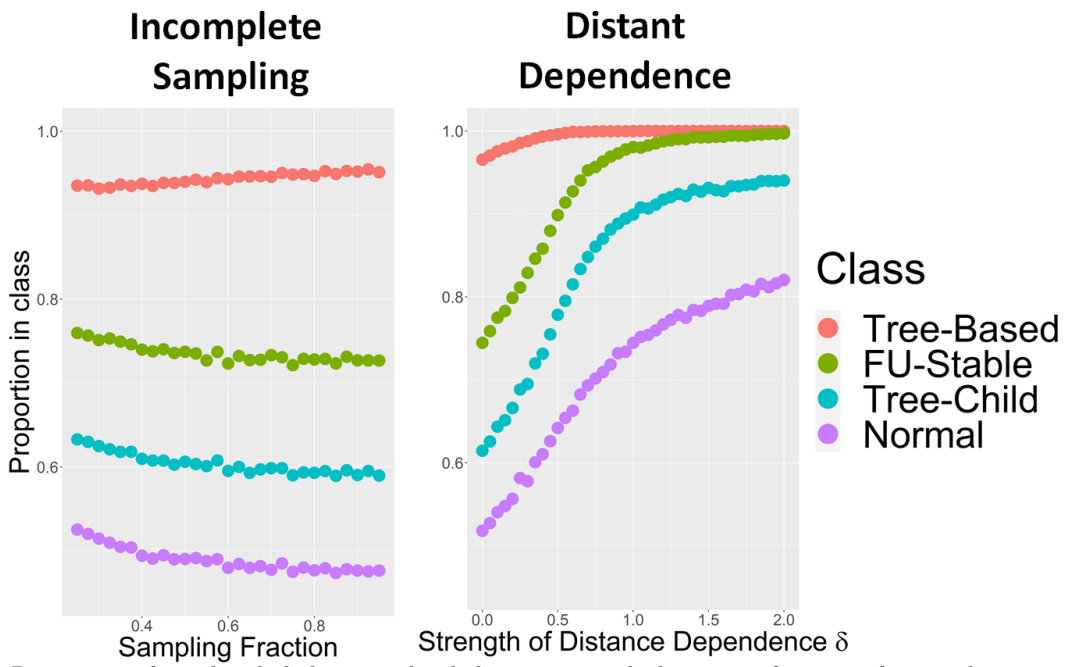

Figure 8: Proportion of simulated phylogenies that belong to network classes as a function of incomplete sampling and genetic distance dependence. Each point represents the class proportions observed from 20,000 simulated phylogenetic networks. The sampling fraction used in the incomplete sampling dataset represents the number of extant tips that are randomly sampled from the network. The value of δ used in the genetic distance-dependant dataset corresponds to the strength of dependence, with higher values of δ indicating that reticulation events primarily only occur between closely related lineages. Extant-only networks were used for the incomplete sampling dataset, while complete phylogenetic networks with extant and extinct species were used for the genetic distance-dependence dataset.

the type of reticulation affects the proportion of networks belonging to the same given classes. Our results are largely concordant with these conclusions. Though, when considering the birth-death-hybridization process, it creates a complex interaction of these factors; the reticulate events that are the best for some phylogenetic network classes (lineage generative; Janssen and Liu 2021) also lead to more lineages and consequently, more reticulations. We found that phylogenetic-network classes responded differently to these factors; tree-based and FU-stable network proportions were lower in the region of the hybrid rate simplex with a high number reticulations, primarily being lineage generative. Tree-child and normal classes had low proportions in the region that had few reticulation events but were mostly lineage degenerative. Thus, it is important to not only consider how each hybridization type affects the diversification and topological properties of systems but also to characterize how each network class individually responds to these factors.

Many network classifications require reticulations to be distant from one another. For example, networks will not be tree-child if reticulations are stacked, or networks are not tree-based if they have a specific zig-zagging pattern that connects reticulations (Zhang 2016)—both only occur if reticulations have few nodes separating them on the phylogenetic network. Reticulation proximity makes it important to consider not only the number of hybridizations when considering topological properties but also their density with respect to the size a phylogenetic network. Both poor taxon sampling and not accounting for extinct lineages with fossil evidence has been shown to reduce the accuracy of phylogenetic inference (Heath et al. 2008a; Nabhan and Sarkar 2012; Warnock et al. 2020). Incomplete species sampling may also pose a challenge for estimating phylogenetic networks as the higher reticulation density of poorly sampled networks compared to those that are completely sampled may induce fewer topological properties that can inform phylogenetic estimates.

Both our simulations and those of Janssen and Liu (2021) with genetic distance dependence also highlight the importance of reticulation proximity. Gene flow reduces the observed divergence between species (Slatkin 1985; Leaché et al. 2014). When considering a fixed number of reticulation events, this creates a positive feedback loop for gene-flow events to occur locally on reticulate lineages, thus decreasing the proportion of networks with certain topological properties. However, these same distance-dependent mechanisms

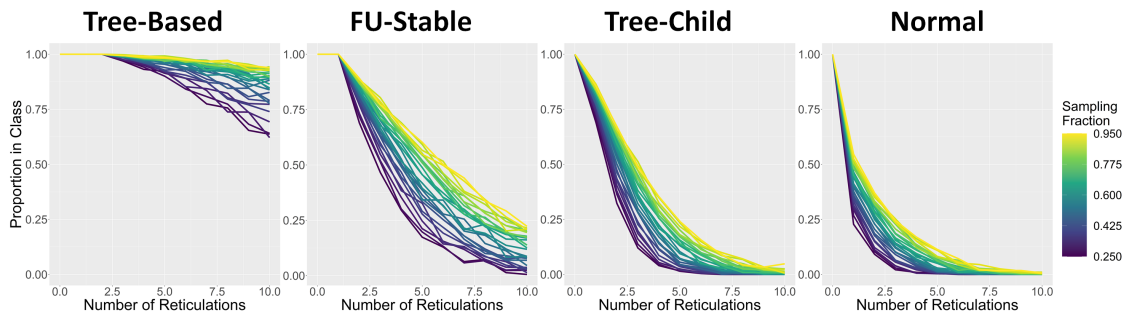

Figure 9: Class membership proportions under incomplete sampling as a function of reticulation number. Replicates are grouped and summarized by the number of reticulations and sampling fraction. The sampling fraction is the proportion of extant lineages sampled from an extant-only phylogenetic network. For each sampling fraction, 20,000 phylogenetic networks were simulated. The proportion in a class represents the proportion of networks with a given number of reticulation events to belong to that class. Points were only drawn when a given distance-dependence strength and number of reticulations had at least 30 replicates.

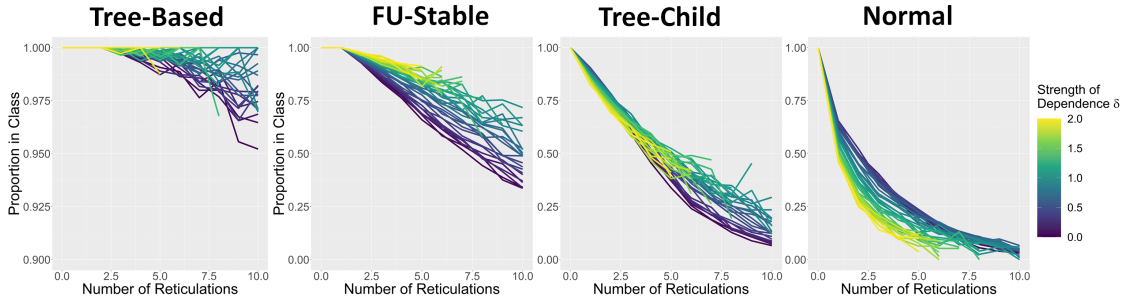

Figure 10: Class membership proportions under genetic-distance dependence and as a function of reticulation number. Phylogenetic networks are grouped by the simulated number of reticulations and strength of distance dependence. For each value of distance dependence (δ), 20,000 phylogenetic networks were simulated. The proportion in a class represents the proportion of networks with a given number of reticulation events to belong to that class.

also reduce the overall number of gene-flow events, potentially mitigating the effects of increased locality. Further, we only considered a scenario where the hybridization rate decreases linearly with genetic distance, which has two potentially important characteristics: (1) the rate of hybridization may decrease slower than other biologically relevant mechanisms for modeling genetic-distance dependence (e.g., exponential or snowballing decay), and (2) after a certain threshold, the probability of two distant lineages hybridizing becomes zero. The former attribute may lead to excessive gene flow, which, due to the homogenizing effect of gene flow, would create a positive feedback loop that keeps the genetic distances between lineages low, resulting in more gene flow. The latter attribute effectively makes distant clades independent with respect to hybridization; this may be an undesirable effect if one is trying to model gene flow between distantly related lineages. Ultimately, if systems with distance-dependent gene flow (e.g., Chapman and Burke 2007; Tea et al. 2020; Barley et al. 2022) have an abundance of events, it may pose challenges to methods that make certain topological assumptions of the underlying phylogenetic network.

The proximity of reticulations to one another also explains why only lineage-generative or neutral hybridization (*i.e.,* $\nu_- = 0$) always produced networks of certain types and why incomplete or extant-only sampling eliminates this pattern (see lemmas 1-4 in Janssen and Liu 2021). Effectively, lineage-generative and neutral hybridization events create extra tree nodes during the reticulation event. These nodes can put space between reticulation nodes, making it more likely that a network meets certain topological criteria for class assignment.

While not all reticulate histories are tree-based and have an underlying tree as a backbone (Francis and Steel 2015), those generated from the our parameterizations of birth-death-hybridization process appear largely tree-based. The tree-based class is conceptually satisfying for phylogenetic networks by allowing

researchers to think of reticulate evolution as phylogenetic trees with additional arcs. In fact, tree inference on reticulate systems has several uses and advantages over network inference (see Blair and Ané 2020). Though, inferring the treelike backbone of a network—if such a tree exists—should be done with caution because topology and divergence estimates can be biased (Leaché et al. 2014). Interestingly however, several recent methods attempt to quickly estimate phylogenetic networks by starting with phylogenetic trees and augmenting them into networks (Cao et al. 2019; Molloy et al. 2021).

Some relationships on phylogenetic trees are even impossible to infer (Slowinski 2001; Lewis et al. 2005), particularly for systems undergoing rapid speciation (Stanley et al. 2011; Suh 2016). Instead, polytomies can be included in an analysis to reflect the lack of phylogenetic resolution (Lewis et al. 2005; Kemp 2009). A similar discourse is occurring in the phylogenetic-network community with many studies finding only specific topologies identifiable and possible to estimate from gene tree topologies (Huber et al. 2015; Pardi and Scornavacca 2015; Solís-Lemus and Ané 2016; Francis and Moulton 2018; Gross and Long 2018; Erdős et al. 2019). Even when sequence alignments are used to alleviate identifiability issues, scalability and computational complexity remains a challenge for phylogenetic network inference. Pardi and Scornavacca (2015) proposed an interesting solution to the identifiability—and potentially scalability—problem: infer distinguishable canonical phylogenetic networks by effectively collapsing some reticulate events into polytomies. There are techniques capable of reconstructing the phylogenetic network if it is of a certain class and returns a reduced form network if not belonging to that class. Normal networks can be reconstructed in polynomial time from their set of displayed trees (Willson 2010a), and if the underlying network is not normal then the algorithm will return a reduced form normal network. Not all reticulations may be present in the reduced normal network, but these constructions have desirable mathematical properties and summarize the history into a simpler structure (Francis et al. 2021). Additionally, the folding operation of Huber and Moulton (2006) can reconstruct FU-stable networks from MUL-trees (Huber et al. 2016). If the underlying network is FU-stable then the method returns the true network, otherwise it produces a minimally reticulate network that unfolds into the same MUL-tree. Although estimating the MUL-tree is not a trivial task (Huber et al. 2008), MUL-trees and their implied FU-stable networks have been used to understand polyploid histories (Brysting et al. 2007). These types of approaches represent a shift from the fully resolved phylogenies that biologists have come to expect, but they have several appealing qualities: (1) they only return what is distinguishable, (2) if the true network has the right topological properties then it can still be recovered, and (3) canonical forms have a reduced topology search space with computationally efficient construction algorithms. Further, these approaches will still be applicable to the appreciable proportion of networks under the birth-death-hybridization process that do not have ideal topological properties or meet strict assumptions.

Work towards characterizing and understanding the identifiability of level-1 networks (Solís-Lemus and Ané 2016; Solis-Lemus et al. 2020; Gross et al. 2021; Allman et al. 2022) has led to many reticulate inference methods assuming an underlying level-1 network (Huber et al. 2010; Solís-Lemus and Ané 2016; Allman et al. 2019; LeMay et al. 2021). These methods have surged in popularity recently, largely due to their computational tractability and—for `SNaQ` (Solís-Lemus and Ané 2016) and `NANUQ` (Allman et al. 2019)—their ability to account for discordance both from gene flow and incomplete lineage sorting. However, the level-1 assumption was routinely violated for networks generated under the birth-death-hybridization process, particularly for highly reticulate networks. Fortunately, in practice, methods that make this assumption are often used to either infer few reticulate events (e.g., Morales-Briones et al. 2018; Karimi et al. 2020; Myers et al. 2022), or researchers assume that gene flow events are primarily isolated within each major clade (e.g., Esquerré et al. 2022)—with both cases being more likely to have level-1 networks. Further, recent identifiability results for level-2 networks (Van Iersel et al. 2009; van Iersel et al. 2020, 2022) may prove useful in extending the level-1 limitation of some methods. Nonetheless, it will be important to assess how robust these methods are to model violations and characterize possible biases when estimating networks that are not level-1.

5 Conclusions

In this study, we have shown that the specific assumptions about reticulation and sampling can have important effects on distribution of generated phylogenetic networks from the birth-death-hybridization process.

There are many mechanisms that leave a specific reticulate pattern (see table in Box 1); both the biological process and the sampling framework affect how reticulate patterns (lineage generative, degenerative, neutral) manifest on phylogenetic networks. Methods that can directly investigate which type of reticulation occurs on phylogenetic networks (Hibbins and Hahn 2019; Flouri et al. 2020) will be vital as we infer patterns of gene flow. Further, system-specific knowledge—factors like life history (Wolf et al. 2001; Rumpho et al. 2008; Wendel et al. 2009; Schulte et al. 2012; Montanari et al. 2016; Meier et al. 2019), geographic distributions (López-Caamal et al. 2014; Ottenburghs et al. 2017; Dolinay et al. 2021), and genomic architecture (Wendel and Cronn 2003; Ottenburghs et al. 2017; Edelman et al. 2019)—should be incorporated to help inform how to model and interpret each type of reticulation. The specific reticulation types affect diversification and topology of phylogenetic networks, potentially violating assumptions of some methods or biasing diversification rate estimates if not taken into account. Overall, our work highlights the importance of thinking carefully and deliberately about how we model gene flow with the birth-death-hybridization process.

6 Additional

6.1 Acknowledgements

We would like to thank members of the Heath Lab for helpful comments and feedback on the manuscript. We would also like to thank Huw Ogilvie and two anonymous reviewers for helpful comments and observations. Lastly, we would like to thank Claudia Solis-Lemus and George Tiley for inviting us to be a part of the special issue.

6.2 Data Availability Statement

All scripts used to simulate data, perform analyses, and create figures are available at: `https://github.com/jjustison/BDH_simulation`. All simulated data and supplemental material are available on Zenodo at: `https://doi.org/10.5281/zenodo.8371004`

6.3 References

References

Allman ES, Baños H, Rhodes JA. 2022. Identifiability of species network topologies from genomic sequences using the logdet distance. Journal of Mathematical Biology. 84:1–38. doi:https://doi.org/10.1007/s00285-022-01734-2.

Allman ES, Baños H, Rhodes JA. 2019. NANUQ: a method for inferring species networks from gene trees under the coalescent model. Algorithms for Molecular Biology. 14:1–25. doi:https://doi.org/10.1186/s13015-019-0159-2.

Anderson WJ. 1992. Continuous-time Markov chains: An applications-oriented approach. Springer Science & Business Media. doi:https://doi.org/10.1007/978-1-4612-3038-0.

Arenas M, Valiente G, Posada D. 2008. Characterization of reticulate networks based on the coalescent with recombination. Molecular Biology and Evolution. 25:2517–2520. doi:https://doi.org/10.1093/molbev/msn219.

Bansaye V, Méléard S. 2015. Birth and death processes. Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior. :7–17doi:https://doi.org/10.1007/978-3-319-21711-6_2.

Barley AJ, Nieto-Montes de Oca A, Manríquez-Morán NL, Thomson RC. 2022. The evolutionary network of whiptail lizards reveals predictable outcomes of hybridization. Science. 377:773–777. doi:https://doi.org/10.1126/science.abn1593.

Betat H, Mede T, Tretbar S, Steiner L, Stadler PF, Mörl M, Prohaska SJ. 2015. The ancestor of modern Holozoa acquired the CCA-adding enzyme from Alphaproteobacteria by horizontal gene transfer. Nucleic Acids Research. 43:6739–6746. doi:https://doi.org/10.1093/nar/gkv631.

Bienvenu F, Lambert A, Steel M. 2022. Combinatorial and stochastic properties of ranked tree-child networks. Random Structures & Algorithms. 60:653–689. doi:https://doi.org/10.1002/rsa.21048.

Blair C, Ané C. 2020. Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. Systematic Biology. 69:593–601. doi:https://doi.org/10.1093/sysbio/syz056.

Bock R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. Trends in Plant Science. 15:11–22. doi:10.1016/j.tplants.2009.10.001.

Bordewich M, Tokac N. 2016. An algorithm for reconstructing ultrametric tree-child networks from inter-taxa distances. Discrete Applied Mathematics. 213:47–59. doi:https://doi.org/10.1016/j.dam.2016.05.011.

Brysting AK, Oxelman B, Huber KT, Moulton V, Brochmann C. 2007. Untangling complex histories of genome mergings in high polyploids. Systematic Biology. 56:467–476. doi:https://doi.org/10.1080/10635150701424553.

Cao Z, Zhu J, Nakhleh L. 2019. Empirical Performance of Tree-Based Inference of Phylogenetic Networks. In: KT Huber, D Gusfield, editors. 19th International Workshop on Algorithms in Bioinformatics (WABI 2019). volume 143 of *Leibniz International Proceedings in Informatics (LIPIcs)*. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. p. 21:1–21:13. doi:10.4230/LIPIcs.WABI.2019.21.

Cardona G, Rosselló F, Valiente G. 2008. Comparison of tree-child phylogenetic networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 6:552–569. doi:https://doi.org/10.1109/tcbb.2007.70270.

Chapman MA, Burke JM. 2007. Genetic divergence and hybrid speciation. Evolution. 61:1773–1780. doi:https://doi.org/10.1111/j.1558-5646.2007.00134.x.

Dolinay M, Nečas T, Zimkus BM, Schmitz A, Fokam EB, Lemmon EM, Lemmon AR, Gvoždík V. 2021. Gene flow in phylogenomics: Sequence capture resolves species limits and biogeography of Afromontane forest endemic frogs from the Cameroon Highlands. Molecular Phylogenetics and Evolution. 163:107258. doi:https://doi.org/10.1016/j.ympev.2021.107258.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. Science. 366:594–599. doi:https://doi.org/10.1126/science.aaw2090.

Elworth RA, Ogilvie HA, Zhu J, Nakhleh L. 2019. Advances in computational methods for phylogenetic networks in the presence of hybridization. In: Bioinformatics and Phylogenetics. Springer. p. 317–360. doi:https://doi.org/10.1007/978-3-030-10837-3_13.

Erdős PL, Semple C, Steel M. 2019. A class of phylogenetic networks reconstructable from ancestral profiles. Mathematical Biosciences. 313:33–40. doi:https://doi.org/10.1016/j.mbs.2019.04.009.

Esquerré D, Keogh JS, Demangel D, Morando M, Avila LJ, Sites Jr JW, Ferri-Yáñez F, Leaché AD. 2022. Rapid radiation and rampant reticulation: Phylogenomics of South American *Liolaemus* lizards. Systematic Biology. 71:286–300. doi:https://doi.org/10.1093/sysbio/syab058.

Etienne RS, Rosindell J. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Systematic Biology. 61:204. doi:https://doi.org/10.1093/sysbio/syr091.

Flouri T, Jiao X, Rannala B, Yang Z. 2020. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Molecular Biology and Evolution. 37:1211–1223. doi:https://doi.org/10.1093/molbev/msz296.

Francis A, Huson DH, Steel M. 2021. Normalising phylogenetic networks. Molecular Phylogenetics and Evolution. 163:107215. doi:https://doi.org/10.1016/j.ympev.2021.107215.

Francis A, Moulton V. 2018. Identifiability of tree-child phylogenetic networks under a probabilistic recombination-mutation model of evolution. Journal of Theoretical Biology. 446:160–167. doi:https://doi.org/10.1016/j.jtbi.2018.03.011.

Francis AR, Steel M. 2015. Which phylogenetic networks are merely trees with additional arcs? Systematic Biology. 64:768–777. doi:https://doi.org/10.1093/sysbio/syv037.

Gourbière S, Mallet J. 2010. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". Evolution. 64:1–24. doi:https://doi.org/10.1111/j.1558-5646.2009.00844.x.

Gross E, Long C. 2018. Distinguishing phylogenetic networks. SIAM Journal on Applied Algebra and Geometry. 2:72–93. doi:10.1137/17M1134238.

Gross E, van Iersel L, Janssen R, Jones M, Long C, Murakami Y. 2021. Distinguishing level-1 phylogenetic networks on the basis of data generated by markov processes. Journal of Mathematical Biology. 83:1–24. doi:https://doi.org/10.1007/s00285-021-01653-8.

Heath TA, Hedtke SM, Hillis DM. 2008a. Taxon sampling and the accuracy of phylogenetic analyses. Journal of Systematics and Evolution. 46:239–257.

Heath TA, Zwickl DJ, Kim J, Hillis DM. 2008b. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Systematic Biology. 57:160–166. doi:https://doi.org/10.1080/10635150701884640.

Hejase HA, Liu KJ. 2016. A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. BMC Bioinformatics. 17:1–12. doi:https://doi.org/10.1186/s12859-016-1277-1.

Hibbins MS, Hahn MW. 2019. The timing and direction of introgression under the multispecies network coalescent. Genetics. 211:1059–1073. doi:https://doi.org/10.1534/genetics.118.301831.

Hibbins MS, Hahn MW. 2022. Phylogenomic approaches to detecting and characterizing introgression. Genetics. 220:iyab173. doi:https://doi.org/10.1093/genetics/iyab173.

Huber KT, Lott M, Moulton V, Spillner A. 2008. The complexity of deriving multi-labeled trees from bipartitions. Journal of Computational Biology. 15:639–651. doi:https://doi.org/10.1089/cmb.2008.0088.

Huber KT, Moulton V. 2006. Phylogenetic networks from multi-labelled trees. Journal of Mathematical Biology. 52:613–632. doi:https://doi.org/10.1007/s00285-005-0365-z.

Huber KT, Moulton V, Steel M, Wu T. 2016. Folding and unfolding phylogenetic trees and networks. Journal of Mathematical Biology. 73:1761–1780. doi:https://doi.org/10.1007/s00285-016-0993-5.

Huber KT, Van Iersel L, Kelk S, Suchecki R. 2010. A practical algorithm for reconstructing level-1 phylogenetic networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 8:635–649. doi:https://doi.ieeecomputersociety.org/10.1109/TCBB.2010.17.

Huber KT, Van Iersel L, Moulton V, Wu T. 2015. How much information is needed to infer reticulate evolutionary histories? Systematic Biology. 64:102–111. doi:https://doi.org/10.1093/sysbio/syu076.

Höhna S, May MR, Moore BR. 2015. TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics. 32:789–791. doi:10.1093/bioinformatics/btv651. https://academic.oup.com/bioinformatics/article-pdf/32/5/789/25588940/btv651.pdf.

Janssen R, Liu P. 2021. Comparing the topology of phylogenetic network generators. Journal of Bioinformatics and Computational Biology. 19:2140012. doi:https://doi.org/10.1142/S0219720021400126.

Jansson J, Sung WK. 2006. Inferring a level-1 phylogenetic network from a dense set of rooted triplets. Theoretical Computer Science. 363:60–68. doi:https://doi.org/10.1016/j.tcs.2006.06.022.

Jones GR. 2011. Tree models for macroevolution and phylogenetic analysis. Systematic Biology. 60:735–746. doi:https://doi.org/10.1093/sysbio/syr086.

Justison JA, Solis-Lemus C, Heath TA. 2023. SiPhyNetwork: An R package for simulating phylogenetic networks. Methods in Ecology and Evolution. 14:1687–1698. doi:10.1111/2041-210X.14116. https://besjournals.onlinelibrary.wiley.com/doi/pdfdirect/10.1111/2041-210X.14116.

Karimi N, Grover CE, Gallagher JP, Wendel JF, Ané C, Baum DA. 2020. Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (adansonia; bombacoideae; malvaceae). Systematic Biology. 69:462–478. doi:https://doi.org/10.1093/sysbio/syz073.

Kemp TS. 2009. Phylogenetic interrelationships and pattern of evolution of the therapsids: testing for polytomy. Palaeontologica Africana. 44:1–12.

Kendall DG. 1948. On the generalized "birth-and-death" process. The Annals of Mathematical Statistics. 19:1–15. doi:10.1214/aoms/1177730285.

Kong S, Pons JC, Kubatko L, Wicke K. 2022. Classes of explicit phylogenetic networks and their biological and mathematical significance. Journal of Mathematical Biology. 84:1–44. doi:https://doi.org/10.1007/s00285-022-01746-y.

Kozak KM, Joron M, McMillan WO, Jiggins CD. 2021. Rampant genome-wide admixture across the *Heliconius* radiation. Genome Biology and Evolution. 13:evab099. doi:https://doi.org/10.1093/gbe/evab099.

Lambert A. 2005. The branching process with logistic growth. The Annals of Applied Probability. 15:1506–1535. doi:10.1214/105051605000000098.

Lambert A, Stadler T. 2013. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. Theoretical Population Biology. 90:113–128. doi:https://doi.org/10.1016/j.tpb.2013.10.002.

Leaché AD, Harris RB, Rannala B, Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Systematic Biology. 63:17–30. doi:https://doi.org/10.1093/sysbio/syt049.

LeMay M, Libeskind-Hadas R, Wu YC. 2021. A polynomial-time algorithm for minimizing the deep coalescence cost for level-1 species networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 19:2642–2653. doi:https://doi.org/10.1109/TCBB.2021.3105922.

Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. Systematic Biology. 54:241–253. doi:https://doi.org/10.1080/10635150590924208.

Linz S, Semple C. 2020. Caterpillars on three and four leaves are sufficient to reconstruct binary normal networks. Journal of Mathematical Biology. 81:961–980. doi:https://doi.org/10.1007/s00285-020-01533-7.

López-Caamal A, Cano-Santana Z, Jiménez-Ramírez J, Ramírez-Rodríguez R, Tovar-Sánchez E. 2014. Is the insular endemic *Psidium socorrense* (Myrtaceae) at risk of extinction through hybridization? Plant Systematics and Evolution. 300:1959–1972. doi:https://doi.org/10.1007/s00606-014-1025-9.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution. 20:229–237. doi:https://doi.org/10.1016/j.tree.2005.02.010.

Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? BioEssays. 38:140–149. doi:10.1002/bies.201500149.

Meier JI, Stelkens RB, Joyce DA, Mwaiko S, Phiri N, Schliewen UK, Selz OM, Wagner CE, Katongo C, Seehausen O. 2019. The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. Nature Communications. 10:1–11. doi:https://doi.org/10.1038/s41467-019-13278-z.

Moen D, Morlon H. 2014. Why does diversification slow down? Trends in Ecology & Evolution. 29:190–197. doi:https://doi.org/10.1016/j.tree.2014.01.010.

Molloy EK, Durvasula A, Sankararaman S. 2021. Advancing admixture graph estimation via maximum likelihood network orientation. Bioinformatics. 37:i142–i150. doi:https://doi.org/10.1093/bioinformatics/btab267.

Montanari SR, Hobbs JPA, Pratchett MS, Van Herwerden L. 2016. The importance of ecological and behavioural data in studies of hybridisation among marine fishes. Reviews in Fish Biology and Fisheries. 26:181–198. doi:https://doi.org/10.1007/s11160-016-9420-7.

Mooers A, Gascuel O, Stadler T, Li H, Steel M. 2012. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. Systematic Biology. 61:195–203. doi:https://doi.org/10.1093/sysbio/syr090.

Mooers AO, Heard SB. 1997. Inferring evolutionary process from phylogenetic tree shape. The Quarterly Review of Biology. 72:31–54. doi:https://doi.org/10.1086/419657.

Morales-Briones DF, Liston A, Tank DC. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). New Phytologist. 218:1668–1684. doi:https://doi.org/10.1111/nph.15099.

Morin MM, Moret BME. 2006. NetGen: generating phylogenetic networks with diploid hybrids. Bioinformatics. 22:1921–1923. doi:10.1093/bioinformatics/btl191.

Morlon H, Potts MD, Plotkin JB. 2010. Inferring the dynamics of diversification: a coalescent approach. PLoS Biology. 8:e1000493. doi:https://doi.org/10.1371/journal.pbio.1000493.

Murakami Y, van Iersel L, Janssen R, Jones M, Moulton V. 2019. Reconstructing tree-child networks from reticulate-edge-deleted subnetworks. Bulletin of Mathematical Biology. 81:3823–3863. doi:https://doi.org/10.1007/s11538-019-00641-w.

Myers EA, Mulcahy DG, Falk B, Johnson K, Carbi M, de Queiroz K. 2022. Interspecific gene flow and mitochondrial genome capture during the radiation of Jamaican *Anolis* lizards (Squamata; Iguanidae). Systematic Biology. 71:501–511. doi:https://doi.org/10.1093/sysbio/syab089.

Nabhan AR, Sarkar IN. 2012. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Briefings in Bioinformatics. 13:122–134. doi:10.1093/bib/bbr014.

Nee S. 2006. Birth-death models in macroevolution. Annual Review of Ecology, Evolution, and Systematics. 37:1–17. doi:10.1146/annurev.ecolsys.37.091305.110035. `https://doi.org/10.1146/annurev.ecolsys.37.091305.110035`.

Nee S, Holmes EC, May RM, Harvey PH. 1994a. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 344:77–82. doi:https://doi.org/10.1098/rstb.1994.0054.

Nee S, May RM, Harvey PH. 1994b. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 344:305–311. doi:https://doi.org/10.1098/rstb.1994.0068.

Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics. 139:1805–1813. doi:https://doi.org/10.1093%2Fgenetics%2F139.4.1805.

Orr HA, Turelli M. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution. 55:1085–1094. doi:https://doi.org/10.1111/j.0014-3820.2001.tb00628.x.

Ottenburghs J. 2020. Ghost introgression: spooky gene flow in the distant past. Bioessays. 42:2000012. doi:https://doi.org/10.1002/bies.202000012.

Ottenburghs J, Kraus RH, van Hooft P, van Wieren SE, Ydenberg RC, Prins HH. 2017. Avian introgression in the genomic era. Avian Research. 8:1–11. doi:https://doi.org/10.1186/s40657-017-0088-z.

Ottenburghs J, van Hooft P, van Wieren SE, Ydenberg RC, Prins HH. 2016. Birds in a bush: Toward an avian phylogenetic network. The Auk: Ornithological Advances. 133:577–582. doi:https://doi.org/10.1642/AUK-16-53.1.

Pardi F, Scornavacca C. 2015. Reconstructible phylogenetic networks: do not distinguish the indistinguishable. PLoS Computational Biology. 11:e1004135. doi:https://doi.org/10.1371/journal.pcbi.1004135.

Pineda-Krch M. 2008. Gillespiessa: implementing the Gillespie stochastic simulation algorithm in R. Journal of Statistical Software. 25:1–18. doi:https://doi.org/10.18637/jss.v025.i12.

Pybus OG, Harvey PH. 2000. Testing macro–evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society of London. Series B: Biological Sciences. 267:2267–2272. doi:https://doi.org/10.1098/rspb.2000.1278.

Rabosky DL, Lovette IJ. 2008. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution. 62:1866–1875. doi:https://doi.org/10.1111/j.1558-5646.2008.00409.x.

Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. Journal of Molecular Evolution. 43:304–311. doi:https://doi.org/10.1007/BF02338839.

Raup DM. 1985. Mathematical models of cladogenesis. Paleobiology. 11:42–52. doi:10.1017/S0094837300011386.

Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR. 2008. Horizontal gene transfer of the algal nuclear gene psbo to the photosynthetic sea slug *Elysia chlorotica*. Proceedings of the National Academy of Sciences. 105:17867–17871. doi:https://doi.org/10.1073/pnas.0804968105.

Salzburger W. 2018. Understanding explosive diversification through cichlid fish genomics. Nature Reviews Genetics. 19:705–717. doi:https://doi.org/10.1038/s41576-018-0043-9.

Schulte U, Veith M, Hochkirch A. 2012. Rapid genetic assimilation of native wall lizard populations (*Podarcis muralis*) through extensive hybridization with introduced lineages. Molecular Ecology. 21:4313–4326. doi:https://doi.org/10.1111/j.1365-294X.2012.05693.x.

Semple C, Toft G. 2021. Trinets encode orchard phylogenetic networks. Journal of Mathematical Biology. 83:1–20. doi:https://doi.org/10.1007/s00285-021-01654-7.

Slatkin M. 1985. Gene flow in natural populations. Annual Review of Ecology and Systematics. 16:393–430. doi:https://doi.org/10.1146/annurev.es.16.110185.002141.

Slowinski JB. 2001. Molecular polytomies. Molecular Phylogenetics and Evolution. 19:114–120. doi:https://doi.org/10.1006/mpev.2000.0897.

Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genetics. 12:e1005896. doi:https://doi.org/10.1371/journal.pgen.1005896.

Solis-Lemus C, Coen A, Ane C. 2020. On the identifiability of phylogenetic networks under a pseudolikelihood model. arXiv preprint arXiv:2010.01758. doi:https://arxiv.org/abs/2010.01758.

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. Annual Review of Plant Biology. 60:561–588. doi:10.1146/annurev.arplant.043008.092039.

Stadler T. 2008. Lineages-through-time plots of neutral models for speciation. Mathematical Biosciences. 216:163–171. doi:https://doi.org/10.1016/j.mbs.2008.09.006.

Stanley EL, Bauer AM, Jackman TR, Branch WR, Mouton PLFN. 2011. Between a rock and a hard polytomy: rapid radiation in the rupicolous girdled lizards (Squamata: Cordylidae). Molecular Phylogenetics and Evolution. 58:53–70. doi:https://doi.org/10.1016/j.ympev.2010.08.024.

Stebbins GL. 1959. The role of hybridization in evolution. Proceedings of the American Philosophical Society. 103:231–251. doi:https://doi.org/10.1046/j.1365-294x.2001.01216.x.

Steel M, Hordijk W, Kauffman SA. 2020. Dynamics of a birth–death process based on combinatorial innovation. Journal of Theoretical Biology. 491:110187. doi:https://doi.org/10.1016/j.jtbi.2020.110187.

Steel M, Mooers A. 2010. The expected length of pendant and interior edges of a Yule tree. Applied Mathematics Letters. 23:1315–1319. doi:https://doi.org/10.1016/j.aml.2010.06.021.

Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. Zoologica Scripta. 45:50–62. doi:https://doi.org/10.1111/zsc.12213.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'agostino ER, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Current Biology. 32:111–123. doi:https://doi.org/10.1016/j.cub.2021.10.052.

Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nature Ecology & Evolution. 3:170–177. doi:https://doi.org/10.1038/s41559-018-0777-y.

Tea YK, Hobbs JPA, Vitelli F, DiBattista JD, Ho SY, Lo N. 2020. Angels in disguise: sympatric hybridization in the marine angelfishes is widespread and occurs between deeply divergent lineages. Proceedings of the Royal Society B. 287:20201459. doi:https://doi.org/10.1098/rspb.2020.1459.

Thawornwattana Y, Dalquen D, Yang Z. 2018. Coalescent analysis of phylogenomic data confidently resolves the species relationships in the *Anopheles gambiae* species complex. Molecular Biology and Evolution. 35:2512–2527. doi:https://doi.org/10.1093/molbev/msy158.

Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, Heredia SM, Hahn MA, Caseys C, Bock DG, Rieseberg LH. 2016. Hybridization and extinction. Evolutionary Applications. 9:892–908. doi:https://doi.org/10.1111/eva.12367.

Tricou T, Tannier E, de Vienne DM. 2022. Ghost lineages highly influence the interpretation of introgression tests. Systematic Biology. 71:1147–1158. doi:https://doi.org/10.1093/sysbio/syac011.

Van Iersel L, Keijsper J, Kelk S, Stougie L, Hagen F, Boekhout T. 2009. Constructing level-2 phylogenetic networks from triplets. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 6:667–681. doi:10.1109/TCBB.2009.22.

van Iersel L, Kole S, Moulton V, Nipius L. 2022. An algorithm for reconstructing level-2 phylogenetic networks from trinets. Information Processing Letters. 178:106300. doi:https://doi.org/10.1016/j.ipl.2022.106300.

Van Iersel L, Moulton V. 2014. Trinets encode tree-child and level-2 phylogenetic networks. Journal of Mathematical Biology. 68:1707–1729. doi:https://doi.org/10.1007/s00285-013-0683-5.

van Iersel L, Moulton V, Murakami Y. 2020. Reconstructibility of unrooted level-k phylogenetic networks from distances. Advances in Applied Mathematics. 120:102075. doi:https://doi.org/10.1016/j.aam.2020.102075.

Velasco JD. 2008. The prior probabilities of phylogenetic trees. Biology & Philosophy. 23:455–473. doi:https://doi.org/10.1007/s10539-007-9105-7.

Wang Y, Zhao Y, Pan Q. 2022. Advances, challenges and opportunities of phylogenetic and social network analysis using COVID-19 data. Briefings in Bioinformatics. 23:bbab406. doi:https://doi.org/10.1093/bib/bbab406.

Warnock RC, Heath TA, Stadler T. 2020. Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates. Paleobiology. 46:137–157. doi:10.1017/pab.2020.12.

Wawerka M, Dąbkowski D, Rutecka N, Mykowiecka A, Górecki P. 2022. Embedding gene trees into phylogenetic networks by conflict resolution algorithms. Algorithms for Molecular Biology. 17:1–23. doi:https://doi.org/10.1186/s13015-022-00218-8.

Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. 2009. Evolution and natural history of the cotton genus. In: Genetics and Genomics of Cotton. Springer. p. 3–22. doi:http://dx.doi.org/10.1007/978-0-387-70810-2_1.

Wendel JF, Cronn RC. 2003. Polyploidy and the evolutionary history of cotton. Advances in Agronomy. 78:139–186.

Willson S. 2010a. Regular networks can be uniquely constructed from their trees. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 8:785–796. doi:10.1109/TCBB.2010.69.

Willson SJ. 2008. Reconstruction of certain phylogenetic networks from the genomes at their leaves. Journal of Theoretical Biology. 252:338–349. doi:https://doi.org/10.1016/j.jtbi.2008.02.015.

Willson SJ. 2010b. Properties of normal phylogenetic networks. Bulletin of Mathematical Biology. 72:340–358. doi:https://doi.org/10.1007/s11538-009-9449-z.

Wolf DE, Takebayashi N, Rieseberg LH. 2001. Predicting the risk of extinction through hybridization. Conservation Biology. 15:1039–1053. doi:https://doi.org/10.1046/j.1523-1739.2001.0150041039.x.

Woodhams MD, Lockhart PJ, Holland BR. 2016. Simulating and summarizing sources of gene tree incongruence. Genome Biology and Evolution. 8:1299–1315. doi:10.1093/gbe/evw065.

Zhang C, Ogilvie HA, Drummond AJ, Stadler T. 2018. Bayesian inference of species networks from multilocus sequence data. Molecular Biology and Evolution. 35:504–517. doi:10.1093/molbev/msx307.

Zhang L. 2016. On tree-based phylogenetic networks. Journal of Computational Biology. 23:553–565. doi:https://doi.org/10.1089/cmb.2015.0228.

Zhang L. 2019. Clusters, trees, and phylogenetic network classes. In: Bioinformatics and Phylogenetics. Springer. p. 277–315. doi:https://doi.org/10.1007/978-3-030-10837-3_12.