Evolutionary Journal of the Linnean Society, 2025, **4(1)**, kzaf019
https://doi.org/10.1093/evolinnean/kzaf019

Original Article

Extracting diamonds: identifiability of 4-node cycles in level-1 phylogenetic networks

George P. Tiley [1,‡], Nan Liu [2,‡], Claudia Solís-Lemus [3,*]

[1]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27606, United States
[2]Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, United States
[3]Department of Plant Pathology, Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53706, United States

*Corresponding author. Department of Plant Pathology, Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53706, United States.
E-mail: solislemus@wisc.edu

‡= Equal contribution.

ABSTRACT

Phylogenetic networks encode a broader picture of evolution by the inclusion of reticulate processes such as hybridization, introgression, or horizontal gene transfer. Each hybridization event is represented by a 'hybridization cycle'. Here, we investigate the statistical identifiability of the position of the hybrid node in a 4-node hybridization cycle in a semi-directed level-1 phylogenetic network. That is, we investigate if our model is able to detect the correct placement of the hybrid node in the hybridization cycle using quartet concordance factors as data. In the current study, we prove that the correct placement of the hybrid node in 4-node hybridization cycles, included in level-1 phylogenetic networks, is generically identifiable if the assumptions are non-restrictive such as $t \in (0, \infty)$ for all branch (or edge) lengths and $\gamma \in (0, 1)$ for the inheritance probability of the hybrid edges. However, simulations show that accurate detection of these cycles can be complicated by inadequate sampling, small sample size, or gene tree estimation error. We identify practical advice for evolutionary biologists on best sampling strategies to improve the detection of this type of hybridization cycle.

Keywords: phylogenetic networks; reticulation; identifiability; coalescent model

INTRODUCTION

The increasing evidence of reticulate processes in the Tree of Life has inspired the development of novel mathematical and statistical methods to infer phylogenetic networks from genomic data (Huson and Bryant 2006, Degnan 2018, Jiao *et al.* 2021, Kong *et al.* 2022). A phylogenetic network is an extension of the tree graph by the inclusion of hybrid nodes that allow for two incoming parental edges which represent genetic material being transferred from two different sources. Methods to infer phylogenetic networks include classes from combinatorial (Grünewald *et al.* 2007, Grunewald *et al.* 2013, Yang *et al.* 2013, Oldman *et al.* 2016, Tan *et al.* 2019), distance-based (Willems *et al.* 2014, Markin *et al.* 2019, Xu and Ané 2022), or model-based under Markov models that use genetic sequences or markers as input (Gross *et al.* 2021, Rabier *et al.* 2021, Lutteropp *et al.* 2022) or under the multispecies coalescent model (Rannala and Yang 2003) that uses gene trees as input (Yu *et al.* 2014, Solís-Lemus and Ané 2016, Zhang *et al.* 2018).

Model-based approaches provide accurate representations of the reticulate evolutionary process among species represented by the estimated phylogenetic network. However, the accurate estimation of this parameter requires proof of identifiability. Indeed, multiple studies have tackled the question of whether phylogenetic networks are identifiable under different models (Pardi and Scornavacca 2015, Solís-Lemus and Ané 2016, Gross and Long 2018, Long and Kubatko 2018, Allman et al. 2019, Baños 2019, Gross *et al.* 2021, Solis-Lemus *et al.* 2020, Xu and Ané 2022, Allman *et al.* 2024), and here, we extend the identifiability proofs in Solís-Lemus and Ané (2016) and Solis-Lemus *et al.* (2020) to further validate the inference of level-1 phylogenetic networks under a pseudolikelihood model built on the multispecies network coalescent model (MSNC). Since identifiability can be unattainable in many cases, in practice, one aims to prove *generic identifiability* which means that the network parameter is almost surely identifiable.

Let $\mathcal{N}$ be a semi-directed level-1 phylogenetic network with h hybridization events and n taxa. We denote this the *network parameter* of the model. Let $\mathcal{G} = \{G_1, G_2, \ldots, G_g\}$ be g estimated gene trees as input for the model. We can define a likelihood as the product of the probability under the

Received 08 February 2025; **Revised** 23 June 2025; **Accepted** 19 August 2025

multispecies network coalescent model (Yu *et al.* 2012) of each gene tree given the network parameter (with its numerical parameters as branch lengths and inheritance probabilities). For the case of a species tree as the model parameter, it has been proven that the distribution of gene trees identifies (or generically identifies) the tree topology, branch lengths and even root position under the multispecies coalescent model in trees (Chang 1996, Allman and Rhodes 2006, Allman *et al.* 2011, Rhodes and Sullivant 2012). The same is not known for the full likelihood model on species networks, but generic identifiability proofs have been derived for the case of a quartet-based pseudolikelihood model (Solís-Lemus and Ané 2016, Baños 2019, Solis-Lemus *et al.* 2020) that assumes probabilities of each observed quartet gene tree under its corresponding species quarnet are independent. Note that pseudolikelihood approaches have been used to infer species tree as well when full likelihood calculations are computationally infeasible (Liu *et al.* 2010). The identifiability results in networks show that the presence of hybridization events in n-taxon level-1 semi-directed phylogenetic networks is generically identifiable from the observed distribution of gene trees under a pseudolikelihood model as described in the next theorem.

Theorem 1 [Generic identifiability of a hybridization event (Solís-Lemus and Ané 2016, Solis-Lemus *et al.* 2020)]. *Let $\mathcal{N}$ be a semi-directed level-1 n-taxon phylogenetic network with one hybridization event. Let k be the number of nodes in the cycle defined by the hybridization event. Then, the hybridization event is generically identifiable if (i) $k > 3$, (ii) $t \in (0, \infty)$ for all branch lengths, and (iii) $\gamma \in (0, 1)$ for the inheritance probability corresponding to the hybridization event.*

However, it remained unproven whether the position of the hybrid node within the hybridization cycle was identifiable. Here, we begin to explore the question of network identifiability by focusing on level-1 phylogenetic networks with hybridization cycles of 4 nodes (4-node cycles). These types of cycles [denoted *diamonds* (Pickrell and Pritchard 2012)] are relevant because in practice they seem to lead to flat pseudolikelihood scores (Solís-Lemus and Ané 2016). That is, 4-node cycles that only differ on the placement of the hybrid node in the cycle can have similar pseudolikelihood scores. Here, we show that a semi-directed network (correct direction of hybrid edges) is generically identifiable from input unrooted gene trees, not only the unrooted network (placement of the cycle). This implies that the direction of the gene flow event (hybrid edge) can be distinguished with the pseudolikelihood model, providing an accurate evolutionary interpretation of the origin and target of hybridizations. Our proof follows the same algebraic geometry techniques as in Solís-Lemus and Ané (2016) and Solis-Lemus *et al.* (2020) such as the definition of the set of polynomial equations under the coalescent model and the search for unique (or finitely many) solutions.

The organization of the manuscript is as follows. In second Section, we establish the models more precisely and the approach used to prove generic identifiability. In third Section, we define the theorems and proofs for the generic identifiability of the placement of the hybrid node in the 4-node hybridization cycle. In fourth and fifth Sections, we illustrate the mathematical findings (and the practical considerations when dealing with limited data) with a simulation study. Finally, we present some discussion in Section 6.

MATERIALS AND METHODS

Semi-directed level-1 explicit phylogenetic networks

Our main parameter of interest is the topology $\mathcal{N}$ of a phylogenetic network (Definition 1) along with the numerical parameters of the vector of branch lengths ($\boldsymbol{t}$) and a vector of inheritance probabilities ($\boldsymbol{\gamma}$). The inheritance probabilities describe the proportion of genes inherited by a hybrid node from one of its hybrid parents (see Fig. 1).

Definition 1. *A rooted explicit phylogenetic network* (Huson et al. 2010) *$\mathcal{N}$ on taxon set X is a connected directed acyclic graph with vertices* $V = \{r\} \cup V_L \cup V_H \cup V_T$,

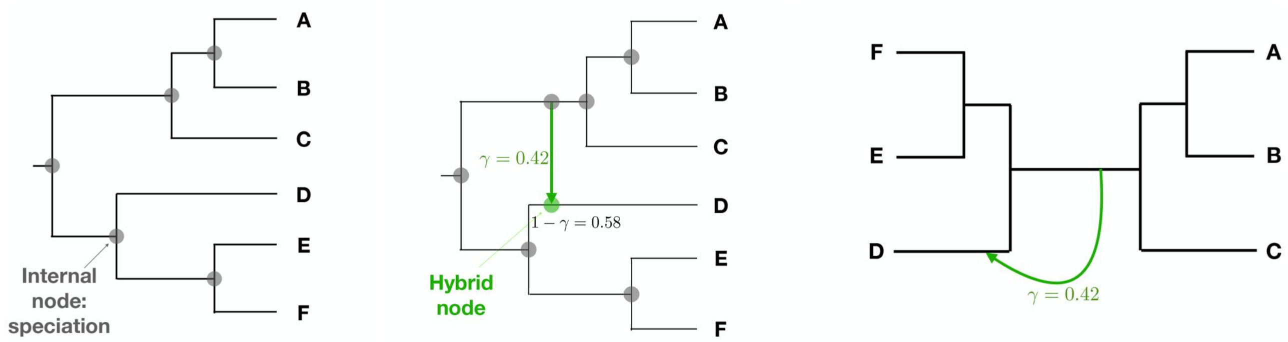


Figure 1. Left: rooted phylogenetic tree of six taxa. Internal nodes (grey circles) represent speciation events and are usually omitted on phylogenetic trees. Centre: rooted phylogenetic network on six taxa with $h = 1$ hybridization event represented by the green edge. The hybrid node (in green) represents a hybridization event. The hybrid node has two parent edges: a minor hybrid edge in green labelled $\gamma = 0.42$ and a major hybrid edge in black labelled $1 - \gamma = 0.58$. The hybridization event can represent different biological processes: hybridization, horizontal gene transfer, or introgression. Right: a semi-directed network for the same biological scenario as in the centre. Although the root location is unknown, its position is constrained by the direction of the hybrid edges. For example, the taxon D cannot be an outgroup.

edges $E = E_H \cup E_T$, and a bijective leaf-labelling function $f : V_L \rightarrow X$ with the following characteristics: (i) the root r has indegree 0 and outdegree 2, (ii) any leaf $v \in V_L$ has indegree 1 and outdegree 0, (iii) any tree node $v \in V_T$ has indegree 1 and outdegree 2, (iv) any hybrid node $v \in V_H$ has indegree 2 and outdegree 1, (v) a tree edge $e \in E_T$ is an edge whose child is a tree node, (vi) a hybrid edge $e \in E_H$ is an edge whose child is a hybrid node, (vii) a hybrid edge $e \in E_H$ has an inheritance probability parameter $0 < \gamma_e < 1$ which represents the proportion of the genetic material that the child hybrid node received from this parent edge. For a tree edge e, $\gamma_e = 1$.

In a rooted explicit network, every internal node represents a biological process: speciation for tree nodes and hybridization for hybrid nodes. However, other types of phylogenetic networks, such as unrooted networks (Huson et al. 2010) and semi-directed networks (Solís-Lemus and Ané 2016, Gross *et al.* 2021) are also useful representations of evolutionary relationships. Unrooted phylogenetic networks are typically obtained by suppressing the root node and the direction of all edges. In semi-directed networks, on the other hand, the root node is suppressed and we ignore the direction of all tree edges, but we maintain the direction of hybrid edges, thus keeping information on which nodes are hybrids. The placement of the root is then constrained, because the direction of the two hybrid edges to a given hybrid node informs the direction of time at this node: the third edge must be a tree edge directed away from the hybrid node and leading to all the hybrid's descendants. Therefore, the root cannot be placed on any descendant of any hybrid node, although it might be placed on some hybrid edges. See Figure 1 for an example of a rooted explicit phylogenetic network and its semi-directed version.

Each hybridization event creates a cycle which represents a subgraph with at least two nodes and no cut-edges. A cut-edge is any edge in the network whose removal disconnects the network. Note that given that phylogenetic networks are acyclic, the term *cycle* refers to sequence of edges that would form a cycle if they were all undirected. We also note that for every hybridization event, there are two parent hybrid edges connected to the hybrid node: (i) a major hybrid edge with inheritance probability $\gamma > 0.5$, and (ii) a minor hybrid edge with inheritance probability $\gamma < 0.5$. Both edges are parametrized with the same γ.

Here, we focus on the case of semi-directed networks. In addition, we assume the following characteristics of the semi-directed network under study.

Assumption 1 *Let $\mathcal{N}$ have n leaves and h hybridization events (that is, $|V_L| = n$ and $|V_h| = h$).*

Assumption 2 We further assume that $\mathcal{N}$ is of level-1 (Huson et al. 2010). That is, we assume that any given edge can be part of at most one cycle which means that there is no overlap between any two cycles (Rosselló and Valiente 2009, Huson et al. 2010).

Thus, our parameters of interest are $(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma})$ where $\mathcal{N}$ is an explicit semi-directed level-1 phylogenetic network that links the n species under study, and has h hybridization events. This network has two vectors of numerical parameters: (i) branch lengths $\boldsymbol{t} \in [0, \infty)^{n_e}$ where n_e is the number of branches in the network, and (ii) inheritance probabilities $\boldsymbol{\gamma} \in (0, 0.5)^{n_h}$ for n_h, the number of minor hybrid edges.

Generic identifiability of the position of the hybrid node

To show that the position of the hybrid node in the hybridization cycle is generically identifiable, we represent every semi-directed level-1 network $\mathcal{N}$ as a set of polynomial quartet concordance factor (CF) equations as in Solís-Lemus and Ané (2016) and Solis-Lemus *et al.* (2020), and we find if both systems of polynomial equations share solutions in the parameter space: $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma}) = CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$. Note that the term concordance factor has been used for measures of agreement for a given bipartition from a collection of gene trees [gene concordance factors (Baum 2007)] or site patterns [site concordance factors (Minh *et al.* 2020)], but our proofs concern only quartet concordance factors (Solís-Lemus and Ané 2016) and the abbreviation CF is used throughout for simplicity.

First, each network can be decomposed into 4-taxon subnetworks [*quarnets* in Huber *et al.* (2018)]. That is, for a given network $\mathcal{N}$ with $n \geq 4$ taxa, we consider all 4-taxon subsets $\mathcal{S} = \{s = \{a, b, c, d\} : a, b, c, d \in X\}$ to define the theoretical CFs expected under the multispecies network coalescent model for each 4-taxon subset. These theoretical CFs are already derived for a species tree in (Allman *et al.* 2011), and for a species network in (Solís-Lemus and Ané 2016). In both cases, the CFs do not depend on the position of the root. For the tree, the major CF is defined for the quartet that agrees with the species tree. That is, if the species tree has the split $ab|cd$ with internal edge t, then the major CF would be $CF_{ab|cd} = 1 - 2/3\exp(-t)$. The CF for the minor resolutions (in disagreement with the species tree $ab|cd$) would then be $CF_{ac|bd} = CF_{ad|bc} = 1/3\exp(-t)$ (Hudson 1983).

For the case of a 4-taxon network, the theoretical CFs are weighted averages of CFs on trees. For example, Figure 2 shows a semi-directed 4-taxon network (▲ = a, ■ = b, ● = c, ● = d) and the two displayed quartet trees depending on which hybrid edge is used (major hybrid edge used corresponds to the quartet on top, and minor hybrid edge used corresponds to the quartet on the bottom). The CFs on this 4-taxon network are then given by a weighted sum of the quartet CFs of the two displayed quartet trees with weights: $1 - \gamma$ and γ. Namely,

- $CF_{ab|cd} = (1-\gamma)(1 - 2/3\exp(-t_{13} - t_{23} - t_2)) + \gamma(1 - 2/3\exp(-t_2))$ for the major resolution, and
- $CF_{ac|bd} = CF_{ad|bc} = (1-\gamma)(1/3\exp(-t_{13} - t_{23} - t_2)) + \gamma(1/3\exp(-t_2))$ for the minor resolutions.

The computation of CF equations when there are more than two leaves beneath a hybrid node is not as straightforward as weights depend on branch lengths and γ parameters. Details about this computation can be found in Solís-Lemus and Ané (2016). We show in the Appendix the system of polynomial equations for the four phylogenetic networks considered in this study: $\mathcal{N}_{down}$, $\mathcal{N}_{right}$, $\mathcal{N}_{left}$, and $\mathcal{N}_{up}$ (Fig. 3). Since each network is obtained by the rotation of the

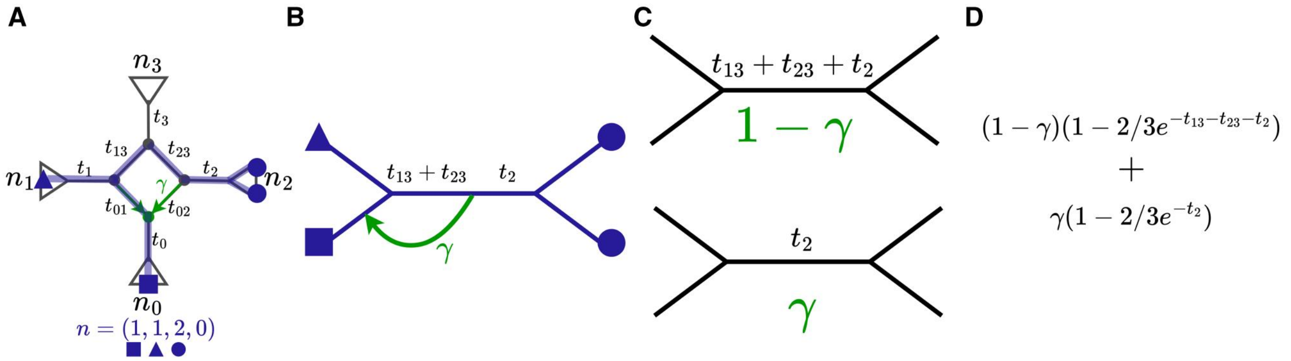


Figure 2. A, a semi-directed network with a given 4-taxon subset $(n_0 = 1, n_1 = 1, n_2 = 2, n_3 = 0)$ highlighted in blue which corresponds to taking one individual in n_0 (blue square), one individual in n_1 (blue triangle), and two individuals in n_2 (blue circles). B, to obtain the CF equations for the resulting quarnet (C) we split it as two quarnet trees weighted by $1-\gamma$ and γ. D, major CF equation for the quarnet as weighted average of CF equations on quarnet trees.

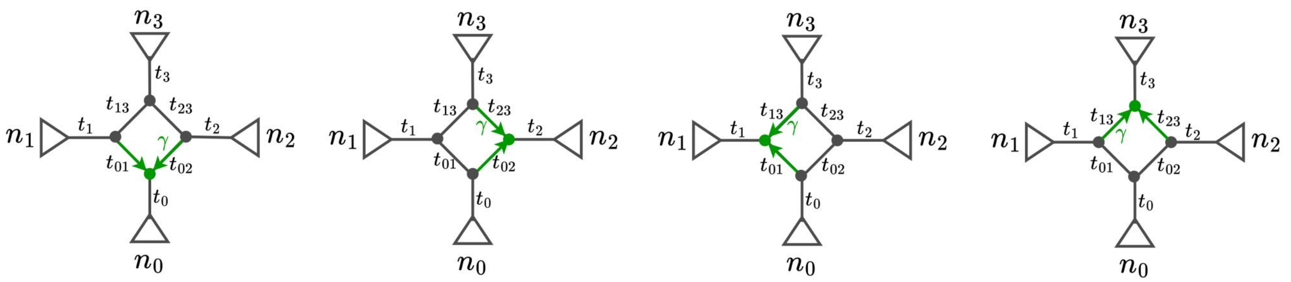


Figure 3. Semi-directed networks with one hybridization cycle with 4 nodes (diamond), but different position of the hybrid node in the cycle. We denote these networks $\mathcal{N}_{down}$, $\mathcal{N}_{right}$, $\mathcal{N}_{left}$, and $\mathcal{N}_{up}$ respectively from left to right.

hybrid node in the hybridization cycle, there is a simple mapping of 4-taxon subsets that allows us to define the CF equations for all networks from the original $\mathcal{N}_{down}$. This process is also described in the Appendix.

Definition 2 *Let $\mathcal{N}$ be n-taxon semi-directed level-1 explicit phylogenetic network with h hybridizations. This network defines a set of $3\binom{n}{4}$ CF equations under the coalescent model with parameters $\boldsymbol{t}$ and $\boldsymbol{\gamma}$. Denote this system of equations as $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma})$. If we change the variables to $z_i = \exp(-t_i)$ for all internal branch lengths, then $CF(\mathcal{N}, \boldsymbol{z}, \boldsymbol{\gamma})$ is a system of polynomial equations.*

Definition 3 (Generic identifiability of hybrid node position). *Let $\mathcal{N}$ be an n-taxon semi-directed level-1 explicit phylogenetic network with h hybridizations. We focus on one hybridization that has 4 nodes in the hybridization cycle. Let $\mathcal{N}'$ be a network with the hybrid node rotated inside the 4-node hybridization cycle. Let $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma})$ be the system of polynomial CF equations defined by $\mathcal{N}$, and let $CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$ be the system of polynomial CF equations defined by $\mathcal{N}'$. We say the position of the hybrid node in the 4-node hybridization cycle in $\mathcal{N}$ is* ***identifiable*** *if the system of $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma}) = CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$ does not have solutions in any set of numerical parameters $(\boldsymbol{t}, \boldsymbol{\gamma}, \boldsymbol{t}', \boldsymbol{\gamma}')$. We say the position of the hybrid node in the 4-node hybridization cycle in $\mathcal{N}$ is* ***generically identifiable*** *if the solution set of the system $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma}) = CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$ has measure zero.*

Identifiability of 4-node cycles in semi-directed level-1 phylogenetic networks

In this paper, we prove the following theorem regarding the generic identifiability of the placement of the hybrid node in the 4-node hybridization cycle in a level-1 network:

Theorem 2 *Let $\mathcal{N}$ be a semi-directed level-1 n-taxon phylogenetic network with one hybridization event that creates a 4-node hybridization cycle. Then, the placement of the hybrid node in the cycle is generically identifiable if (i) $n > 5$, (ii) $t \in (0, \infty)$ for all branch lengths, and (iii) $\gamma \in (0, 1)$ for the inheritance probability corresponding to the hybridization event.*

To prove this theorem, we prove three separate theorems: $\mathcal{N}_{down}$ vs. $\mathcal{N}_{right}$ (Theorem 3), $\mathcal{N}_{down}$ vs. $\mathcal{N}_{left}$ (Theorem 4), and $\mathcal{N}_{down}$ vs. $\mathcal{N}_{up}$ (Theorem 5). In this section, we show Theorem 3 and its proof, and the other two theorems and proofs are placed in the Appendix. For these three theorems, we assume the networks only have six taxa and one hybridization event. We generalize to n taxa and h hybridization events in Remark 2.

Theorem 3 *Let $\mathcal{N}_{down}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes. Without loss of generality, let the taxa be partitioned among clades as $n_0 = 1, n_1 = 2, n_2 = 1, n_3 = 2$ (Fig. 3). Let the hybrid node be ancestral to the clade n_0. Let $\mathcal{N}_{right}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes such that the unrooted version of $\mathcal{N}_{right}$ agrees with the unrooted version of $\mathcal{N}_{down}$. Let the hybrid node in the hybridization cycle in $\mathcal{N}_{right}$ be ancestral to the clade n_2. Then, $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$ are identifiable if $t_1 < \infty, t_{13} > 0, t_3 < \infty$, and $\gamma \in (0,1)$.*

Proof. Let $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ be the system of CF polynomial equations defined by $\mathcal{N}_{down}$ and let $CF(\mathcal{N}_{right}, \boldsymbol{z}', \boldsymbol{\gamma}')$ be the system of CF polynomial equations defined by $\mathcal{N}_{right}$. Both systems of equations can be found in the Appendix (Tables 1–14).

Let $\mathcal{P} = \{p(\boldsymbol{z}, \boldsymbol{\gamma}) - q(\boldsymbol{z}', \boldsymbol{\gamma}') : p(\boldsymbol{z}, \boldsymbol{\gamma}) \in CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma}), q(\boldsymbol{z}', \boldsymbol{\gamma}') \in CF(\mathcal{N}_{right}, \boldsymbol{z}', \boldsymbol{\gamma}')\}$ be the set of polynomial equations resulting from matching $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ to $CF(\mathcal{N}_{right}, \boldsymbol{z}', \boldsymbol{\gamma}')$ for every 4-taxon subset.

Using Macaulay2 (Grayson and Stillman undated), we compute the Gröbner basis of $\mathcal{P}$ on the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables by any elimination order. All Macaulay2 scripts are available at https://github.com/gtiley/diamond-identifiability.

The resulting ideal is given by:

$$z_1(z_{13} - 1)^2 z_3 \gamma (1 - \gamma) = 0$$

$$z_1(z_{13} - 1) z_3 (z_{23} - 1) \gamma (1 - \gamma) = 0$$

which represent the conditions that the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables need to satisfy for the polynomial set $\mathcal{P}$ to vanish to zero.

Thus, $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$ are not identifiable in the subset of parameter space corresponding to $\{z_1 = 0\} \cup \{z_{13} = 1\} \cup \{z_3 = 0\} \cup \{\gamma = 0\} \cup \{\gamma = 1\}$. ■

Remark 1. *We note that by assuming that $t \in (0, \infty)$ for all branch lengths, and $\gamma \in (0,1)$ for the inheritance probabilities, we can guarantee generic identifiability of the placement of the hybrid node in the 4-node hybridization cycle.*

Remark 2. *The identifiability of the position of the hybrid node in 4-node cycles in n-taxon level-1 phylogenetic networks is obtained by noticing that we need at most two taxa per clade to define all the CF polynomial equations [Lemma 1 in Solis-Lemus et al. (2020)], and thus, if the hybridization cycles are identifiable with only one taxon in some clades, the addition of a second taxon will only reduce the set in the parameter space where the two networks are not (generically) identifiable.*

Remark 3. *We did not explore the identifiability in the case of $n = 5$ because it was already discovered in Solís-Lemus and Ané (2016) and Solis-Lemus et al. (2020) that some hybridization events with only one taxon sampled in one of the four taxa partitions (labelled n_0, n_1, n_2, n_3 in Fig. 2) are either not generically detectable or the numerical parameters are not generically identifiable. These cases were denoted bad diamonds, and were excluded for the new proofs here.*

Simulations without gene tree estimation error

We simulated gene trees under the MSNC using BPP v.4.1.4 (Flouri *et al.* 2020). The level-1 network under investigation contained eight taxa and one 4-node hybridization cycle (diamond) (Fig. 4A). To simulate gene trees under the MSNC, we required a rooted evolutionary history, and thus, we investigated three potential root placements: (i) a balanced root included in the cycle (r_1 in Fig. 4B), (ii) a ladderized root with the root in the cycle (r_2 in Fig. 4C), and (iii) a ladderized root such that the cycle occurs in the ingroup (r_3 in Fig. 4D). The root height is not the same between the balanced root and ladderized rootings, but the age of episodic gene flow is $1\frac{1}{2}$ units of θ, population size measured in expected nucleotide diversity. Edge lengths for the networks used in simulations are in substitutions per site and the inheritance probability (γ) was 0.5 for all simulations. A θ of 0.01 was constant over all nodes. Simulations were carried out for nine diamonds for the three rooting scenarios, such that we changed if lineages were represented by a clade of two taxa or a single taxon (Fig. 4E).

Sequences were also simulated with BPP v.4.1.4, which were used in later comparisons between true gene trees and estimated gene trees. Sequence simulation used the HKY model of nucleotide substitution (Hasegawa *et al.* 1985) with a transition-transversion rate-ratio of 3 and equilibrium frequencies at $\pi_A = \pi_T = 0.3$ and $\pi_C = \pi_G = 0.2$ was selected. A low degree of among site rate variation was incorporated using a gamma distribution (Yang 1994) with a shape and rate of 0.6. A strict clock was used among branches and there is no among-locus rate variation. One-hundred simulations were performed for each of the 27 rooted networks, using 100, 500, 1000, and 5000 loci. Each locus had 1000 base pairs.

We then selected the ladderized root (root 3 in Fig. 4D) to assess the changes in performance due to the value of the inheritance parameter. We varied the number of gene trees and the inheritance probability across nine diamonds, as illustrated in Figure 4E. Specifically, we examined nine parameter combinations using the following values: 250, 1000, and 4000 gene trees, paired with an inheritance probability of 0.05, 0.25, and 0.5. For each diamond and parameter combination, we performed 100 replicate simulations. Scripts for reproducing simulation experiments are available at https://github.com/gtiley/diamond-identifiability.

Incorporating gene tree error into simulation experiments

Biological data are often messy with many potential sources of gene tree estimation error. For example, deep paralogy can

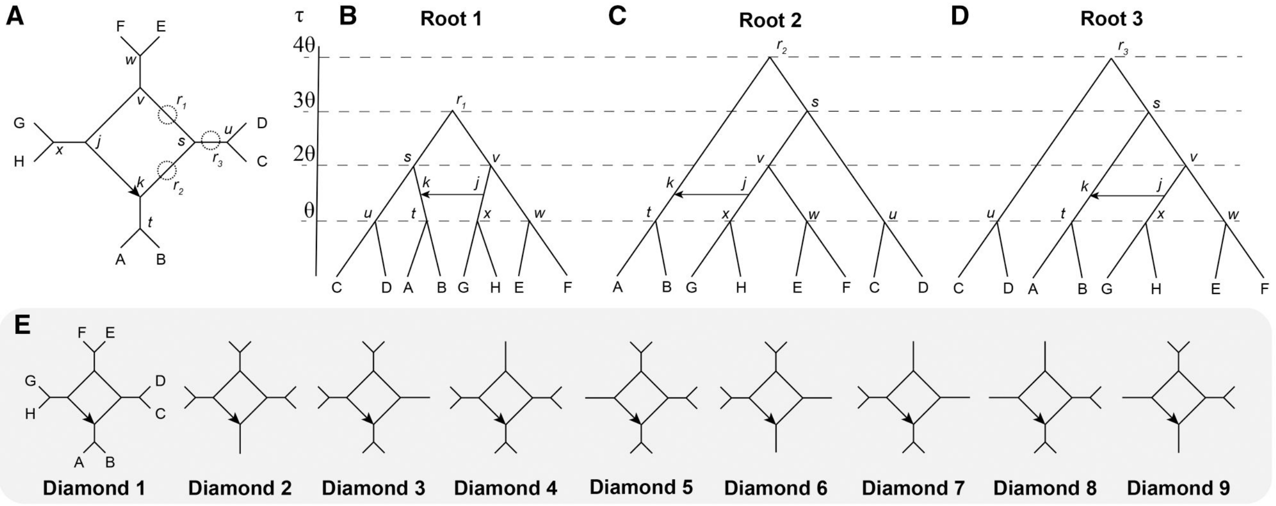


Figure 4. Semi-directed level-1 network topology with a diamond hybridization cycle. A, the semi-directed diamond has tip labels in capital letters and node labels in lower case italicized letters. The hybrid node is labelled as *k*. The locations of the three possible roots are indicated by r_1, r_2, r_3. B–D, node heights for the three rooted networks in units of θ population size. The direction of episodic gene flow is shown with an arrow. E, for each root, nine networks were investigated, the first including all tips, the next four reducing one cherry to a single branch in an anticlockwise fashion, the next four reducing two cherries to two lineages with a single taxon.

lead to gene trees that conflict with species trees (Maddison 1997) and lead to rooting errors when using an outgroup (Huelsenbeck *et al.* 2002, Holland *et al.* 2003). Such gene tree errors can be prevalent in groups with histories of large-scale gene duplication and loss such as plants (Leebens-Mack et al. 2019) and insects (Li *et al.* 2018). These rooting errors may negatively affect the performance of inference methods, especially those based on rooted triples. Therefore, we estimated networks while incorporating gene tree error with two different strategies. First, for each simulation condition, we randomly re-rooted 10% and 30% of the gene trees. Because rooting errors may also be accompanied by other topological errors, we estimated networks from the simulated data incorporating a random re-rooting and a random nearest neighbour interchange (NNI) move to 10% and 30% of gene trees. Second, because our NNI approach may explore a relatively limited set of gene tree errors, we re-estimated gene trees for a subset of simulation conditions over a range of alignment lengths as a proxy for gene tree information. Based on the first set of simulations with rooting and NNI errors, we identified five diamonds (d_1, d_2, d_3, d_8, d_9) that appeared representative for all diamonds. We then fixed the number of gene trees to 1000 and the inheritance probability at 0.25, while varying the sequence length across 250, 1000, and 4000 for each of the five selected diamonds. Gene trees were then estimated with IQ-TREE v.2.3.6 (Minh *et al.* 2020) from gene sequence data using the same model used in the simulations (HKY+G), to avoid incorporating any model mis-specification error. Every scenario was replicated 20 times. Scripts for introducing random roots and NNI moves to trees are available at https://github.com/gtiley/diamond-identifiability.

Network estimation

Networks were estimated from the simulated or estimated gene trees using pseudolikelihood methods that utilize the gene tree topologies only. We used two methods: (i) the Species Networks applying Quartets (SNaQ) function (Solís-Lemus and Ané 2016) implemented through the PhyloNetworks v.0.15.0 Julia package (Solís-Lemus *et al.* 2017) in Julia v.1.6.5 where SNaQ implements statistical inference by involving numerical optimization of branch lengths and inheritance probabilities, and a heuristic search in the space of level-1 phylogenetic networks; and (ii) the InferNetwork_MPL function (Yu and Nakhleh 2015) implemented through PhyloNet v.3.8.3 (Wen *et al.* 2018). Notably, SNaQ uses unrooted quartets as data while InferNetwork_MPL uses rooted triples (more on rooting error later). Both analyses used 10 independent runs with the maximum number of allowed hybridization events set to zero, one, or two.

Three summary statistics were used to evaluate the performance of the estimations. First, we calculated the proportion of simulations where only one hybridization event was detected based on a two-point pseudolikelihood score difference. Although the pseudolikelihood does not have straight-forward implementation of model selection such as the Akaike Information Criterion (Akaike 1974), we used differences in pseudolikelihood scores as some operational criteria to evaluate detection of hybridization edges among the replicates in lieu of slope heuristics (Baudry *et al.* 2012) or goodness-of-fit tests (Cai and Ané 2021) that might be more appropriate for an empirical investigation. Second, we checked if the network estimated when allowing one hybridization event had the same topology as the true network using the hardwiredClusterDistance function (Huson et al. 2010) in PhyloNetworks. While the absolute distance between networks is difficult to interpret, a distance of 0 means the network topologies are identical. The proportion of correct networks was calculated only for networks with one hybridization event regardless of how many events were preferred via model selection, as different selection criteria and additional lines of evidence could help guide decision making. Third, we checked if the true minor hybrid edge is present in the estimated network with only one estimated hybridization event. This is done by drawing bootstrap support from the estimated network onto the true network with the hybridBootstrapSupport function from PhyloNetworks. A bootstrap support of 100% shows that the true hybrid edge is present in the estimated network, even if some other aspects of the network are incorrect. Scripts for calculating these summary statistics are available at https://github.com/gtiley/diamond-identifiability.

RESULTS

Simulation results: error-free gene trees

The pseudolikelihood methods implemented in PhyloNetworks and PhyloNet generally performed well across all simulation conditions. SNaQ was capable of correctly identifying the presence of one hybridization event, recovering the correct network, and recovering the hybridization node in all simulations as the number of gene trees increased (Supporting Information, Figs S1–S3). For diamonds d_6 and d_9 when the rooting was ladderized and in the cycle (r_2) or ladderized and out of the cycle (r_3), SNaQ sometimes missed the correct network or hybridization node when the number of gene trees was low, but converged to the correct network as gene trees increased from 100 to 5000. InferNetwork_MPL was more efficient with respect to the number of gene trees for d_6 or d_9 and performed well for r_1 (Supporting Information, Fig. S4) and r_2 (Supporting Information, Fig. S5), but sometimes struggled with r_3 (Supporting Information, Fig. S6). Across all diamonds for r_3, InferNetwork_MPL did not always detect one hybridization edge, such that two hybridization edges were preferred over none. Even when considering the estimated networks that only allowed one hybridization event, the network was not always correct; however, the correct hybridization node was almost always recovered (Supporting Information, Fig. S6). If the recovery of the hybridization node was the most favourable criteria of an estimator, both SNaQ and InferNetwork_MPL performed well with error-free gene trees, with SNaQ requiring more gene trees in cases where the hybrid node had a single taxon for a descendant instead of a cherry (Fig. 5).

When the inheritance probability was either 0.25 and 0.5, SNaQ was capable of identifying the presence of one hybridization event, recovering the hybridization node, and recovering the correct network when the number of gene trees increased (Fig. 6). For diamond d_5, d_6 and d_9, the correct network or hybridization node was sometimes missed when the number of gene trees was low (250), but converged to the correct network as the gene trees increased to 1000 and 4000. When the inheritance probability was 0.05, recovering the correct network and recovering the hybridization node had an increasing trend across all diamonds while the number of gene trees increased. Diamond d_9 could only be recovered about 50 percent even with 4000 gene trees. Furthermore, SNaQ could identify the presence of the one hybridization event for all diamonds excluding d_6 and d_9, whose presence rate was decreasing while the number of gene trees increased. Since d_6 and d_9 were symmetric, their results both reflected false negatives when including more gene trees. We also noticed that CorrectNetwork and HasHybridEdge were identical, which was a special case. In other general cases, these two evaluations could perform differently.

Simulation results: gene trees with error

SNaQ tolerated rooting errors well when they were present in both 10% (Supporting Information, Figs S7–S9) and 30% (Supporting Information, Figs S10–S12) of gene trees. The patterns for d_6 and d_9 with r_2 and r_3 from error-free gene trees were observed again, which was expected since SNaQ used unrooted quartets as data. InferNetwork_MPL was largely robust to root error in 10% of the gene trees for r_1 and r_2, with the exception of sometimes detecting more than one hybridization edge (Supporting Information, Figs S13, S14). However, it was difficult to recover the correct network for r_3 and the proportion of incorrect networks sometimes increased with the number of gene trees (Supporting Information, Fig. S15). The lowest proportions of correct network estimations were observed for diamonds d_4, d_7, and d_9, which were not all cases where the hybrid node had a single taxon as a descendant. The overall network being incorrect did not always mean that the estimated hybridization node was wrong, but the decreased performance compared to simulations without gene tree error was drastic (Supporting Information, Table S1). As the proportion of gene tree errors increased from 10% to 30%, similar patterns were observed for InferNetwork_MPL, such that d_4, d_7, and d_9 performed poorly for r_3, but with a higher proportion of network estimation errors (Supporting Information, Figs S16–S18). For example, the hybridization node was never recovered for d_9 with r_3 and 5000 gene trees (Supporting Information, Fig. S18). Introducing an additional NNI move error yielded qualitatively similar simulation results (Supporting Information, Figs. S19–S30), implying both methods are robust to small degrees of estimation error.

When utilizing estimated gene trees, accurate estimation improves as sequence length increases as is expected (Figure 8). Only d9 performed a different trend as the fraction dropped for large sequence length.

DISCUSSION

Mathematical insights

We have shown that the placement of the hybrid node in a 4-node hybridization cycle is generically identifiable in semi-directed level-1 phylogenetic networks. This result has important biological implications such as correct rooting of a semi-directed network which provides more information about the actual speciation process from the origin of the clade. More work is needed to understand the identifiability of larger (and smaller) cycles. We decided to focus on the case of the 4-node hybridization cycle because of empirical evidence suggesting flat pseudolikelihood on these hybridization events. However, smaller cycles are also biologically interesting as they describe hybridization events between more closely related populations, but it is not always true under incomplete taxon sampling or in the presence of extinct taxa. In such cases, the resulting cycle may also be small but with longer edge lengths. It has already been proven that 2-node cycles are not identifiable, but 3-node cycles are whenever we have sufficient

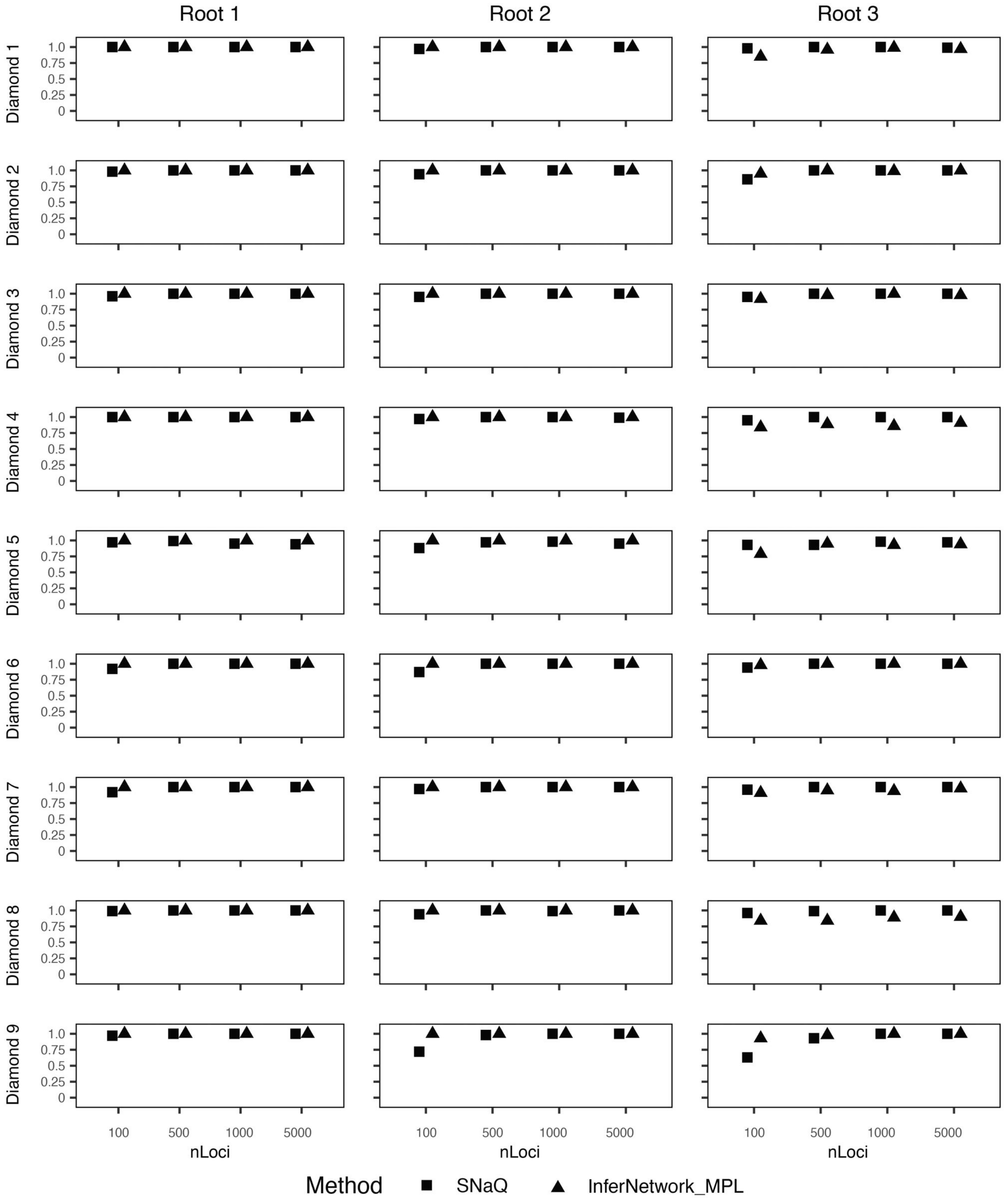


Figure 5. Proportion of 100 simulation replicates where the true hybridization edge is recovered by SNaQ (squares) or InferNetwork_MPL (triangles). The x-axis corresponds to the number of gene trees in the input sample. The diamond and root numbers correspond to Figure 4.

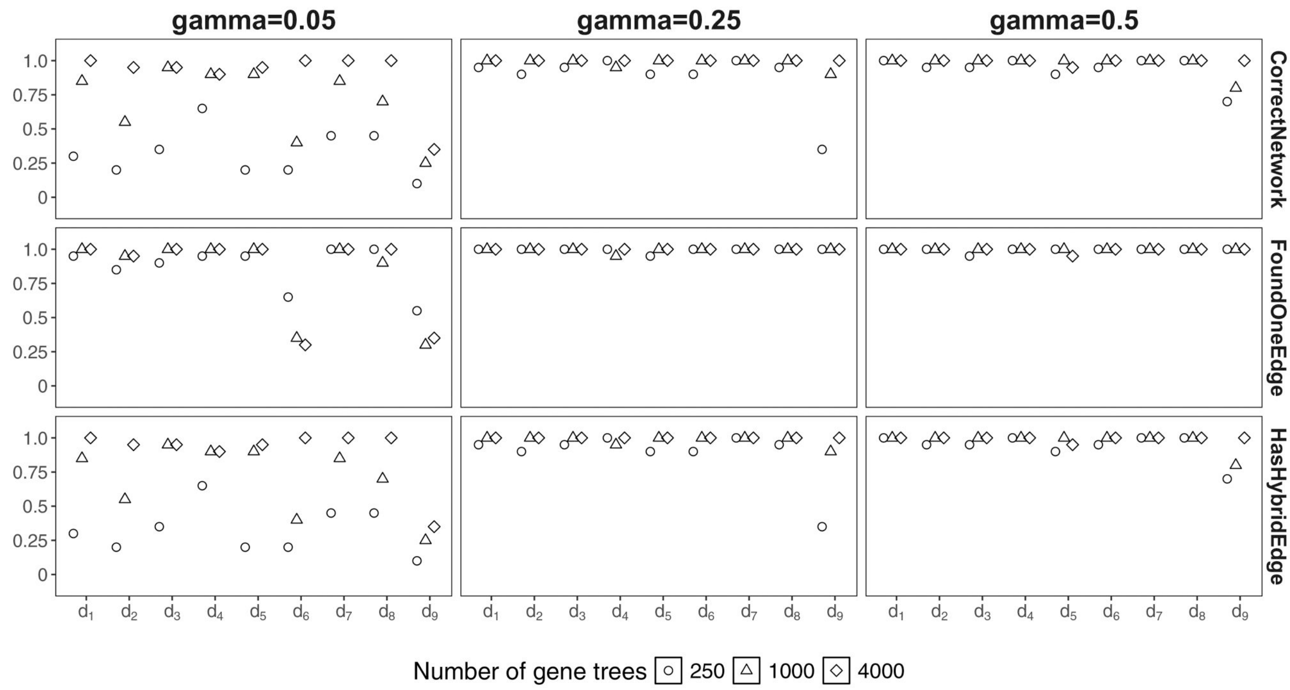


Figure 6. SNaQ results for root 3 with different value of the number of gene trees paired with the inheritance probability. *FoundOneEdge* is the proportion out of 20 replicates that one reticulate edge was correctly inferred by the pseudolikelihood scores. *CorrectNetwork* is the proportion of 20 replicates where the estimated topology when allowing only one hybridization event is identical to the true topology. *HasHybridEdge* is the proportion of 20 replicates where the correct hybridization edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one hybridization event. Gamma represents the inheritance probability. Diamond d_1 through tod_9 correspond to Figure 4E. Note that the plots in *CorrectNetwork* and *HasHybridEdge* are identical which suggests that when the correct hybrid edges are recovered, the true network is estimated.

sampling from the hybrid and sister clades (Solís-Lemus and Ané 2016, Solis-Lemus *et al.* 2020). Thus, we can explore the identifiability of the placement of the hybrid node in the hybridization 3-node cycle as future work. In particular, we want to answer the following question:

> **Question 1.** *Let $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3$ be the three rotations of the n-taxon semi-directed level-1 phylogenetic networks with one hybridization event producing a hybridization cycle of 3 nodes. Are $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3$ (generically) identifiable?*

Furthermore, all existing identifiability studies on semi-directed phylogenetic networks are restricted to a level-1 network (Solís-Lemus and Ané 2016, Gross and Long 2018, Baños 2019, Gross *et al.* 2021, Solis-Lemus *et al.* 2020, Allman *et al.* 2024). Questions on the identifiability of higher-level networks remain open.

Biological insights from simulations and best practices

Pseudolikelihood methods can perform well for estimating a level-1 network from gene tree distributions in the absence of gene tree error, consistent with previous simulations and empirical analyses (Yu and Nakhleh 2015, Solís-Lemus and Ané 2016). PhyloNet's InferNetwork_MPL was more efficient than PhyloNetwork's SNaQ for diamonds d_2, d_6, and d_9, capable of recovering the correct network with as few as 100 gene trees where SNaQ would require at least 500 (Fig. 5). However, InferNetwork_MPL was not always perfect even with error-free gene trees such as d_8 for r_3. In addition, SNaQ was far more robust to the types of gene tree errors explored here than InferNetwork_MPL, capable of detecting one hybridization event and recovering the correct network in most conditions when provided a sufficient number of gene trees (Fig. 7). InferNetwork_MPL could still detect the presence of a hybridization edge more often than not depending on the diamond analysed with r_3, but the estimated network was frequently incorrect and if one were using unknown empirical data, there is a risk of estimating additional false positive hybridization nodes in the presence of gene tree rooting errors.

The diamonds and rootings where network estimation errors were most prevalent have implications for empirical analyses. SNaQ errors primarily occurred for d_2, d_6, and d_9, all diamonds where the hybrid node descendant was a single taxon (Fig. 4E). This implies that sampling two or more individuals for the lineage of interest could provide an increase in statistical power over a single taxon, at least for SNaQ and methods that use unrooted gene trees. Interestingly, network estimation errors were most prevalent when the cycle was restricted to an ingroup (r_3). InferNetwork_MPL especially struggled with d_4, d_7, and d_9 in the presence of errors, as it appears more data would not be helpful. Precisely why these

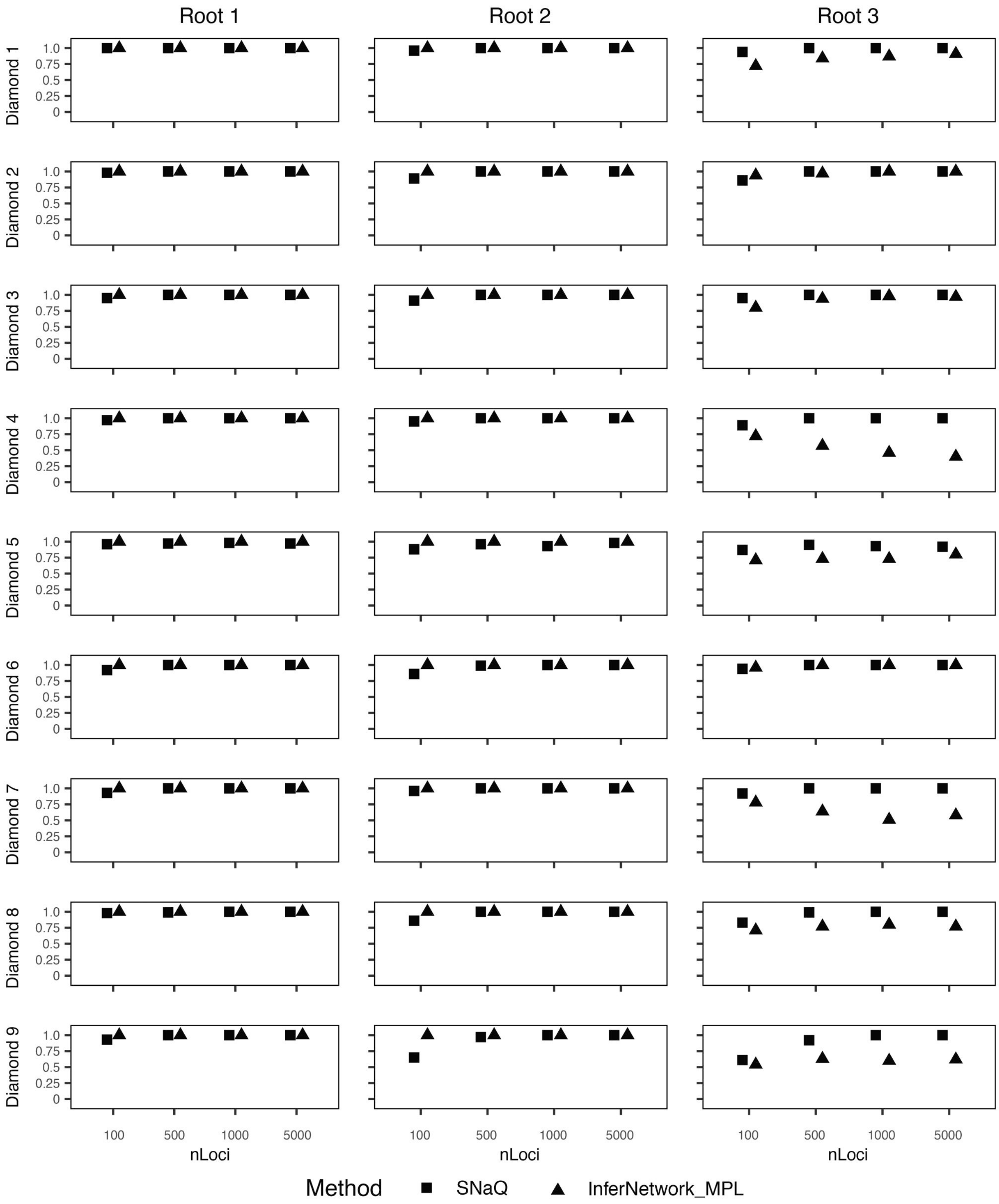


Figure 7. Proportion of 100 simulation replicates where the hybridization node was recovered by SNaQ (squares) or InferNetwork_MPL (triangles). The x-axis corresponds to the number of gene trees in the input sample. The diamond and root numbers correspond to Figure 4. Error was introduced into 10% of gene trees for each round of simulation by randomly re-rooting the gene tree and performing a random NNI move.

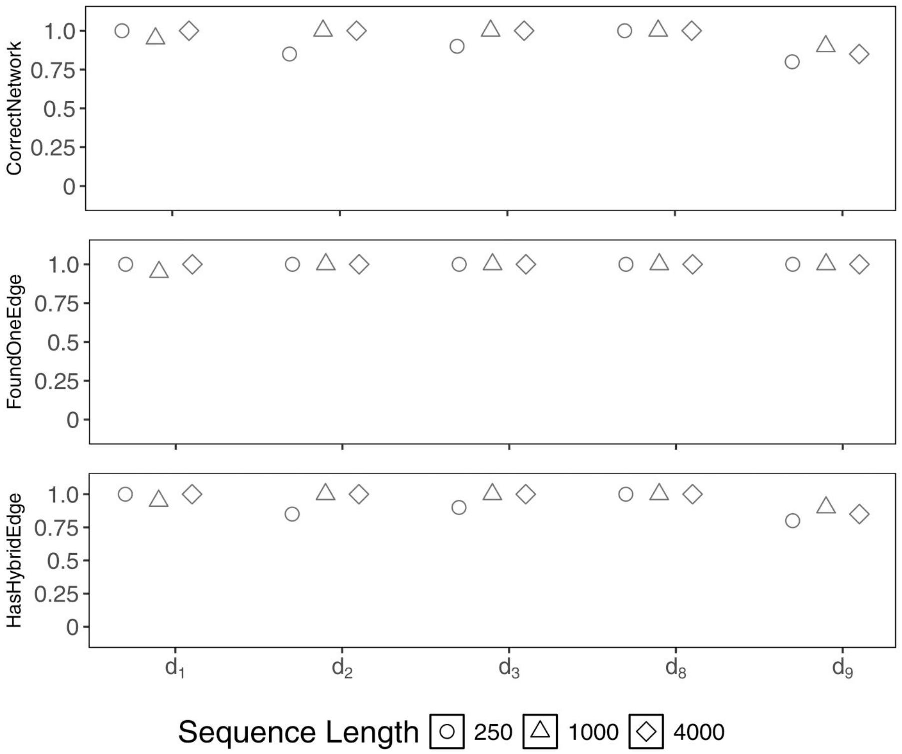


Figure 8. SNaQ results for root 3 with gene tree estimation error. *FoundOneEdge* is the proportion out of 20 replicates that one reticulate edge was correctly inferred by the pseudolikelihood scores. *CorrectNetwork* is the proportion of 20 replicates where the estimated topology when allowing only one hybridization event is identical to the true topology. *HasHybridEdge* is the proportion of 20 replicates where the correct hybridization edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one hybridization event. Diamonds correspond to Figure 4E.

diamonds performed poorly is unclear, but it could be an artifact of how errors were introduced to simulations. For example, d_4 and d_7 both have cherries for the descendants of the hybrid node. Since the outgroup errors were done randomly, this created more chances for a descendant of the hybrid node to be the outgroup. Outgroup lineages are typically sampled at an appropriate evolutionary distance to polarize phylogenetic relationships or site patterns in the context of *D*-statistics (Green *et al.* 2010, Durand *et al.* 2011, Pease and Hahn 2015), but our results show that rooting errors and potentially other sources of topological errors (e.g. deep paralogy, assembly, or genotyping, biases in starting material such as museum tissue vs. fresh) could be misleading for estimating level-1 networks with rooted triples. An unrooted quartet-based method such as SNaQ should be more robust to gene tree error, but rooted triple-based methods should perform well and can be beneficial for a low number of gene trees (e.g. 100) if gene tree quality can be reliably evaluated.

AUTHOR CONTRIBUTIONS

George Paul Tiley [Conceptualization (Equal], Formal Analysis (Equal), Methodology (Equal), Visualization (Equal), Writing—original draft (Equal), Writing—review & editing], Nan Liu [Formal Analysis (Equal), Investigation (Equal), Visualization (Equal), Writing—review & editing (Equal)], Claudia Solis-Lemus (Conceptualization (Equal), Formal Analysis (Equal), Funding Acquisition (Equal), Methodology (Equal), Project Administration (Equal), Writing—original draft (Equal), Writing—review & editing]

SUPPORTING INFORMATION

Supplementary data is available at *Evolutionary Journal of the Linnean Society* online.

CONFLICT OF INTEREST

None declared.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 101026923, awarded to G.P.T. This work was also funded by the National Science Foundation (DEB-2144367 to C.S-L.).

DATA AVAILABILITY

No new data were generated or analysed in support of this research. All simulated data was generated with the scripts publicly available in https://github.com/gtiley/diamond-identifiability.

REFERENCES

Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19:716–23.

Allman ES, Baños H, Garrote-Lopez M *et al.* Identifiability of level-1 species networks from gene tree quartets. ArXiv, https://arxiv.org/abs/2401.06290, 2024, preprint: not peer reviewed.

Allman ES, Baños H, Rhodes JA. Nanuq: a method for inferring species networks from gene trees under the coalescent model. *Algorithms for Molecular Biology* 2019;14:1–25.

Allman ES, Degnan JH, Rhodes JA. Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent. *Journal of Mathematical Biology* 2011;62:833–62.

Allman ES, Petrović S, Rhodes JA *et al. Identifiability of two-tree mixtures for group-based models.* IEEE/ACM Transactions on Computational Biology and Bioinformatics 2011;8:710–22.

Allman ES, Rhodes JA. The identifiability of tree topology for phylogenetic models, including covarion and mixture models. *Journal of Computational Biology: a Journal of Computational Molecular Cell Biology* 2006;13:1101–13.

Baños H. Identifying species network features from gene tree quartets under the coalescent model. *Bulletin of Mathematical Biology* 2019;81:494–534.

Baudry J-P, Maugis C, Michel B. Slope heuristics: overview and implementation. *Statistics and Computing* 2012;22:455–70.

Baum DA. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *TAXON* 2007;56:417–26. http://www.jstor.org/stable/25065797.

Cai R, Ané C. Assessing the fit of the multi-species network coalescent to multi-locus data. *Bioinformatics* 2021;37:634–41.

Chang JT. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Mathematical Biosciences* 1996;137:51–73.

Degnan JH. Modeling hybridization under the network multispecies coalescent. *Systematic Biology* 2018;67:786–99.

Durand EY, Patterson N, Reich D *et al. Testing for ancient admixture between closely related populations.* Molecular Biology and Evolution 2011;28:2239–52.

Flouri T, Jiao X, Rannala B *et al. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis.* Molecular Biology and Evolution 2020;37:1211–23.

Grayson DR, Stillman ME. Macaulay2, a software system for research in algebraic geometry, November 2024. https://macaulay2.com/

Green RE, Krause J, Briggs AW *et al. A draft sequence of the Neandertal genome.* Science 2010;328:710–22.

Gross E, Long C. Distinguishing phylogenetic networks. *SIAM Journal on Applied Algebra and Geometry* 2018;2:72–93.

Gross E, van Iersel L, Janssen R *et al. Distinguishing level-1 phylogenetic networks on the basis of data generated by Markov processes.* Journal of Mathematical Biology 2021;**83**: 32.

Grünewald S, Forslund K, Dress A *et al. QNet: an agglomerative method for the construction of phylogenetic networks from weighted quartets.* Molecular Biology and Evolution 2007;24:532–8.

Grunewald S, Spillner A, Bastkowski S *et al. SuperQ: computing supernetworks from quartets.* IEEE/ACM Transactions on Computational Biology and Bioinformatics 2013;10:151–60.

Hasegawa M, Kishino H, Yano T-A Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 1985;22:160–74.

Holland BR, Penny D, Hendy MD. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—a simulation study. *Systematic Biology* 2003;52:229–38.

Huber KT, Moulton V, Semple C *et al. Quarnet inference rules for level-1 networks.* Bulletin of Mathematical Biology 2018;80:2137–53.

Hudson RR. Testing the constant-rate neutral allele model with protein sequence data. *Evolution; International Journal of Organic Evolution* 1983;37:203–17. pages

Huelsenbeck JP, Bollback JP, Levine AM. Inferring the root of a phylogenetic tree. *Systematic Biology* 2002;51:32–43.

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 2006;23:254–67.

Huson DH, Rupp R, Scornavacca C. *Phylogenetic networks: concepts, algorithms and applications.* New York, USA: Cambridge University Press, 2010.

Jiao X, Flouri T, Yang Z. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *National Science Review* 2021;8:nwab127.

Kong S, Pons JC, Kubatko L *et al. Classes of explicit phylogenetic networks and their biological and mathematical significance.* Journal of Mathematical Biology 2022;84:47.

Leebens-Mack JH, Barker MS, Carpenter EJ *et al. One thousand plant transcriptomes and the phylogenomics of green plants.* Nature 2019;574:679–85.

Li Z, Tiley GP, Galuska SR *et al. Multiple large-scale gene and genome duplications during the evolution of hexapods.* Proceedings of the National Academy of Sciences 2018;115:4713–8.

Liu L, Yu L, Edwards SV. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* 2010;10:302–18.

Long C, Kubatko L. The effect of gene flow on coalescent-based species-tree inference. *Systematic Biology* 2018;67:770–85.

Lutteropp S, Scornavacca C, Kozlov AM *et al. NetRAX: accurate and fast maximum likelihood phylogenetic network inference.* Bioinformatics 2022;38:3725–33.

Maddison WP. Gene trees in species trees. *Systematic Biology* 1997;46:523–36.

Markin A, Anderson TK, Teja Vadali VSK, Eulenstein O. Robinson-foulds reticulation networks. In: *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*, Barcelona, Spain. 2019, 77–86.

Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 2020;37:2727–33. ISSN 0737-4038. doi: 10.1093/molbev/msaa106.

Oldman J, Wu T, Van Iersel L *et al. TriLoNet: piecing together small networks to reconstruct reticulate evolutionary histories.* Molecular Biology and Evolution 2016;33:2151–62.

Pardi F, Scornavacca C. Reconstructible phylogenetic networks: do not distinguish the indistinguishable. *PLoS Computational Biology* 2015;11:e1004135.

Pease JB, Hahn MW. Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology* 2015;64:651–62.

Pickrell J, Pritchard J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings* 2012;**8**: e1002967.

Rabier C-E, Berry V, Stoltz M *et al. On the inference of complex phylogenetic networks by Markov Chain Monte-Carlo.* PLoS Computational Biology 2021;17:e1008380.

Rannala B, Yang Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 2003;164:1645–56.

Rhodes JA, Sullivant S. Identifiability of large phylogenetic mixture models. *Bulletin of Mathematical Biology* 2012;74:212–31.

Rosselló F, Valiente G. All that glisters is not galled. *Mathematical Biosciences* 2009;221:54–9.

Solís-Lemus C, Ané C. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* 2016;12:e1005896.

Solís-Lemus C, Bastide P, Ané C. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* 2017;34:3292–8.

Solis-Lemus C, Coen A, Ane C. On the identifiability of phylogenetic networks under a pseudolikelihood model. Arxiv, https://arxiv.org/abs/2010.01758, October 2020, preprint: not peer reviewed.

Tan M, Long H, Liao B *et al. QS-Net: reconstructing phylogenetic networks based on quartet and sextet.* Frontiers in Genetics 2019;10:607.

Wen D, Yu Y, Zhu J *et al. Inferring phylogenetic networks using PhyloNet.* Systematic Biology 2018;67:735–40.

Willems M, Tahiri N, Makarenkov V. A new efficient algorithm for inferring explicit hybridization networks following the neighbor-joining principle. *Journal of Bioinformatics and Computational Biology* 2014;12:1450024.

Xu J, Ané C. Identifiability of local and global features of phylogenetic networks from average distances. *Journal of Mathematical Biology* 2022;86:12.

Yang J, Grünewald S, Wan X-F. Quartet-net: a quartet-based method to reconstruct phylogenetic networks. *Molecular Biology and Evolution* 2013;30:1206–17.

Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* 1994;39:306–14.

Yu Y, Degnan JH, Nakhleh L. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genetics* 2012;8:e1002660.

Yu Y, Dong J, Liu KJ *et al. Maximum likelihood inference of reticulate evolutionary histories.* Proceedings of the National Academy of Sciences of the United States of America 2014;111:16448–53.

Yu Y, Nakhleh L. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* 2015;16(Suppl 10): S10–10.

Zhang C, Ogilvie HA, Drummond AJ *et al. Bayesian inference of species networks from multilocus sequence data.* Molecular Biology and Evolution 2018;35:504–17.

Appendix A: Phylogenetic networks as a set of polynomial equations

A.1 CF equations for $\mathcal{N}_{down}$: $\boldsymbol{CF}(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$.

Table 1. CF equations for $\mathcal{N}_{down}$: $\boldsymbol{CF}(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$

n	**CF formula**
$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$
	$\frac{1}{3}z_2z_{23}z_3$
	$\frac{1}{3}z_2z_{23}z_3$
$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$
	$\frac{1}{3}z_{23}z_2$
	$\frac{1}{3}z_{23}z_2$
$(0,1,1,2)$	$1-\frac{2}{3}z_3$
	$\frac{1}{3}z_3$
	$\frac{1}{3}z_3$
$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$
	$\frac{1}{3}z_2z_{23}z_{13}z_1$
	$\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
	$\frac{1}{3}z_{13}z_1$
	$\frac{1}{3}z_{13}z_1$
$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
	$\frac{1}{3}z_3z_{13}z_1$
	$\frac{1}{3}z_3z_{13}z_1$
$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
	$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
	$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
	$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
	$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
	$(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
	$(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
	$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
	$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$

(continued)

Table 1. (continued)

n	CF formula
$(2,0,2,0)$	$(1-\gamma)^2\left(1-\tfrac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\tfrac{2}{3}z_2z_0\right)+\gamma^2\left(1-\tfrac{2}{3}z_2z_0z_{02}\right)$
	$(1-\gamma)^2\tfrac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\tfrac{1}{3}z_2z_0+\gamma^2\tfrac{1}{3}z_2z_0z_{02}$
	$(1-\gamma)^2\tfrac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\tfrac{1}{3}z_2z_0+\gamma^2\tfrac{1}{3}z_2z_0z_{02}$
$(2,0,1,1)$	$(1-\gamma)^2\left(1-\tfrac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\tfrac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\tfrac{2}{3}z_0z_{02}\right)$
	$(1-\gamma)^2\tfrac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\tfrac{1}{3}z_{23}\right)+\gamma^2\tfrac{1}{3}z_0z_{02}$
	$(1-\gamma)^2\tfrac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\tfrac{1}{3}z_{23}\right)+\gamma^2\tfrac{1}{3}z_0z_{02}$
$(2,0,0,2)$	$(1-\gamma)^2\left(1-\tfrac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\tfrac{2}{3}z_3z_0\right)+\gamma^2\left(1-\tfrac{2}{3}z_3z_0z_{23}z_{02}\right)$
	$(1-\gamma)^2\tfrac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\tfrac{1}{3}z_3z_0+\gamma^2\tfrac{1}{3}z_3z_0z_{23}z_{02}$
	$(1-\gamma)^2\tfrac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\tfrac{1}{3}z_3z_0+\gamma^2\tfrac{1}{3}z_3z_0z_{23}z_{02}$
$(2,1,1,0)$	$(1-\gamma)^2\left(1-\tfrac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\tfrac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\tfrac{2}{3}z_0z_{02}\right)$
	$(1-\gamma)^2\tfrac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\tfrac{1}{3}z_{23}z_{13}\right)+\gamma^2\tfrac{1}{3}z_0z_{02}$
	$(1-\gamma)^2\tfrac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\tfrac{1}{3}z_{23}z_{13}\right)+\gamma^2\tfrac{1}{3}z_0z_{02}$
$(2,1,0,1)$	$(1-\gamma)^2\left(1-\tfrac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\tfrac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\tfrac{2}{3}z_0z_{02}z_{23}\right)$
	$(1-\gamma)^2\tfrac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\tfrac{1}{3}z_{13}\right)+\gamma^2\tfrac{1}{3}z_0z_{02}z_{23}$
	$(1-\gamma)^2\tfrac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\tfrac{1}{3}z_{13}\right)+\gamma^2\tfrac{1}{3}z_0z_{02}z_{23}$
$(2,2,0,0)$	$(1-\gamma)^2\left(1-\tfrac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\tfrac{2}{3}z_1z_0\right)+\gamma^2\left(1-\tfrac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$
	$(1-\gamma)^2\tfrac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\tfrac{1}{3}z_1z_0+\gamma^2\tfrac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
	$(1-\gamma)^2\tfrac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\tfrac{1}{3}z_1z_0+\gamma^2\tfrac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,1,1,1)$	$(1-\gamma)\left(1-\tfrac{2}{3}z_{13}\right)+\gamma\tfrac{1}{3}z_{23}$
	$(1-\gamma)\tfrac{1}{3}z_{13}+\gamma\left(1-\tfrac{2}{3}z_{23}\right)$
	$(1-\gamma)\tfrac{1}{3}z_{13}+\gamma\tfrac{1}{3}z_{23}$

A.2 Definition of CF equations for network obtained by the rotation of hybrid in hybridization cycle.

We start by noticing that there is a correspondence between a quartet in the network $\mathcal{N}_{down}$ and a quartet in any other network. For example, in Figure 9, we want to match the equations corresponding to $n=(0,1,2,1)$ in $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$. We can see that the equations for the quartet $n=(0,1,2,1)$ in $\mathcal{N}_{right}$ can be obtained from the equations in quartet $n=(2,0,1,1)$ in $\mathcal{N}_{down}$.

Therefore, in order to get the equations for $\mathcal{N}_{right}$, we need to identify to which quartets they correspond in the $\mathcal{N}_{down}$.

Note that only the last quartet $(1,1,1,1)$ is not a mere rotation. We have to compute the CFs for the specific network (Fig. 10).

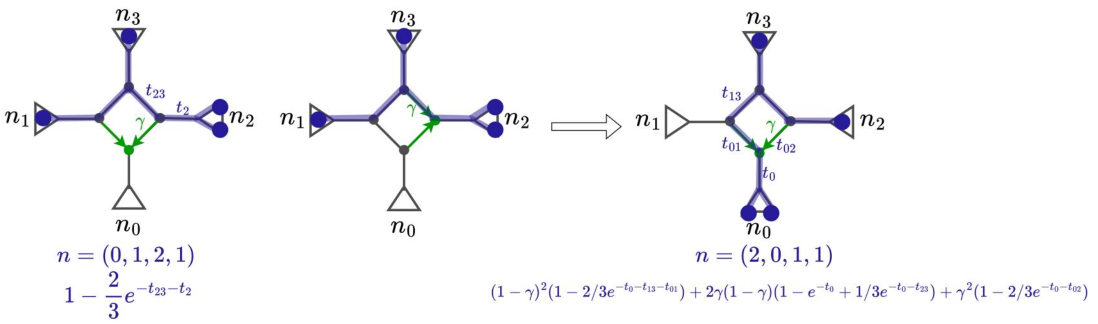
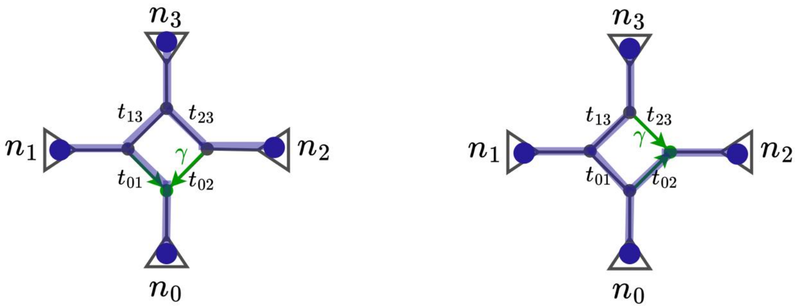


Figure 9. Correspondence between the quartet $n=(0,1,2,1)$ in $\mathcal{N}_{down}$ (left), the same quartet in $\mathcal{N}_{right}$ (centre), and the quartet $n=(2,0,1,1)$ in $\mathcal{N}_{down}$ (right).



$$CF_{01|23} = (1-\gamma)(1-2/3e^{-t_{13}}) + \gamma 1/3e^{-t_{23}} \quad CF_{01|23} = (1-\gamma)1/3e^{-t_{01}} + \gamma(1-2/3e^{-t_{13}})$$
$$CF_{02|13} = (1-\gamma)1/3e^{-t_{13}} + \gamma(1-2/3e^{-t_{23}}) \quad CF_{02|13} = (1-\gamma)(1-2/3e^{-t_{01}}) + \gamma 1/3e^{-t_{13}}$$
$$CF_{03|12} = (1-\gamma)1/3e^{-t_{13}} + \gamma 1/3e^{-t_{23}} \quad CF_{03|12} = (1-\gamma)1/3e^{-t_{01}} + \gamma 1/3e^{-t_{13}}$$

Figure 10. CF equations for quartet $(1,1,1,1)$ comparing the $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$ networks.

A.3 CF equations for $\mathcal{N}_{right}$: $\boldsymbol{CF}(\mathcal{N}_{right}, \boldsymbol{z}, \boldsymbol{\gamma})$.

Table 2. CF equations for $\mathcal{N}_{right}$: $\boldsymbol{CF}(\mathcal{N}_{right}, \boldsymbol{z}, \boldsymbol{\gamma})$

$\mathbf{n} \in \mathcal{N}_{right}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,0,2,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(0,1,2,1)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,1,1,2)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(0,2,0,2)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(1,0,2,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,0,1,2)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,2,0)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$

(continued)

Table 2. (continued)

$\mathbf{n} \in \mathcal{N}_{right}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(1,1,0,2)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$
		$\frac{1}{3}z_{23}z_2$
		$\frac{1}{3}z_{23}z_2$
$(1,2,1,0)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,0,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$
		$\frac{1}{3}z_3$
		$\frac{1}{3}z_3$
$(2,0,2,0)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(2,0,1,1)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(2,0,0,2)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$
		$\frac{1}{3}z_2z_{23}z_{13}z_1$
		$\frac{1}{3}z_2z_{23}z_{13}z_1$
$(2,1,1,0)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(2,1,0,1)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
$(2,2,0,0)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{01}+\gamma(1-\frac{2}{3}z_{13})$
		$(1-\gamma)(1-\frac{2}{3}z_{01})+\gamma\frac{1}{3}z_{13}$
		$(1-\gamma)\frac{1}{3}z_{01}+\gamma\frac{1}{3}z_{13}$

A.4 CF equations for $\mathcal{N}_{left}$: $CF(\mathcal{N}_{left}, \boldsymbol{z}, \boldsymbol{\gamma})$.

Table 3. CF equations for $\mathcal{N}_{left}$: $CF(\mathcal{N}_{left}, \boldsymbol{z}, \boldsymbol{\gamma})$

$\mathbf{n} \in \mathcal{N}_{left}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,0,2,2)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
$(0,1,2,1)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(0,1,1,2)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(0,2,0,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,0,2,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$
		$\frac{1}{3}z_3$
		$\frac{1}{3}z_3$
$(1,0,1,2)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
$(1,1,2,0)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,0,2)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,1,0)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,0,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(2,0,2,0)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$
		$\frac{1}{3}z_2z_{23}z_3$
		$\frac{1}{3}z_2z_{23}z_3$

(continued)

Table 3. (continued)

$\mathbf{n} \in \mathcal{N}_{left}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(2,0,1,1)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(2,0,0,2)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(2,1,1,0)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(2,1,0,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(2,2,0,0)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{23}+\gamma(1-\frac{2}{3}z_{02})$ $(1-\gamma)(1-\frac{2}{3}z_{23})+\gamma\frac{1}{3}z_{02}$ $(1-\gamma)\frac{1}{3}z_{23}+\gamma\frac{1}{3}z_{02}$

A.5 CF equations for $\mathcal{N}_{up}$: $CF(\mathcal{N}_{up}, z, \gamma)$.

Table 4. CF equations for $\mathcal{N}_{up}$: $CF(\mathcal{N}_{up}, z, \gamma)$

$\mathbf{n} \in \mathcal{N}_{\mathbf{up}}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,0,2,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(0,1,2,1)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(0,1,1,2)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,2,2,0)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,0,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$

(continued)

Table 4. (continued)

$\mathbf{n} \in \mathcal{N}_{\mathbf{up}}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(1,0,2,1)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,0,1,2)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,2,0)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
$(1,1,0,2)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,1,0)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$
		$\frac{1}{3}z_{23}z_2$
		$\frac{1}{3}z_{23}z_2$
$(1,2,0,1)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
		$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
		$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(2,0,2,0)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
$(2,0,1,1)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(2,0,0,2)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(2,1,1,0)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$
		$\frac{1}{3}z_3$
		$\frac{1}{3}z_3$
$(2,1,0,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(2,2,0,0)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$
		$\frac{1}{3}z_2z_{23}z_3$
		$\frac{1}{3}z_2z_{23}z_3$
$(1,1,1,1)$		$(1-\gamma)(1-\frac{2}{3}z_{02})+\gamma\frac{1}{3}z_{01}$
		$(1-\gamma)\frac{1}{3}z_{02}+\gamma(1-\frac{2}{3}z_{01})$
		$(1-\gamma)\frac{1}{3}z_{02}+\gamma\frac{1}{3}z_{01}$

Appendix B: Phylogenetic networks as a set of polynomial equations: $n = 6$ taxa

B.1 CF equations for 4-cycle network $\mathcal{N}_{down}$ for $N = 1212$ (Diamond 6).

Table 5. CF equations for 4-cycle network $\mathcal{N}_{down}$ for $N = 1212$ (Diamond 6)

n	CF formula
$(0,1,1,2)$	$1-\frac{2}{3}z_3$
	$\frac{1}{3}z_3$
	$\frac{1}{3}z_3$
$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
	$\frac{1}{3}z_{13}z_1$
	$\frac{1}{3}z_{13}z_1$
$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
	$\frac{1}{3}z_3z_{13}z_1$
	$\frac{1}{3}z_3z_{13}z_1$
$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
	$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
	$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
	$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
	$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,1,1,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}\right)+\gamma\frac{1}{3}z_{23}$
	$(1-\gamma)\frac{1}{3}z_{13}+\gamma\left(1-\frac{2}{3}z_{23}\right)$
	$(1-\gamma)\frac{1}{3}z_{13}+\gamma\frac{1}{3}z_{23}$

B.2 CF equations for 4-cycle network $\mathcal{N}_{right}$ for $N = 1122$ (Diamond 9).

Table 6. CF equations for 4-cycle network $\mathcal{N}_{right}$ for $N = 1122$ (Diamond 9)

$\mathbf{n} \in \mathcal{N}_{right}$	$\mathbf{n} \in \mathcal{N}_{down}$	CF formula
$(0,1,1,2)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
		$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
		$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,1,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$

(continued)

Table 6. (continued)

$\mathbf{n} \in \mathcal{N}_{right}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,2,0,2)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$
		$\frac{1}{3}z_2z_{23}z_3$
		$\frac{1}{3}z_2z_{23}z_3$
$(1,0,1,2)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
		$(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
		$(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,0,2)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$
		$\frac{1}{3}z_{23}z_2$
		$\frac{1}{3}z_{23}z_2$
$(1,2,1,0)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,0,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$
		$\frac{1}{3}z_3$
		$\frac{1}{3}z_3$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{01}+\gamma(1-\frac{2}{3}z_{13})$
		$(1-\gamma)(1-\frac{2}{3}z_{01})+\gamma\frac{1}{3}z_{13}$
		$(1-\gamma)\frac{1}{3}z_{01}+\gamma\frac{1}{3}z_{13}$

B.3 CF equations for 4-cycle network $\mathcal{N}_{left}$ for $N = 2211$ (Diamond 7).

Table 7. CF equations for 4-cycle network $\mathcal{N}_{left}$ for $N = 2211$ (Diamond 7)

$\mathbf{n} \in \mathcal{N}_{left}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,1,1,2)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(0,2,1,1)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(0,2,0,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,0,1,2)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
$(1,1,0,2)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$

(continued)

Table 7. (continued)

$\mathbf{n} \in \mathcal{N}_{left}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(1,2,1,0)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,0,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{23}+\gamma(1-\frac{2}{3}z_{02})$ $(1-\gamma)(1-\frac{2}{3}z_{23})+\gamma\frac{1}{3}z_{02}$ $(1-\gamma)\frac{1}{3}z_{23}+\gamma\frac{1}{3}z_{02}$

B.4 CF equations for 4-cycle network $\mathcal{N}_{up}$ for $N = 2121$ (Diamond 8).

Table 8. CF equations for 4-cycle network $\mathcal{N}_{up}$ for $N = 2121$ (Diamond 8)

$\mathbf{n} \in \mathcal{N}_{\mathbf{up}}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,1,1,2)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,2,1,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,0,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,0,1,2)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,0,2)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,1,0)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,0,1)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,1,1)$		$(1-\gamma)(1-\frac{2}{3}z_{02})+\gamma\frac{1}{3}z_{01}$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma(1-\frac{2}{3}z_{01})$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma\frac{1}{3}z_{01}$

Appendix C: Phylogenetic networks as a set of polynomial equations: $n = 7$ taxa

C.1 CF equations for $\mathcal{N}_{down}$ for $N = 1222$ (Diamond 2).

Table 9. CF equations for $\mathcal{N}_{down}$ for $N = 1222$ (Diamond 2)

n	**CF formula**
$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$
	$\frac{1}{3}z_2z_{23}z_3$
	$\frac{1}{3}z_2z_{23}z_3$
$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$
	$\frac{1}{3}z_{23}z_2$
	$\frac{1}{3}z_{23}z_2$
$(0,1,1,2)$	$1-\frac{2}{3}z_3$
	$\frac{1}{3}z_3$
	$\frac{1}{3}z_3$
$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$
	$\frac{1}{3}z_2z_{23}z_{13}z_1$
	$\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
	$\frac{1}{3}z_{13}z_1$
	$\frac{1}{3}z_{13}z_1$
$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
	$\frac{1}{3}z_3z_{13}z_1$
	$\frac{1}{3}z_3z_{13}z_1$
$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
	$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
	$(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
	$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
	$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$
	$(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
	$(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
	$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
	$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
	$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,1,1,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}\right)+\gamma\frac{1}{3}z_{23}$
	$(1-\gamma)\frac{1}{3}z_{13}+\gamma\left(1-\frac{2}{3}z_{23}\right)$
	$(1-\gamma)\frac{1}{3}z_{13}+\gamma\frac{1}{3}z_{23}$

C.2 CF equations for $\mathcal{N}_{right}$ for $N = 2122$ (Diamond 5).

Table 10. CF equations for $\mathcal{N}_{right}$ for $N = 2122$ (Diamond 5)

$\mathbf{n} \in \mathcal{N}_{right}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,0,2,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(0,1,2,1)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,1,1,2)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(0,2,0,2)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(1,0,2,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,0,1,2)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,2,0)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,0,2)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,1,0)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,0,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{01}+\gamma(1-\frac{2}{3}z_{13})$ $(1-\gamma)(1-\frac{2}{3}z_{01})+\gamma\frac{1}{3}z_{13}$ $(1-\gamma)\frac{1}{3}z_{01}+\gamma\frac{1}{3}z_{13}$

C.3 CF equations for $\mathcal{N}_{left}$ for $N = 2212$ (Diamond 3).

Table 11. CF equations for $\mathcal{N}_{left}$ for $N = 2212$ (Diamond 3)

$\mathbf{n} \in \mathcal{N}_{left}$	$\mathbf{n} \in \mathcal{N}_{down}$	**CF formula**
$(0,0,2,2)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
		$\frac{1}{3}z_3z_{13}z_1$
$(0,1,2,1)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
		$(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(0,1,1,2)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(0,2,0,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
		$(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,0,2,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$
		$\frac{1}{3}z_3$
		$\frac{1}{3}z_3$
$(1,0,1,2)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
		$\frac{1}{3}z_{13}z_1$
$(1,1,2,0)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
		$(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,0,2)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,1,0)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,0,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
		$(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{23}+\gamma(1-\frac{2}{3}z_{02})$
		$(1-\gamma)(1-\frac{2}{3}z_{23})+\gamma\frac{1}{3}z_{02}$
		$(1-\gamma)\frac{1}{3}z_{23}+\gamma\frac{1}{3}z_{02}$

C.4 CF equations for $\mathcal{N}_{up}$ for $N = 2221$ (Diamond 4).

Table 12. CF equations for $\mathcal{N}_{up}$ for $N = 2221$ (Diamond 4)

$\mathbf{n} \in \mathcal{N}_{\mathbf{up}}$	$\mathbf{n} \in \mathcal{N}_{down}$	CF formula
$(0,0,2,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(0,1,2,1)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(0,1,1,2)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,2,2,0)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,0,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,0,2,1)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,0,1,2)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,2,0)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(1,1,0,2)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,1,0)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,0,1)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,1,1)$		$(1-\gamma)(1-\frac{2}{3}z_{02})+\gamma\frac{1}{3}z_{01}$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma(1-\frac{2}{3}z_{01})$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma\frac{1}{3}z_{01}$

Appendix D: Theorems and proofs for generic identifiability of 4-node cycles in semi-directed level-1 phylogenetic networks

Theorem 4. *Let $\mathcal{N}_{down}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes. Without loss of generality, let the taxa be partitioned among clades as $n_0 = 1, n_1 = 2, n_2 = 1, n_3 = 2$ (Fig. 3). Let the hybrid node be ancestral to the clade n_0. Let $\mathcal{N}_{left}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes such that the unrooted version of $\mathcal{N}_{left}$ agrees with the unrooted version of $\mathcal{N}_{down}$. Let the hybrid node in the hybridization cycle in $\mathcal{N}_{left}$ be ancestral to the clade n_1. Then, $\mathcal{N}_{down}$ and $\mathcal{N}_{left}$ are generically identifiable if $t_1 < \infty, t_{13} > 0, t_3 < \infty$, and $\gamma \in (0,1)$.*

Proof.

The structure of the proof is the same as for Theorem 3, but we repeat it for completeness. Let $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ be the system of CF polynomial equations defined by $\mathcal{N}_{down}$ and let $CF(\mathcal{N}_{left}, \boldsymbol{z}', \boldsymbol{\gamma}')$ be the system of CF polynomial equations defined by $\mathcal{N}_{left}$. Both systems of equations can be found in the Appendix (Tables 1–14).

Let $\mathcal{P} = \{p(\boldsymbol{z}, \boldsymbol{\gamma}) - q(\boldsymbol{z}', \boldsymbol{\gamma}') : p(\boldsymbol{z}, \boldsymbol{\gamma}) \in CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma}),\ q(\boldsymbol{z}', \boldsymbol{\gamma}') \in CF(\mathcal{N}_{left}, \boldsymbol{z}', \boldsymbol{\gamma}')\}$ be the set of polynomial equations resulting from matching $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ to $CF(\mathcal{N}_{left}, \boldsymbol{z}', \boldsymbol{\gamma}')$.

Using Macaulay2 (Grayson and Stillman undated), we compute the Gröbner basis of $\mathcal{P}$ on the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables by any elimination order. All Macaulay2 scripts are available at https://github.com/gtiley/diamond-identifiability.

Let $\mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}') : \boldsymbol{z} \in [0,\infty)^{n-3}, \boldsymbol{\gamma} \in [0,1]^h, \boldsymbol{z}' \in [0,\infty)^{n-3}, \boldsymbol{\gamma}' \in [0,1]^h]$ be the set of all polynomials on the $(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')$ variables.

The resulting ideal is given by:

$$\{(z_1)(z_{13}-1)(z_3)(\gamma_z)(\gamma_z-1)p_i(\boldsymbol{z}, \boldsymbol{\gamma}) : p_i \in \mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')]; \deg(p_i) = 7; i = 1, \ldots, 8\}$$

which represents the conditions that the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables need to satisfy for the polynomial set $\mathcal{P}$ to vanish to zero.

As the polynomials $\{p_i(\boldsymbol{z}, \boldsymbol{\gamma})\}_{i=1}^{8}$ have the Lebesgue measure zero, $\mathcal{N}_{down}$ and $\mathcal{N}_{left}$ are not generically identifiable in the subset of parameter space corresponding to $\{z_1 = 0\} \cup \{z_{13} = 1\} \cup \{z_3 = 0\} \cup \{\gamma = 0\} \cup \{\gamma = 1\}$. ■.

Theorem 5. *Let $\mathcal{N}_{down}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes. Without loss of generality, let the taxa be partitioned among clades as $n_0 = 1, n_1 = 2, n_2 = 1, n_3 = 2$ (Fig. 3). Let the hybrid node be ancestral to the clade n_0. Let $\mathcal{N}_{up}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes such that the unrooted version of $\mathcal{N}_{up}$ agrees with the unrooted version of $\mathcal{N}_{down}$. Let the hybrid node in the hybridization cycle in $\mathcal{N}_{up}$ be ancestral to the clade n_3. Then, $\mathcal{N}_{down}$ and $\mathcal{N}_{up}$ are generically identifiable if $t_1 < \infty, t_{13} > 0, t_3 < \infty$, and $\gamma \in (0,1)$.*

Proof.

The structure of the proof is the same as for Theorem 4, but we repeat it for completeness. Let $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ be the system of CF polynomial equations defined by $\mathcal{N}_{down}$ and let $CF(\mathcal{N}_{up}, \boldsymbol{z}', \boldsymbol{\gamma}')$ be the system of CF polynomial equations defined by $\mathcal{N}_{up}$. Both systems of equations can be found in the Appendix.

Let $\mathcal{P} = \{p(\boldsymbol{z}, \boldsymbol{\gamma}) - q(\boldsymbol{z}', \boldsymbol{\gamma}') : p(\boldsymbol{z}, \boldsymbol{\gamma}) \in CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma}), q(\boldsymbol{z}', \boldsymbol{\gamma}') \in CF(\mathcal{N}_{up}, \boldsymbol{z}', \boldsymbol{\gamma}')\}$ be the set of polynomial equations resulting from matching $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ to $CF(\mathcal{N}_{up}, \boldsymbol{z}', \boldsymbol{\gamma}')$.

Using Macaulay2 (Grayson and Stillman undated), we compute the Gröbner basis of $\mathcal{P}$ on the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables by any elimination order. All Macaulay2 scripts are available at https://github.com/gtiley/diamond-identifiability.

Let $\mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}') : \boldsymbol{z} \in [0,\infty)^{n-3}, \boldsymbol{\gamma} \in [0,1]^h, \boldsymbol{z}' \in [0,\infty)^{n-3}, \boldsymbol{\gamma}' \in [0,1]^h]$ be the set of all polynomials on the $(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')$ variables.

The resulting ideal is given by:

$$\{(z_1)(z_{13}-1)(z_3)(\gamma_z)(\gamma_z-1)p_i(\boldsymbol{z}, \boldsymbol{\gamma}) : p_i \in \mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')]; \deg(p_i) = 10; i = 1, \ldots, 12\}$$

which represents the conditions that the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables need to satisfy for the polynomial set $\mathcal{P}$ to vanish to zero.

As the polynomials $\{p_i(\boldsymbol{z}, \boldsymbol{\gamma})\}_{i=1}^{12}$ have the Lebesgue measure zero, $\mathcal{N}_{down}$ and $\mathcal{N}_{up}$ are not generically identifiable in the subset of parameter space corresponding to $\{z_1 = 0\} \cup \{z_{13} = 1\} \cup \{z_3 = 0\} \cup \{\gamma = 0\} \cup \{\gamma = 1\}$. ■.

Appendix E: Subset of CF equations used in Macaulay2

We present below the table with the major CF for all 4-taxon subsets, except (1,1,1,1) for which we present all equations. All equations are multiplied by 3 for conciseness. All Macaulay2 scripts can be found in https://github.com/gtiley/diamond-identifiability/tree/main/scripts/macaulay2.

Table 13. Subset of CF equations used in Macaulay2 (Part 1)

$(\mathbf{n_0},\mathbf{n_1},\mathbf{n_2},\mathbf{n_3})$	$\mathcal{N}_{down}$	$\mathcal{N}_{right}$
$(0,0,2,2)$	$3-2z_2z_{23}z_3$	$(1-\gamma_u)^2(3-2u_2u_0u_{01}u_{13}u_{23})$ $+2\gamma_u(1-\gamma_u)(3-2u_2u_0)$ $+\gamma_u^2(3-2u_2u_0u_{02})$
$(0,1,2,1)$	$3-2z_{23}z_2$	$(1-\gamma_u)^2(3-2u_0u_{13}u_{01})$ $+2\gamma_u(1-\gamma_u)(3-3u_0+u_0u_{23})$ $+\gamma_u^2(3-2u_0u_{02})$
$(0,1,1,2)$	$3-2z_3$	$(1-\gamma_u)(3-2u_{23}u_2)+\gamma_u(3-2u_2)$
$(0,2,2,0)$	$3-2z_2z_{23}z_{13}z_1$	$(1-\gamma_u)^2(3-2u_3u_0u_{13}u_{01})$ $+2\gamma_u(1-\gamma_u)(3-2u_3u_0)$ $+\gamma_u^2(3-2u_3u_0u_{23}u_{02})$
$(0,2,1,1)$	$3-2z_{13}z_1$	$(1-\gamma_u)(3-2z_3)+\gamma_u(3-2z_{23}z_3)$
$(0,2,0,2)$	$3-2z_3z_{13}z_1$	$3-2z_2z_{23}z_3$
$(1,0,2,1)$	$(1-\gamma)(3-2z_{23}z_2)+\gamma(3-2z_2)$	$(1-\gamma_u)^2(3-2u_0u_{01})$ $+2\gamma_u(1-\gamma_u)(3-3u_0+u_0u_{23}u_{13})$ $+\gamma_u^2(3-2u_0u_{02})$
$(1,0,1,2)$	$(1-\gamma)(3-2z_3)+\gamma(3-2z_{23}z_3)$	$(1-\gamma_u)(3-2u_{13}u_{23}u_2)+\gamma_u(3-2u_2)$
$(1,1,2,0)$	$(1-\gamma)(3-2z_{13}z_{23}z_2)+\gamma(3-2z_2)$	$(1-\gamma_u)^2(3-2u_0u_{01})$ $+2\gamma_u(1-\gamma_u)(3-3u_0+u_0u_{13})$ $+\gamma_u^2(3-2u_0u_{02}u_{23})$
$(1,1,0,2)$	$(1-\gamma)(3-2z_{13}z_3)+\gamma(3-2z_3)$	$3-2u_{23}u_2$
$(1,2,1,0)$	$(1-\gamma)(3-2z_1)+\gamma(3-2z_{23}z_{13}z_1)$	$(1-\gamma_u)(3-2u_{13}u_3)+\gamma_u(3-2u_3)$
$(1,2,0,1)$	$(1-\gamma)(3-2z_1)+\gamma(3-2z_{13}z_1)$	$3-2u_3$
$(2,0,2,0)$	$(1-\gamma)^2(3-2z_2z_0z_{01}z_{13}z_{23})$ $+2\gamma(1-\gamma)(3-2z_2z_0)$ $+\gamma^2(3-2z_2z_0z_{02})$	$(1-\gamma_u)^2(3-2u_1u_0u_{01})$ $+2\gamma_u(1-\gamma_u)(3-2u_1u_0)$ $+\gamma_u^2(3-2u_1u_0u_{02}u_{23}u_{13})$
$(2,0,1,1)$	$(1-\gamma)^2(3-2z_0z_{13}z_{01})$ $+2\gamma(1-\gamma)(3-3z_0+z_0z_{23})$ $+\gamma^2(3-2z_0z_{02})$	$(1-\gamma_u)(3-2u_1)+\gamma_u(3-2u_{23}u_{13}u_1)$
$(2,0,0,2)$	$(1-\gamma)^2(3-2z_3z_0z_{13}z_{01})$ $+2\gamma(1-\gamma)(3-2z_3z_0)$ $+\gamma^2(3-2z_3z_0z_{23}z_{02})$	$3-2u_2u_{23}u_{13}u_1$
$(2,1,1,0)$	$(1-\gamma)^2(3-2z_0z_{01})$ $+2\gamma(1-\gamma)(3-3z_0+z_0z_{23}z_{13})$ $+\gamma^2(3-2z_0z_{02})$	$(1-\gamma_u)(3-2u_1)+\gamma_u(3-2u_{13}u_1)$

(continued)

Table 13. (continued)

$(\mathbf{n_0}, \mathbf{n_1}, \mathbf{n_2}, \mathbf{n_3})$	$\mathcal{N}_{down}$	$\mathcal{N}_{right}$
$(2, 1, 0, 1)$	$(1-\gamma)^2(3-2z_0z_{01})$ $+2\gamma(1-\gamma)(3-3z_0+z_0z_{13})$ $+\gamma^2(3-2z_0z_{02}z_{23})$	$3-2u_{13}u_1$
$(2, 2, 0, 0)$	$(1-\gamma)^2(3-2z_1z_0z_{01})$ $+2\gamma(1-\gamma)(3-2z_1z_0)$ $+\gamma^2(3-2z_1z_0z_{02}z_{23}z_{13})$	$3-2u_3u_{13}u_1$
$(1, 1, 1, 1)$	$(1-\gamma)(3-2z_{13})+\gamma z_{23}$	$(1-\gamma_u)u_{01}+\gamma_u(3-2u_{13})$
	$(1-\gamma)z_{13}+\gamma(3-2z_{23})$	$(1-\gamma_u)(3-2u_{01})+\gamma_u u_{13}$
	$(1-\gamma)z_{13}+\gamma z_{23}$	$(1-\gamma_u)u_{01}+\gamma_u u_{13}$

Table 14. Subset of CF equations used in Macaulay2 (Part 2)

$(\mathbf{n_0}, \mathbf{n_1}, \mathbf{n_2}, \mathbf{n_3})$	$\mathcal{N}_{left}$	$\mathcal{N}_{\mathbf{up}}$
$(0, 0, 2, 2)$	$3-2v_3v_{13}v_1$	$(1-\gamma_w)^2(3-2w_1w_0w_{01})$ $+2\gamma_w(1-\gamma_w)(3-2w_1w_0)$ $+\gamma_w^2(3-2w_1w_0w_{02}w_{23}w_{13})$
$(0, 1, 2, 1)$	$(1-\gamma_v)(3-2v_{13}v_3)+\gamma_v(3-2v_3)$	$(1-\gamma_w)(3-2w_1)+\gamma_w(3-2w_{23}w_{13}w_1)$
$(0, 1, 1, 2)$	$(1-\gamma_v)(3-2v_1)+\gamma_v(3-2v_{13}v_1)$	$(1-\gamma_w)^2(3-2w_0w_{01})$ $+2\gamma_w(1-\gamma_w)(3-3w_0+w_0w_{23}w_{13})$ $+\gamma_w^2(3-2w_0w_{02})$
$(0, 2, 2, 0)$	$(1-\gamma_v)^2(3-2v_3v_0v_{13}v_{01})$ $+2\gamma_v(1-\gamma_v)(3-2v_3v_0)$ $+\gamma_v^2(3-2v_3v_0v_{23}v_{02})$	$3-2w_2w_{23}w_{13}w_1$
$(0, 2, 1, 1)$	$(1-\gamma_v)^2(3-2v_0v_{01})$ $+2\gamma_v(1-\gamma_v)(3-3v_0+v_0v_{13})$ $+\gamma_v^2(3-2v_0v_{02}v_{23})$	$(1-\gamma_w)(3-2w_{13}w_{23}w_2)+\gamma_w(3-2w_2)$
$(0, 2, 0, 2)$	$(1-\gamma_v)^2(3-2v_1v_0v_{01})$ $+2\gamma_v(1-\gamma_v)(3-2v_1v_0)$ $+\gamma_v^2(3-2v_1v_0v_{02}v_{23}v_{13})$	$(1-\gamma_w)^2(3-2w_2w_0w_{01}w_{13}w_{23})$ $+2\gamma_w(1-\gamma_w)(3-2w_2w_0)$ $+\gamma_w^2(3-2w_2w_0w_{02})$
$(1, 0, 2, 1)$	$3-2v_3$	$(1-\gamma_w)(3-2w_1)+\gamma_w(3-2w_{13}w_1)$
$(1, 0, 1, 2)$	$3-2v_{13}v_1$	$(1-\gamma_w)^2(3-2w_0w_{01})$ $+2\gamma_w(1-\gamma_w)(3-3w_0+w_0w_{13})$ $+\gamma_w^2(3-2w_0w_{02}w_{23})$
$(1, 1, 2, 0)$	$(1-\gamma_v)(3-2v_3)+\gamma_v(3-2v_{23}v_3)$	$3-2w_{13}w_1$
$(1, 1, 0, 2)$	$(1-\gamma_v)(3-2v_1)+\gamma_v(3-2v_{23}v_{13}v_1)$	$(1-\gamma_w)^2(3-2w_0w_{13}w_{01})$ $+2\gamma_w(1-\gamma_w)(3-3w_0+w_0w_{23})$ $+\gamma_w^2(3-2w_0w_{02})$
$(1, 2, 1, 0)$	$(1-\gamma_v)^2(3-2v_0v_{13}v_{01})$ $+2\gamma_v(1-\gamma_v)(3-3v_0+v_0v_{23})$ $+\gamma_v^2(3-2v_0v_{02})$	$3-2w_{23}w_2$

(continued)

Table 14. (continued)

$(\mathbf{n_0}, \mathbf{n_1}, \mathbf{n_2}, \mathbf{n_3})$	$\mathcal{N}_{left}$	$\mathcal{N}_{\mathbf{up}}$
$(1,2,0,1)$	$(1-\gamma_v)^2(3-2v_0v_{01})$ $+2\gamma_v(1-\gamma_v)(3-3v_0+v_0v_{23}v_{13})$ $+\gamma_v^2(3-2v_0v_{02})$	$(1-\gamma_w)(3-2w_{23}w_2)+\gamma_w(3-2w_2)$
$(2,0,2,0)$	$3-2v_2v_{23}v_3$	$3-2w_3w_{13}w_1$
$(2,0,1,1)$	$3-2v_{23}v_2$	$(1-\gamma_w)(3-2w_{13}w_3)+\gamma_w(3-2w_3)$
$(2,0,0,2)$	$3-2v_2v_{23}v_{13}v_1$	$(1-\gamma_w)^2(3-2w_3w_0w_{13}w_{01})$ $+2\gamma_w(1-\gamma_w)(3-2w_3w_0)$ $+\gamma_w^2(3-2w_3w_0w_{23}w_{02})$
$(2,1,1,0)$	$(1-\gamma_v)(3-2v_{23}v_2)+\gamma_v(3-2v_2)$	$3-2w_3$
$(2,1,0,1)$	$(1-\gamma_v)(3-2v_{13}v_{23}v_2)+\gamma_v(3-2v_2)$	$(1-\gamma_w)(3-2w_3)+\gamma_w(3-2w_{23}w_3)$
$(2,2,0,0)$	$(1-\gamma_v)^2(3-2v_2v_0v_{01}v_{13}v_{23})$ $+2\gamma_v(1-\gamma_v)(3-2v_2v_0)$ $+\gamma_v^2(3-2v_2v_0v_{02})$	$3-2w_2w_{23}w_3$
$(1,1,1,1)$	$(1-\gamma_v)v_{23}+\gamma_v(3-2v_{02})$	$(1-\gamma_w)(3-2w_{02})+\gamma_w w_{01}$
	$(1-\gamma_v)(3-2v_{23})+\gamma_v v_{02}$	$(1-\gamma_w)w_{02}+\gamma_w(3-2w_{01})$
	$(1-\gamma_v)v_{23}+\gamma_v v_{02}$	$(1-\gamma_w)w_{02}+\gamma_w w_{01}$

Evolutionary Journal of the Linnean Society, 2025, 4, 1–29
https://doi.org/10.1093/evolinnean/kzaf019
Original Article