Online appendix for "Convergence-divergence models: Generalized phylogenetic trees modeling gene flow over time" for peer review

Jonathan D. Mitchell[1,2*] and Barbara R. Holland[1,2]

[1]School of Natural Sciences (Mathematics), University of Tasmania, Hobart, TAS, Australia.
[2]ARC Centre of Excellence for Plant Success in Nature and Agriculture, University of Tasmania, Hobart, TAS, Australia.

*Corresponding author(s). E-mail(s): jonathanmitchell88@gmail.com;

Sections, algorithms, definitions, theorems, propositions, figures and expressions introduced in the main text are labeled numerically. Any introduced in the online appendix are labeled with a number followed by "A".

1A Parameter identifiability

Recall that we decompose edges of the principal tree into "diverging sections" and "converging sections". Converging sections span only a single epoch, while diverging sections may span multiple epochs.

Recall that rates and epoch times cannot be identified individually; only their products can be identified. A parameter associated with a diverging section of an

¹⁹ edge than spans a single epoch cannot be identified. Instead, an "average" over the

²⁰ maximum number of epochs a contiguous diverging section can span can be identified.

²¹ For example, suppose rate matrix $\boldsymbol{Q}_1$ applies over epoch time t_1 to a diverging section

²² of an edge immediately before an event and rate matrix $\boldsymbol{Q}_2$ applies over epoch time

²³ t_2 to a diverging section of the edge immediately after the event. Then for the 2-state

²⁴ general Markov model, $\exp\left(\widehat{\boldsymbol{Q}}\left(t_1 + t_2\right)\right) = \exp\left(\boldsymbol{Q}_2 t_2\right)\exp\left(\boldsymbol{Q}_1 t_1\right)$, where $\widehat{\boldsymbol{Q}}$ is again

²⁵ a rate matrix from the 2-state general Markov model. Thus, we apply rate matrix $\widehat{\boldsymbol{Q}}$

²⁶ to both diverging sections of the edge.

²⁷ This lack of identifiability result follows from the 2-state general Markov model

²⁸ forming a Lie algebra, sufficient for multiplicative closure of the model class (Sumner

²⁹ et al. 2012a). Suppose α_1 and β_1 and α_2 and β_2 correspond with rate matrices $\boldsymbol{Q}_1$

³⁰ and $\boldsymbol{Q}_2$, respectively. Then by Definition 4, $\frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2}$. It is straightforward to show

³¹ that if $\widehat{\alpha}$ and $\widehat{\beta}$ are associated with $\widehat{\boldsymbol{Q}}$, then $\frac{\widehat{\alpha}}{\widehat{\beta}} = \frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2}$. Thus, the product of the

³² two transition matrices is replaced by a single "average" transition matrix.

³³ All parameters except for those corresponding to the root distribution are of the

³⁴ form $l_i = a_i + b_i = \alpha_i t_i + \beta_i t_i = a_i\left(1 + \frac{b_j}{a_j}\right)$, where i and j are arbitrary parameter

³⁵ indices, $a_i = \alpha_i t_i$ and $b_i = \beta_i t_i$. (Note that these parameters are scalars, whereas

³⁶ $\exp\left(\widehat{\boldsymbol{Q}}\left(t_1 + t_2\right)\right)$ is a matrix.) Since $\frac{\alpha_i}{\beta_i} = \frac{\alpha_j}{\beta_j}$, it follows that $\frac{a_i}{b_i} = \frac{a_j}{b_j}$.

³⁷ Contiguous diverging sections of an edge — not separated by a converging section

³⁸ — each have a single associated parameter l_i. Furthermore, each convergence group

³⁹ has an associated parameter l_j, in common for all converging sections of edges in

⁴⁰ the convergence group. In addition to parameters describing the convergence groups

⁴¹ and contiguous diverging sections, there is a parameter $\gamma = [\boldsymbol{\Pi}]_0 - [\boldsymbol{\Pi}]_1 = \frac{-a_i + b_i}{a_i + b_i}$

⁴² describing the difference in probabilities of states 0 and 1 on the root taxon.

⁴³ To form the set of parameters of a CDM, we consider a particular unique set of

⁴⁴ diverging and converging sections. Since differences in parameters between contiguous

⁴⁵ diverging sections cannot be identified, the diverging sections we consider are those

sections on the principal tree between a node or converging section and another node or converging section. Furthermore, since the exact root location on the outgroup edge is not identifiable, we consider one diverging section to be the entire outgroup edge when the principal tree of the CDM is unrooted. The converging sections correspond to individual epochs where there is a convergence group. Converging sections correspond to convergence parameters and diverging sections correspond to divergence parameters.

Note that although this is the general formulation of the parameter space, on a given CDM not all parameters are necessarily identifiable. To obtain an identifiable set of parameters some combinations of the divergence parameters may be required, which we describe in Section 3A.1. For the following sections, the parameters $x_i = \exp(-l_i) \in (0,1)$ and the variants y_i and z_i are used for establishing identifiability and distinguishability of CDMs.

2A Limiting behavior of converging taxa

2A.1 Proof of Proposition 1

Before proving the claim, we introduce some notation. Using similar notation to Sumner et al. (2012b), for some arbitrary integer $l \geq 1$ and $\boldsymbol{X} \in \{\boldsymbol{L}_\alpha, \boldsymbol{L}_\beta\}$,

$$\boldsymbol{X}^{(A)} = \prod_{i \in A} \boldsymbol{X}^{(i)},$$

where $\boldsymbol{X}^{(i)} = \boldsymbol{I} \otimes \boldsymbol{I} \otimes \ldots \otimes \boldsymbol{X} \otimes \boldsymbol{I} \otimes \ldots \otimes \boldsymbol{I}$ has $\boldsymbol{X}$ in the ith position and $\boldsymbol{I}$ in all $l-1$ other positions, $\otimes$ is the Kronecker product, $A \subseteq [l] = \{1, 2, \ldots, l\}$ and

$$\boldsymbol{L}_\alpha = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{L}_\beta = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \quad \text{and } \boldsymbol{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

3

Then we define

$$\boldsymbol{\mathcal{L}}_\alpha^{[l]} = \sum_{B \subseteq [l]} \boldsymbol{L}_\alpha^{(B)} = (\boldsymbol{L}_\alpha + \boldsymbol{I})^{\otimes l},$$

$$\boldsymbol{\mathcal{L}}_\beta^{[l]} = \sum_{B \subseteq [l]} \boldsymbol{L}_\beta^{(B)} = (\boldsymbol{L}_\beta + \boldsymbol{I})^{\otimes l},$$

where $(\boldsymbol{X} + \boldsymbol{I})^{\otimes l} = (\boldsymbol{X} + \boldsymbol{I}) \otimes (\boldsymbol{X} + \boldsymbol{I}) \otimes \ldots \otimes (\boldsymbol{X} + \boldsymbol{I})$ involves $\boldsymbol{X} + \boldsymbol{I}$ a total of l times.

Note that this definition is very similar to the rate matrix of Sumner et al. (2012b) for all l taxa present in an epoch also being in a convergence group,

$$\boldsymbol{Q}^{[l]} = \alpha \boldsymbol{\mathfrak{L}}_\alpha^{[l]} + \beta \boldsymbol{\mathfrak{L}}_\beta^{[l]},$$

where

$$\boldsymbol{\mathfrak{L}}_\alpha^{[l]} = \sum_{B \subseteq [l], B \neq \emptyset} \boldsymbol{L}_\alpha^{(B)}, \quad \boldsymbol{\mathfrak{L}}_\beta^{[l]} = \sum_{B \subseteq [l], B \neq \emptyset} \boldsymbol{L}_\beta^{(B)}.$$

Then

$$\boldsymbol{\mathcal{L}}_\alpha^{[l]} = \boldsymbol{\mathfrak{L}}_\alpha^{[l]} + \boldsymbol{I}^{\otimes l}, \quad \boldsymbol{\mathcal{L}}_\beta^{[l]} = \boldsymbol{\mathfrak{L}}_\beta^{[l]} + \boldsymbol{I}^{\otimes l},$$

where $\boldsymbol{I}^{\otimes l} = \boldsymbol{I} \otimes \boldsymbol{I} \otimes \ldots \otimes \boldsymbol{I}$ involves $\boldsymbol{I}$ a total of l times.

The proof is split into four parts. We determine the rate matrix for an arbitrary epoch in each part of the proof. 1) Instead of having N taxa, we assume that $\mathcal{N}$ has l taxa, where $l \in \{1, 2, \ldots, N\}$, all in the same convergence-divergence group in some arbitrary epoch. 2) We assume that $\mathcal{N}$ has N taxa, with the first l — according to the indices $i = i_1 i_2 \ldots i_N$ and $j = j_1 j_2 \ldots j_N$ — in the same convergence-divergence

4

group in the epoch. 3) We determine the rate matrix corresponding to an arbitrary convergence-divergence group with l taxa — still assuming $\mathcal{N}$ has N taxa — by permuting the taxon labels, which corresponds to permuting the indices i and j. 4) The rate matrix for the epoch is determined by summing the rate matrices corresponding to all convergence-divergence groups in arbitrary epoch — all N taxa are in exactly one convergence-divergence group.

Proposition 1 *Suppose a tip epoch of CDM $\mathcal{N}$ with leaf taxon set X and $|X| = N$ corresponds to a set of sets of taxa in each convergence-divergence group $\mathcal{C} = \{C_1, C_2, \ldots C_k\}$. Suppose $\boldsymbol{Q}^{[\mathcal{C}]}$ is the $2^N \times 2^N$ rate matrix representing $\mathcal{C}$. Then*

$$
\left[\boldsymbol{Q}^{[\mathcal{C}]}\right]_{ij} = \begin{cases} \alpha_r & \text{if for some } C_r \in \mathcal{C}, \quad \prod_{a \in C_r} j_a = 0, \\ & \prod_{a \in C_r} i_a = 1 \text{ and } i_a = j_a \text{ for all } a \in X \setminus C_r, \\ \beta_r & \text{if for some } C_r \in \mathcal{C}, \quad \prod_{a \in C_r} (1 - j_a) = 0, \\ & \prod_{a \in C_r} (1 - i_a) = 1 \text{ and } i_a = j_a \text{ for all } a \in X \setminus C_r, \\ 0 & \text{otherwise if } i \neq j, \\ -\sum_{s=1, s \neq j}^{2^N} \left[\boldsymbol{Q}^{[\mathcal{C}]}\right]_{sj} & \text{if } i = j, \end{cases}
$$

where $\alpha_r, \beta_r > 0$.

Proof 1) Suppose $\mathcal{N}$ has only l taxa, where $l \in \{1, 2, \ldots, N\}$, all in the same convergence-divergence group in some epoch. Then from Sumner et al. (2012b), the rate matrix for the epoch is

$$
\boldsymbol{Q}^{[l]} = \alpha \boldsymbol{\mathfrak{L}}_\alpha^{[l]} + \beta \boldsymbol{\mathfrak{L}}_\beta^{[l]},
$$

where $\alpha, \beta > 0$.

We first prove that

$$
\left[\boldsymbol{Q}^{[l]} \right]_{ij} = \begin{cases} \alpha & \text{if } \prod_{a=1}^{l} i_a = 1 \text{ and } \prod_{a=1}^{l} j_a = 0, \\ \beta & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1 \text{ and } \prod_{a=1}^{l} (1 - j_a) = 0, \\ 0 & \text{otherwise if } i \neq j. \end{cases}
$$

Note that we have not declared the diagonal elements of $\boldsymbol{Q}^{[l]}$, which are determined in
the next part of the proof.

We first define $\widetilde{\boldsymbol{Q}}^{[l]} = \alpha \boldsymbol{\mathcal{L}}_{\alpha}^{[l]} + \beta \boldsymbol{\mathcal{L}}_{\beta}^{[l]}$ and determine its elements by mathematical induction.
It is straightforward to show that

$$
\boldsymbol{L}_{\alpha} + \boldsymbol{I} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{L}_{\beta} + \boldsymbol{I} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}
$$

and in turn,

$$
\widetilde{\boldsymbol{Q}}^{[1]} = \alpha \left(\boldsymbol{L}_{\alpha} + \boldsymbol{I} \right) + \beta \left(\boldsymbol{L}_{\beta} + \boldsymbol{I} \right) = \begin{bmatrix} \beta & \beta \\ \alpha & \alpha \end{bmatrix}.
$$

We claim that for some integer $m \geq 1$,

$$
\left[\widetilde{\boldsymbol{Q}}^{[m]} \right]_{ij} = \begin{cases} \beta & \text{if } i = 1, \\ \alpha & \text{if } i = 2^m, \\ 0 & \text{otherwise.} \end{cases}
$$

We establish that if the claim is true, then

$$
\left[\widetilde{\boldsymbol{Q}}^{[m+1]} \right]_{ij} = \begin{cases} \beta & \text{if } i = 1, \\ \alpha & \text{if } i = 2^{m+1}, \\ 0 & \text{otherwise.} \end{cases}
$$

Since

$$
\begin{aligned} \widetilde{\boldsymbol{Q}}^{[m]} &= \alpha \boldsymbol{\mathcal{L}}_{\alpha}^{[m]} + \beta \boldsymbol{\mathcal{L}}_{\beta}^{[m]} \\ &= \alpha \left(\boldsymbol{L}_{\alpha} + \boldsymbol{I} \right)^{\otimes m} + \beta \left(\boldsymbol{L}_{\beta} + \boldsymbol{I} \right)^{\otimes m}, \end{aligned}
$$

6

101 it follows that

$$\widetilde{\boldsymbol{Q}}^{[m+1]} = \alpha \left(\boldsymbol{L}_\alpha + \boldsymbol{I}\right) \otimes \boldsymbol{\mathcal{L}}_\alpha^{[m]} + \beta \left(\boldsymbol{L}_\beta + \boldsymbol{I}\right) \otimes \boldsymbol{\mathcal{L}}_\beta^{[m]}$$

102 and thus the claim is established.

103   Then since

$$\begin{aligned}\boldsymbol{Q}^{[l]} =& \alpha \boldsymbol{\mathfrak{L}}_\alpha^{[l]} + \beta \boldsymbol{\mathfrak{L}}_\beta^{[l]} \\ =& \widetilde{\boldsymbol{Q}}^{[l]} - (\alpha + \beta)\, \boldsymbol{I}^{\otimes l},\end{aligned}$$

104 it follows that

$$\left[\boldsymbol{Q}^{[l]}\right]_{ij} = \begin{cases} \alpha & \text{if } \prod_{a=1}^{l} i_a = 1 \text{ and } \prod_{a=1}^{l} j_a = 0, \\ \beta & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1 \text{ and } \prod_{a=1}^{l} (1 - j_a) = 0, \\ 0 & \text{otherwise if } i \neq j. \end{cases}$$

105   Note that $\sum_{i=1}^{2^l} \left[\widetilde{\boldsymbol{Q}}^{[l]}\right]_{ij} = \alpha + \beta$. Thus, $\sum_{i=1}^{2^l} \left[\boldsymbol{Q}^{[l]}\right]_{ij} = 0$. Summarizing,

$$\left[\boldsymbol{Q}^{[l]}\right]_{ij} = \begin{cases} \alpha & \text{if } \prod_{a=1}^{l} i_a = 1 \text{ and } \prod_{a=1}^{l} j_a = 0, \\ \beta & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1 \text{ and } \prod_{a=1}^{l} (1 - j_a) = 0, \\ 0 & \text{otherwise if } i \neq j, \\ -\sum_{s=1, s\neq j}^{2^l} \left[\boldsymbol{Q}^{[l]}\right]_{sj} & \text{if } i = j. \end{cases}$$

106   From here onwards, we let $*$ represent the negative of the sum of all non-diagonal elements
107 of column j of the rate matrix.

108   2) Suppose that $\mathcal{N}$ has N taxa, with the first l — according to the indices $i = i_1 i_2 \ldots i_N$
109 and $j = j_1 j_2 \ldots j_N$ — in a convergence-divergence group. Assume that this is the only
110 convergence-divergence group in the epoch. That is, in the epoch the last $N - l$ taxa are not
111 in any convergence-divergence group; they will be assigned to convergence-divergence groups
112 at later steps in the proof. Then by Sumner et al. (2012b), the rate matrix for the last $N - l$
113 taxa is $I^{\otimes N - l}$; that is, the last $N - l$ taxa are not evolving. Note that since the first l taxa
114 are diverging independently from the last $N - l$ taxa, the rate matrix for all N taxa is the

7

Kronecker product of the rate matrix for the first l taxa and the rate matrix for the last $N - l$ taxa. Then it follows directly from 1) that the rate matrix has elements

$$
\left[\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{\otimes N-l} \right]_{ij} =
\begin{cases}
\alpha & \text{if } \prod_{a=1}^{l} i_a = 1, \ \prod_{a=1}^{l} j_a = 0 \\
& \text{and } i_a = j_a \text{ for all } a \in \{l+1, l+2, \ldots, N\}, \\
\beta & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1, \ \prod_{a=1}^{l} (1 - j_a) = 0 \\
& \text{and } i_a = j_a \text{ for all } a \in \{l+1, l+2, \ldots, N\}, \\
0 & \text{otherwise if } i \neq j, \\
* & \text{if } i = j.
\end{cases}
$$

3) Consider a permutation of leaf taxon X. This is an element of the symmetric group $\mathfrak{S}_N$ on X. An arbitrary such permutation corresponds to an arbitrary re-order of the leaf taxa. We allow the permutation to act on $V^{\otimes N}$ (the tensor product space of Sumner et al. (2012b)). For some arbitrary convergence-divergence group $C_r \in \mathcal{C}$ involving l taxa, there exists a permutation $\sigma \in \mathfrak{S}_N$ such that $\sigma \left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{\otimes N-l} \right) = \boldsymbol{Q}^{[C_r]}$, where $\boldsymbol{Q}^{[C_r]}$ is the rate matrix for the convergence-divergence group C_r and all other leaf taxa belonging to no convergence-divergence group. Then by Sumner et al. (2012b),

$$
\boldsymbol{Q}^{[C_r]} = \alpha \mathfrak{L}_\alpha^{[C_r]} + \beta \mathfrak{L}_\beta^{[C_r]},
$$

which is obtained from $\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{\otimes N-l}$ by the permutation σ on the slots of the Kronecker products of each term of $\mathfrak{L}_\alpha^{[l]} \otimes \boldsymbol{I}^{\otimes N-l}$ and $\mathfrak{L}_\beta^{[l]} \otimes \boldsymbol{I}^{\otimes N-l}$.

Then it follows directly from 2) that

$$
\left[\boldsymbol{Q}^{[C_r]} \right]_{ij} =
\begin{cases}
\alpha & \text{if } \prod_{a \in C_r} i_a = 1, \ \prod_{a \in C_r} j_a = 0 \\
& \text{and } i_a = j_a \text{ for all } a \in [N] \setminus C_r, \\
\beta & \text{if } \prod_{a \in C_r} (1 - i_a) = 1, \ \prod_{a \in C_r} (1 - j_a) = 0 \\
& \text{and } i_a = j_a \text{ for all } a \in [N] \setminus C_r, \\
0 & \text{otherwise if } i \neq j, \\
* & \text{if } i = j.
\end{cases}
$$

8

127   4) Now suppose the substitution rates for convergence-divergence group C_r are $\alpha_r, \beta_r > 0$.

128   Then since $\boldsymbol{Q}^{[\mathcal{C}]} = \sum_{C_r \in \mathcal{C}} \boldsymbol{Q}^{[C_r]}$, it follows directly from 3) that

$$
\left[\boldsymbol{Q}^{[\mathcal{C}]}\right]_{ij} = \begin{cases} \alpha_r & \text{if for some } C_r \in \mathcal{C}, \ \prod_{a \in C_r} i_a = 1, \ \prod_{a \in C_r} j_a = 0 \\ & \text{and } i_a = j_a \text{ for all } a \in [N] \setminus C_r, \\ \beta_r & \text{if for some } C_r \in \mathcal{C}, \ \prod_{a \in C_r} (1 - i_a) = 1, \ \prod_{a \in C_r} (1 - j_a) = 0 \\ & \text{and } i_a = j_a \text{ for all } a \in [N] \setminus C_r, \\ 0 & \text{otherwise if } i \neq j, \\ * & \text{if } i = j. \end{cases}
$$

129    $\square$

130   2A.2 Proof of Theorem 2

131   For the proof, we assume the phylogenetic tensor at the beginning of the tip epoch

132   takes an arbitrary form. We then consider the effect of convergence-divergence groups

133   in the tip epoch on the phylogenetic tensor. As in the proof of Proposition 2A.1,

134   the proof is split into several parts. 1) We assume $\mathcal{N}$ has N taxa, but $\mathcal{C}$ has only

135   one convergence-divergence group C_a involving the first l taxa in the tip epoch.

136   We determine the elements of the transition matrix $\boldsymbol{M}^{[C_a]} = \exp\left(\boldsymbol{Q}^{[l]} t\right) \otimes \boldsymbol{I}^{N-l}$.

137   2) We determine the phylogenetic tensor in the limit as the epoch time of the

138   tip epoch diverges. 3) We assume $\mathcal{C}$ has $k \geq 1$ convergence-divergence groups,

139    $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$, and determine the phylogenetic tensor in the limit as the epoch

140   time of the tip epoch diverges.

141   **Theorem 2** *Suppose an arbitrary epoch of CDM $\mathcal{N}$ corresponds to set of sets of taxa in*

142   *each convergence-divergence group $\mathcal{C} = \{C_1, C_2, \ldots C_k\}$. Then if $a, b \in C_i$, as tip epoch length*

143   *$t \to \infty$, a and b become identical.*

9

¹⁴⁴ *Proof* 1) As in the proof of Proposition 2A.1, assume that the convergence-divergence
¹⁴⁵ group C_a involves the first l taxa. We assume $\mathcal{N}$ has N taxa, unlike in 1) of the proof of
¹⁴⁶ Proposition 2A.1. Suppose $\boldsymbol{P}'$ is the phylogenetic tensor representing the probabilities of
¹⁴⁷ combinations of states immediately before the tip epoch. Then let

$$\widetilde{\boldsymbol{P}} = \exp\left(\boldsymbol{Q}^{[C_a]}t\right) \cdot \boldsymbol{P}',$$

¹⁴⁸ where

$$\boldsymbol{Q}^{[C_a]} = \boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}.$$

¹⁴⁹ To find an expression for $\exp\left(\boldsymbol{Q}^{[C_a]}t\right)$, we use the Taylor series,

$$
\begin{aligned}
\exp\left(\boldsymbol{Q}^{[C_a]}t\right) &= \exp\left(\left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right)t\right) \\
&= \boldsymbol{I}^{\otimes N} + \left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right)t + \frac{1}{2!}\left(\left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right)t\right)^2 + \dots \\
&= \boldsymbol{I}^{\otimes N} + \left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right)t + \frac{t^2}{2!}\left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right) \cdot \left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right) + \dots \\
&= \boldsymbol{I}^{\otimes N} + \left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right)t + \frac{t^2}{2!}\left(\boldsymbol{Q}^{[l]} \cdot \boldsymbol{Q}^{[l]}\right) \otimes \left(\boldsymbol{I}^{N-l} \cdot \boldsymbol{I}^{N-l}\right) + \dots \\
&= \boldsymbol{I}^{\otimes N} + \left(\boldsymbol{Q}^{[l]} \otimes \boldsymbol{I}^{N-l}\right)t + \frac{t^2}{2!}\left(\boldsymbol{Q}^{[l]}\right)^2 \otimes \boldsymbol{I}^{N-l} + \dots \\
&= \left(\boldsymbol{I}^{\otimes l} + \boldsymbol{Q}^{[l]}t + \frac{t^2}{2!}\left(\boldsymbol{Q}^{[l]}\right)^2 + \dots\right) \otimes \boldsymbol{I}^{N-l} \\
&= \exp\left(\boldsymbol{Q}^{[l]}t\right) \otimes \boldsymbol{I}^{N-l}.
\end{aligned}
$$

¹⁵⁰ Then

$$\widetilde{\boldsymbol{P}} = \left(\exp\left(\boldsymbol{Q}^{[l]}t\right) \otimes \boldsymbol{I}^{N-l}\right) \cdot \boldsymbol{P}'.$$

¹⁵¹ Now focusing on $\exp\left(\boldsymbol{Q}^{[l]}t\right)$, again using a Taylor series,

$$\exp\left(\boldsymbol{Q}^{[l]}t\right) = \boldsymbol{I}^{\otimes l} + \boldsymbol{Q}^{[l]}t + \frac{1}{2!}\left(\boldsymbol{Q}^{[l]}\right)^2 t^2 + \dots.$$

¹⁵² Focusing on $\left(\boldsymbol{Q}^{[l]}\right)^2$,

$$
\begin{aligned}
\left(\boldsymbol{Q}^{[l]}\right)^2 &= \left(\alpha\mathfrak{L}_\alpha^{[l]} + \beta\mathfrak{L}_\beta^{[l]}\right) \cdot \left(\alpha\mathfrak{L}_\alpha^{[l]} + \beta\mathfrak{L}_\beta^{[l]}\right) \\
&= \alpha^2\left(\mathfrak{L}_\alpha^{[l]}\right)^2 + \alpha\beta\left(\mathfrak{L}_\alpha^{[l]}\mathfrak{L}_\beta^{[l]} + \mathfrak{L}_\beta^{[l]}\mathfrak{L}_\alpha^{[l]}\right) + \beta^2\left(\mathfrak{L}_\beta^{[l]}\right)^2,
\end{aligned}
$$

10

153   where $\alpha, \beta > 0$.

154   From Sumner et al. (2012b),

$$\left(\boldsymbol{\mathfrak{L}}_\alpha^{[l]}\right)^2 = -\boldsymbol{\mathfrak{L}}_\alpha^{[l]}, \quad \boldsymbol{\mathfrak{L}}_\alpha^{[l]}\boldsymbol{\mathfrak{L}}_\beta^{[l]} = -\boldsymbol{\mathfrak{L}}_\beta^{[l]}, \quad \boldsymbol{\mathfrak{L}}_\beta^{[l]}\boldsymbol{\mathfrak{L}}_\alpha^{[l]} = -\boldsymbol{\mathfrak{L}}_\alpha^{[l]}, \quad \left(\boldsymbol{\mathfrak{L}}_\beta^{[l]}\right)^2 = -\boldsymbol{\mathfrak{L}}_\beta^{[l]}.$$

155   Then

$$
\begin{aligned}
\left(\boldsymbol{Q}^{[l]}\right)^2 &= -\alpha^2 \boldsymbol{\mathfrak{L}}_\alpha^{[l]} - \alpha\beta\left(\boldsymbol{\mathfrak{L}}_\alpha^{[l]} + \boldsymbol{\mathfrak{L}}_\beta^{[l]}\right) - \beta^2 \boldsymbol{\mathfrak{L}}_\beta^{[l]} \\
&= -(\alpha + \beta)\left(\alpha\boldsymbol{\mathfrak{L}}_\alpha^{[l]} + \beta\boldsymbol{\mathfrak{L}}_\beta^{[l]}\right) \\
&= -(\alpha + \beta)\,\boldsymbol{Q}^{[l]}.
\end{aligned}
$$

156   It follows that

$$\left(\boldsymbol{Q}^{[l]}\right)^u = (-1)^{u-1}(\alpha + \beta)\,\boldsymbol{Q}^{[l]}$$

157   for any positive integer $u \geq 2$.

158   Returning to the Taylor series,

$$
\begin{aligned}
\exp\left(\boldsymbol{Q}^{[l]}t\right) &= \boldsymbol{I}^{\otimes l} + \boldsymbol{Q}^{[l]}t - \frac{1}{2!}(\alpha + \beta)\,\boldsymbol{Q}^{[l]}t^2 + \dots \\
&= \boldsymbol{I}^{\otimes l} + \left(t - \frac{(\alpha + \beta)\,t^2}{2} + \dots\right)\boldsymbol{Q}^{[l]} \\
&= \boldsymbol{I}^{\otimes l} + \frac{1}{\alpha + \beta}\left((\alpha + \beta)\,t - \frac{(\alpha + \beta)^2\,t^2}{2} + \dots\right)\boldsymbol{Q}^{[l]} \\
&= \boldsymbol{I}^{\otimes l} - \frac{1}{\alpha + \beta}\left(-(\alpha + \beta)\,t + \frac{(\alpha + \beta)^2\,t^2}{2} - \dots\right)\boldsymbol{Q}^{[l]} \\
&= \boldsymbol{I}^{\otimes l} - \frac{1}{\alpha + \beta}\left(1 - (\alpha + \beta)\,t + \frac{(\alpha + \beta)^2\,t^2}{2} - \dots\right)\boldsymbol{Q}^{[l]} + \frac{1}{\alpha + \beta}\boldsymbol{Q}^{[l]} \\
&= \boldsymbol{I}^{\otimes l} + \frac{1}{\alpha + \beta}\left(1 - \exp\left(-(\alpha + \beta)\,t\right)\right)\boldsymbol{Q}^{[l]}.
\end{aligned}
$$

159   Now recall from the proof of Proposition 2A.1 that

$$
\left[\boldsymbol{Q}^{[l]}\right]_{ij} = \begin{cases}
\alpha & \text{if } \prod_{a=1}^l i_a = 1 \text{ and } \prod_{a=1}^l j_a = 0, \\[2mm]
\beta & \text{if } \prod_{a=1}^l (1 - i_a) = 1 \text{ and } \prod_{a=1}^l (1 - j_a) = 0, \\[2mm]
0 & \text{otherwise if } i \neq j, \\[2mm]
* & \text{otherwise if } i = j.
\end{cases}
$$

11

160   Thus, for columns to sum to zero,

$$
\left[\boldsymbol{Q}^{[l]}\right]_{ij} =
\begin{cases}
\alpha & \text{if } \prod_{a=1}^{l} i_a = 1 \text{ and } \prod_{a=1}^{l} j_a = 0, \\[2mm]
\beta & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1 \text{ and } \prod_{a=1}^{l} (1 - j_a) = 0, \\[2mm]
0 & \text{otherwise if } i \neq j, \\[2mm]
-\alpha & \text{if } \prod_{a=1}^{l} (1 - i_a) = \prod_{a=1}^{l} (1 - j_a) = 1, \\[2mm]
-\beta & \text{if } \prod_{a=1}^{l} i_a = \prod_{a=1}^{l} j_a = 1, \\[2mm]
-(\alpha + \beta) & \text{otherwise.}
\end{cases}
$$

161   Letting $\boldsymbol{M}^{[l]} = \exp\left(\boldsymbol{Q}^{[l]} t\right)$,

$$
\left[\boldsymbol{M}^{[l]}\right]_{ij} =
\begin{cases}
\frac{\alpha}{\alpha+\beta} \left(1 - \exp\left(-(\alpha+\beta)\,t\right)\right) & \text{if } \prod_{a=1}^{l} i_a = 1 \text{ and } \prod_{a=1}^{l} j_a = 0, \\[2mm]
\frac{\beta}{\alpha+\beta} \left(1 - \exp\left(-(\alpha+\beta)\,t\right)\right) & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1 \text{ and } \prod_{a=1}^{l} (1 - j_a) = 0, \\[2mm]
0 & \text{otherwise if } i \neq j, \\[2mm]
1 - \frac{\alpha}{\alpha+\beta} \left(1 - \exp\left(-(\alpha+\beta)\,t\right)\right) & \text{if } \prod_{a=1}^{l} (1 - i_a) = \prod_{a=1}^{l} (1 - j_a) = 1, \\[2mm]
1 - \frac{\beta}{\alpha+\beta} \left(1 - \exp\left(-(\alpha+\beta)\,t\right)\right) & \text{if } \prod_{a=1}^{l} i_a = \prod_{a=1}^{l} j_a = 1, \\[2mm]
\exp\left(-(\alpha+\beta)\,t\right) & \text{otherwise.}
\end{cases}
$$

162   Next, take the limit as $t \to \infty$. Then

$$
\lim_{t \to \infty} \left[\boldsymbol{M}^{[l]}\right]_{ij} =
\begin{cases}
\frac{\alpha}{\alpha+\beta} & \text{if } \prod_{a=1}^{l} i_a = 1, \\[2mm]
\frac{\beta}{\alpha+\beta} & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1, \\[2mm]
0 & \text{otherwise.}
\end{cases}
$$

12

Now let $\boldsymbol{M}^{[C_a]} = \exp\left(\boldsymbol{Q}^{[l]}t\right) \otimes \boldsymbol{I}^{N-l}$. Then

$$
\lim_{t\to\infty} \left[\boldsymbol{M}^{[C_a]}\right]_{ij} = \begin{cases} \frac{\alpha}{\alpha+\beta} & \text{if } \prod_{a=1}^{l} i_a = 1 \\ & \text{and } i_a = j_a \text{ for all } a \in \{l+1, l+2, \ldots, N\}, \\ \frac{\beta}{\alpha+\beta} & \text{if } \prod_{a=1}^{l} (1 - i_a) = 1 \\ & \text{and } i_a = j_a \text{ for all } a \in \{l+1, l+2, \ldots, N\}, \\ 0 & \text{otherwise.} \end{cases}
$$

In summary, in the limit, the only rows of $\boldsymbol{M}^{[C_a]}$ with non-zero elements have the first l indices being either all 0 or all 1. Then in the limit, the only non-zero elements of $\widetilde{\boldsymbol{P}}$ also have the first l indices being either all 0 or all 1.

2) We let the substitution rates for convergence group C_k be $\alpha_k, \beta_k > 0$ and recognize that since $\frac{\alpha_k}{\beta_k} = \frac{\alpha}{\beta}$, then $\frac{\alpha_k}{\alpha_k+\beta_k} = \frac{\alpha}{\alpha+\beta}$ and $\frac{\beta_k}{\alpha_k+\beta_k} = \frac{\beta}{\alpha+\beta}$. Then using the same arguments as in Proposition 2A.1, $\sigma\left(\boldsymbol{M}^{[C_a]}\right) = \boldsymbol{M}^{[C_k]}$ and

$$
\lim_{t\to\infty} \left[\boldsymbol{M}^{[C_k]}\right]_{ij} = \begin{cases} \frac{\alpha}{\alpha+\beta} & \text{if } \prod_{a \in C_k} i_a = 1 \\ & \text{and } i_a = j_a \text{ for all } a \in [N] \setminus C_k, \\ \frac{\beta}{\alpha+\beta} & \text{if } \prod_{a \in C_k} (1 - i_a) = 1 \\ & \text{and } i_a = j_a \text{ for all } a \in [N] \setminus C_k, \\ 0 & \text{otherwise} \end{cases}
$$

and

$$
\lim_{t\to\infty} \sum_{j=1}^{2^N} \left[\boldsymbol{M}^{[C_k]}\right]_{ij} \left[\boldsymbol{P}'\right]_j = \begin{cases} c_i^{[C_k]} > 0 & \text{if } \prod_{a \in C_k} i_a = 1 \text{ or } \prod_{a \in C_k} (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}
$$

Note that $c_i^{[C_k]} > 0$ being strictly positive follows from Assumption 11 of Section 3.2.

13

172   3) $\mathcal{N}$ has a tip epoch with epoch time t and set of convergence groups $\mathcal{C} =$

173    $\{C_1, C_2, \ldots, C_k\}$. Then the phylogenetic tensor $\boldsymbol{P}$ representing the probabilities of combina-

174   tions of states at the leaves of the principal tree can be expressed as

$$\boldsymbol{P} = \exp\left(\boldsymbol{Q}^{[\mathcal{C}]} t\right) \cdot \boldsymbol{P}'$$
$$= \prod_{r=1}^{k} \boldsymbol{M}^{[C_r]} \cdot \boldsymbol{P}'.$$

175   We prove that all elements of $\boldsymbol{P}$ converge to 0 except those where, for each $C_r \in \mathcal{C}$, all

176   taxa in C_r are in the same state.

177   In the limit as the epoch time of the tip epoch diverges, the phylogenetic tensor is

$$\lim_{t \to \infty} \boldsymbol{P} = \lim_{t \to \infty} \prod_{r=1}^{k} \boldsymbol{M}^{[C_r]} \cdot \boldsymbol{P}' = \prod_{r=1}^{k} \lim_{t \to \infty} \boldsymbol{M}^{[C_r]} \cdot \boldsymbol{P}'.$$

178   We prove that

$$\lim_{t \to \infty} [\boldsymbol{P}]_i = \begin{cases} c_i^{[\mathcal{C}]} > 0 & \text{if for all } C_r \in \mathcal{C}, \ \prod_{a \in C_r} i_a = 1 \text{ or } \prod_{a \in C_r} (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

179   We prove this claim by induction on the k convergence groups. Note that in 2)

180   we have already proven the claim for the first convergence-divergence group C_k applied

181   to $\boldsymbol{P}'$. Thus, all that remains it to prove that given the claim is true for $\boldsymbol{P}^{k-v+2} =$

182    $\boldsymbol{M}^{[C_v]} \cdots \boldsymbol{M}^{[C_{v+1}]} \cdots \boldsymbol{M}^{[C_k]} \cdot \boldsymbol{P}'$ for some $v \in \{2, \ldots, k\}$, it must be true for $\boldsymbol{P}^{k-v+3} =$

183    $\boldsymbol{M}^{[C_{v-1}]} \cdots \boldsymbol{M}^{[C_v]} \cdots \boldsymbol{M}^{[C_k]} \cdot \boldsymbol{P}'$.

184   We assume that

$$\lim_{t \to \infty} \left[\boldsymbol{P}^{k-v+2}\right]_i = \begin{cases} c_i^{\left[\cup_{r=v}^{k} C_r\right]} > 0 & \text{if for all } r \in \{v, v+1, \ldots, k\}, \ \prod_{a \in C_r} i_a = 1 \\ & \text{or } \prod_{a \in C_r} (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

185   Then in the limit as $t \to \infty$, all elements of $\boldsymbol{P}^{k-v+2}$ are 0 except those where for all

186    $C_r \in \mathcal{C}$, $r \in \{v, v+1, \ldots, k\}$, all taxa in C_r are in the same state.

187   For $\left[\boldsymbol{P}^{k-v+3}\right]_i$ to be non-zero, there must exist some index s, such that

$$\left[\boldsymbol{M}^{[C_{v-1}]}\right]_{is} > 0 \tag{1A}$$

14

and

$$\left[\boldsymbol{P}^{k-v+2}\right]_s > 0. \tag{2A}$$

For Equation (1A) to be true,

$$\begin{cases} \prod_{a\in C_{v-1}} i_a = 1 \text{ or } \prod_{a\in C_{v-1}} (1 - i_a) = 1, \\ \\ i_a = s_a \text{ for all } a \in [N] \setminus C_{v-1}. \end{cases} \tag{3A}$$

For Equation (2A) to be true, by assumption,

$$\text{for all } r \in \{v, v+1, \ldots, k\}, \ \prod_{a\in C_r} s_a = 1 \text{ or } \prod_{a\in C_r} (1 - s_a) = 1. \tag{4A}$$

Combining Constraints (3A) and Constraints (4A),

$$\begin{cases} \prod_{a\in C_{v-1}} i_a = 1 \text{ or } \prod_{a\in C_{v-1}} (1 - i_a) = 1, \\ \\ \text{for all } r \in \{v, v+1, \ldots, k\}, \ \prod_{a\in C_r} i_a = 1 \text{ or } \prod_{a\in C_r} (1 - i_a) = 1. \end{cases} \tag{5A}$$

Constraint (5A) can be simplified to

$$\text{for all } r \in \{v-1, v, \ldots, k\}, \ \prod_{a\in C_r} i_a = 1 \text{ or } \prod_{a\in C_r} (1 - i_a) = 1.$$

In summary,

$$\lim_{t\to\infty} [\boldsymbol{P}]_i = \begin{cases} c_i^{\left[\cup_{r=1}^k C_r\right]} > 0 & \text{if for all } r \in \{1, 2, \ldots, k\}, \\ & \qquad \prod_{a\in C_r} i_a = 1 \text{ or } \prod_{a\in C_r} (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

□

3A Identifiability of 4-taxon CDMs

Sumner et al. (2012b) formally describe phylogenetic epoch models in their Definition 6.1 and introduce notation to compute the phylogenetic tensors. We use the same notation for our CDMs.

15

For each 4-taxon CDM, the phylogenetic tensor $\boldsymbol{P}$ is transformed into the Hadamard basis $\widehat{\boldsymbol{P}}$ by multiplying by $\boldsymbol{H}_{16} = \boldsymbol{H}_2^{\otimes 4}$, where

$$\boldsymbol{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

In the Hadamard basis, the phylogenetic tensor for CDM 5 is

$$
\widehat{\boldsymbol{P}} = \boldsymbol{H}_{16} \cdot \boldsymbol{P} =
\begin{bmatrix}
q_{0000} \\
q_{0001} \\
q_{0010} \\
q_{0011} \\
q_{0100} \\
q_{0101} \\
q_{0110} \\
q_{0111} \\
q_{1000} \\
q_{1001} \\
q_{1010} \\
q_{1011} \\
q_{1100} \\
q_{1101} \\
q_{1110} \\
q_{1111}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
1 \\
\gamma \\
\gamma \\
\gamma^2 + \left(1 - \gamma^2\right) r_{0011} \\
\gamma \\
\gamma^2 + \left(1 - \gamma^2\right) r_{0101} \\
\gamma^2 + \left(1 - \gamma^2\right) r_{0110} \\
\gamma \left(\gamma^2 + \left(1 - \gamma^2\right) \left(r_{0011} + r_{0101} + r_{0110} - 2r_{0111}\right)\right) \\
\gamma \\
\gamma^2 + \left(1 - \gamma^2\right) r_{1001} \\
\gamma^2 + \left(1 - \gamma^2\right) r_{1010} \\
\gamma \left(\gamma^2 + \left(1 - \gamma^2\right) \left(r_{0011} + r_{1001} + r_{1010} - 2r_{1011}\right)\right) \\
\gamma^2 + \left(1 - \gamma^2\right) r_{1100} \\
\gamma \left(\gamma^2 + \left(1 - \gamma^2\right) \left(r_{0101} + r_{1001} + r_{1100} - 2r_{1101}\right)\right) \\
\gamma \left(\gamma^2 + \left(1 - \gamma^2\right) \left(r_{0110} + r_{1010} + r_{1100} - 2r_{1110}\right)\right) \\
\gamma^2 \left(\gamma^2 + \left(1 - \gamma^2\right) \left(r_{0011} + r_{0101} + r_{0110} + r_{1001} + r_{1010} + r_{1100} \right. \right. \\
\left. \left. -2\left(r_{0111} + r_{1011} + r_{1101} + r_{1110} - 2\delta\right)\right)\right) + \left(1 - \gamma^2\right)^2 r_{1111}
\end{bmatrix}.
$$

(6A)

See Mathematica file S2.nb (text version S3.txt) on https://github.com/
jonathanmitchell88/CDMsSI for a derivation of Equation (6A) and equations for r_{0011},
r_{0101}, ..., r_{1111} and δ in terms of x_i and y_i for CDM 5. CDMs $1-4$ are all nested in
CDM 5. Thus, their phylogenetic tensors are also in the form of Equation (6A). How-
ever, the equations for r_{0011}, r_{0101}, ..., r_{1111} and δ involve different expressions of x_i
and y_i.

For the proof that follows, the order of parameters is as in Figure 3, with $x_i = \exp\left(-\left(a_i + b_i\right)\right) \in (0, 1)$ for all $i \in \{1, 2, \ldots, 11\}$. Note again that the exact location of
the root on the outgroup edge is not identifiable; t_1 corresponds to the sum of epoch
times of epochs from the root to the outgroup added to the first epoch time below the
root.

To establish whether a CDM is identifiable or not, we must first determine a
maximal set of independent elements of the transformed phylogenetic tensor. That
is, a set with maximum cardinality such that there are no algebraic equations —
equalities — involving multiple elements of the set. If the cardinality of the set equals
the number of parameters, then the CDM is identifiable. For example, we can see that
invariants $q_{0001} = q_{0010} = q_{0100} = q_{1000} = \gamma$ are equalities on all CDMs. Thus, we can
only include one of q_{0001}, q_{0010}, q_{0100} and q_{1000} in the set.

To determine all such equalities, for a given CDM with $l + 1$ parameters
$x_1, x_2, \ldots, x_l, \gamma$, we construct the ideal,

$$
\begin{aligned}
I = &\langle r_{0011} - f_{0011}\left(x_1, x_2, \ldots, x_l\right), r_{0101} - f_{0101}\left(x_1, x_2, \ldots, x_l\right), \ldots, \\
& r_{1111} - f_{1111}\left(x_1, x_2, \ldots, x_l\right), \delta - f_\delta\left(x_1, x_2, \ldots, x_l\right)\rangle \\
& \subseteq \mathbb{Q}\left[x_1, x_2, \ldots, x_l, r_{0011}, r_{0101}, \ldots, r_{1111}, \delta\right],
\end{aligned}
$$

where each $r_{ijkl} - f_{ijkl}\left(x_1, x_2, \ldots, x_l\right)$ and $\delta - f_\delta\left(x_1, x_2, \ldots, x_l\right)$ is identically zero.
(We can ignore γ since $q_{ijkl} = \gamma^2 + \left(1 - \gamma^2\right) r_{ijkl}$ and including any of these invariants

²²⁴ does not help us to determine any invariants involving multiple variables r_{0011}, r_{0101},
²²⁵ ..., r_{1111}, δ.)

²²⁶ In the Macaulay2 file S4.m2 (output file S5.txt) on https://github.com/
²²⁷ jonathanmitchell88/CDMsSI we derive the (reduced) Gröbner basis for this ideal for
²²⁸ a particular monomial order for CDM 5. Below we outline how this Gröbner basis is
²²⁹ computed.

²³⁰ In the Mathematica file S2.nb (text version S3.txt) we derive the following
²³¹ equations to input into the generators of the ideal:

$$
\begin{cases}
f_{0011} = x_4 x_5 x_6 x_7 x_9 x_{10}, \\[4pt]
f_{0101} = x_{10} x_{11} \left(1 - x_9 \left(1 - x_2 x_3 x_4 x_6 x_8\right)\right), \\[4pt]
f_{0110} = x_7 x_8 x_9 x_{11} \left(1 - x_6 \left(1 - x_2 x_3 x_5\right)\right), \\[4pt]
f_{0111} = x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11}, \\[4pt]
f_{1001} = x_1 x_2 x_4 x_9 x_{10}, \\[4pt]
f_{1010} = x_1 x_2 x_5 x_6 x_7, \\[4pt]
f_{1011} = x_1 x_2 x_4 x_5 x_6 x_7 x_9 x_{10}, \\[4pt]
f_{1100} = x_1 x_3 x_6 x_8 x_9 x_{11}, \\[4pt]
f_{1101} = x_1 x_2 x_3 x_4 x_6 x_8 x_9 x_{10} x_{11}, \\[4pt]
f_{1110} = x_1 x_2 x_3 x_5 x_6 x_7 x_8 x_9 x_{11}, \\[4pt]
f_{1111} = x_1 x_7 x_{10} x_{11} \left(x_4 x_8 x_9 \left(x_2 \left(1 - x_6\right) + x_3 x_5 x_6\right) + x_2 x_5 x_6 \left(1 - x_9\right)\right), \\[4pt]
\quad f_{\delta} = x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11}.
\end{cases}
\tag{7A}
$$

²³² The functions $f_{0011} = f_{0011}(x_1, x_2, \ldots, x_l)$, $f_{0101} = f_{0101}(x_1, x_2, \ldots, x_l)$, ...,
²³³ $f_{1111} = f_{1111}(x_1, x_2, \ldots, x_l)$ and $f_{\delta} = f_{\delta}(x_1, x_2, \ldots, x_l)$ depend on the CDM in
²³⁴ question, for example, CDM 5.

19

The monomial order is the elimination order, eliminating the block $x_1, x_2, \ldots, x_l$, with graded reverse lexicographic order on each block, $x_1 > x_2 > \ldots > x_l$ and $r_{0011} > r_{0101} > \ldots > r_{1111} > \delta$.

Next, we compute the (reduced) Gröbner basis I_G of I. Then $I_{G,q} = I_G \cap \mathbb{R}[r_{0011}, r_{0101}, \ldots, r_{1111}, \delta]$ is a Gröbner basis for the elimination ideal involving only $r_{0011}, r_{0101}, \ldots, r_{1111}, \delta$.

Note that q_{1111} is a function of both r_{1111} and δ, the only element of $\widehat{\boldsymbol{P}}$ that is a function of either. Thus, the maximum cardinality set can include at most one of r_{1111} and δ. In S4.m2 we find that when eliminating r_{1111} there are no generators that involve δ. Thus, r_{1111} is eliminated and δ is another independent variable of the system when r_{1111} is eliminated.

Note that there are still some algebraic equations — equalities — involving some elements of $\{r_{0011}, r_{0101}, \ldots, r_{1110}, \delta\}$. In S4.m2 (output file S5.txt) we find the largest cardinality subset with no algebraic equations involving multiple elements. This cardinality, plus one for γ, is the degrees of freedom of the phylogenetic tensor. Given a set of parameters of the CDM, if this degrees of freedom is less than the number of parameters, then the system of polynomial equations is underdetermined and that set of parameters is not identifiable. (Note that some individual parameters may still be indentifiable.) Otherwise, the set of parameters is identifiable. If that set of parameters is not identifiable, it may be possible to combine the parameters in a such a way that the new set of parameters is identifiable.

3A.1 Proof of Proposition 3

See S4.m2 (output file S5.txt) and S6.m2 (output file S7.txt) on https://github.com/jonathanmitchell88/CDMsSI for the computations of the (reduced) Gröbner bases of the ideals in this proof.

Proposition 3 *The parameter set for CDM 5 is identifiable.*

20

₂₆₁ *Proof* In S4.m2 (output file S5.txt), we see that there are 9 elements of

₂₆₂ $\{r_{0011}, r_{0101}, \dots, r_{1110}, \delta\}$ that are free to vary. However, CDM 5 has 11 parameters exclud-

₂₆₃ ing γ. Thus, this set of parameters is not identifiable. However, recall in Section 1A that

₂₆₄ taking some products of x_i parameters may be required to obtain a set of identifiable

₂₆₅ parameters. Since there are 9 elements of $\{r_{0011}, r_{0101}, \dots, r_{1110}, \delta\}$ that are free to vary, we

₂₆₆ desire a set of 9 parameters.

₂₆₇ In S2.nb (text version S3.txt), we express $f_{0011}, f_{0101}, \dots, f_{1111}, \delta$ in terms of the set of

₂₆₈ parameters $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9\}$. Precisely,

$$\begin{cases} y_1 = x_1, \\ y_2 = x_2, \\ y_3 = x_3 x_8 x_{11}, \\ y_4 = x_4 x_{10}, \\ y_5 = x_5 x_7, \\ y_6 = x_6, \\ y_7 = x_7 x_8 x_{11}, \\ y_8 = x_9, \\ y_9 = x_{10} x_{11}. \end{cases}$$

₂₆₉ In S6.m2 (output file S7.txt), we see that this set of parameters is identifiable. We note

₂₇₀ that $x_i \in (0, 1)$ for all $i \in \{1, 2, \dots, 11\}$. It follows that $r_{0011}, r_{0101}, \dots, r_{1111}, \delta \in (0, 1)$ and

₂₇₁ $y_i \in (0, 1)$ for all $i \in \{1, 2, \dots, 9\}$. In S2.nb (text version S3.txt), we see that the solutions to

21

the system are

$$\begin{cases} y_1 = \frac{\delta}{r_{0111}}, \\[6pt] y_2 = \frac{r_{0111}\sqrt{r_{1001}r_{1010}}}{\delta\sqrt{r_{0011}}}, \\[6pt] y_3 = \frac{\delta}{\sqrt{r_{0011}r_{1001}r_{1010}}}, \\[6pt] y_4 = \frac{r_{1101}\delta\sqrt{r_{0011}}}{r_{0111}r_{1100}\sqrt{r_{1001}r_{1010}}}, \\[6pt] y_5 = \frac{\delta}{r_{1101}}, \\[6pt] y_6 = \frac{r_{1101}\sqrt{r_{0011}r_{1010}}}{\delta\sqrt{r_{1001}}}, \\[6pt] y_7 = \frac{\delta\left(r_{0110}r_{1101}\delta\sqrt{r_{0011}}-r_{0111}^2 r_{1100}\sqrt{r_{1001}r_{1010}}\right)}{r_{0111}r_{1100}\sqrt{r_{0011}r_{1001}}(\delta\sqrt{r_{1001}}-r_{1101}\sqrt{r_{0011}r_{1010}})}, \\[6pt] y_8 = \frac{r_{0111}r_{1001}r_{1100}}{r_{1101}\delta}, \\[6pt] y_9 = \frac{r_{1101}(r_{0101}\delta-r_{0111}r_{1101})}{r_{1101}\delta-r_{0111}r_{1001}r_{1100}}. \end{cases} \tag{8A}$$

Thus, the parameter set $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, \gamma\}$ on CDM 5 is identifiable.

$\square$

Since CDMs $1-4$ are all nested in CDM 5, the transformed phylogenetic tensors of CDMs $1-4$ can be determined directly from that of CDM 5 by setting some parameters x_i to 1. Similarly, it is straightforward to prove that the equivalent sets of y_i parameters are identifiable for each of CDMs $1-4$. The numbers of degrees of freedom for the phylogenetic tensors of CDMs $1-5$ are 6, 7, 8, 9 and 10, respectively.

4A Proof of Theorem 5

For a robust proof, we could consider the (reduced) Gröbner bases of the ideals representing the parameter spaces of the CDMs and show that each CDM has a unique Gröbner basis. The Gröbner basis for CDM 5 has already been computed in Section 3A. However, computation of the Gröbner bases is slow and some bases contain many generators. Instead, it is sufficient to consider only a few constraints for each parameter space that exist for some CDMs and not others, greatly simplifying the proof.

22

287 **Theorem 5** *All pairs of 4-taxon leaf-labeled CDMs of Section 5.2 are distinguishable.*

288 *Proof* By Proposition 4, if two CDMs have parameter spaces with different dimensions, then
289 they are distinguishable from each other. CDMs 1, 2, 3, 4 and 5 have parameter space
290 dimensions 6, 7, 8, 9 and 10 respectively, corresponding to the numbers of free parameters.
291   All that is left to prove is that any two CDMs that differ only in their leaf labelings are
292 distinguishable. The notation that follows is consistent with that of Section 3A. Recall that
293 $y_i \in (0, 1)$ for all $i \in \{1, 2, \ldots, 9\}$.

CDM 5

295 See S8.nb (text version S9.txt) on https://github.com/jonathanmitchell88/CDMsSI for
296 proofs of the following claims.
297   For leaf labelings $(o, (a, (b, c)))$ and $(o, (a, (c, b)))$,

$$r_{0011}r_{1001}r_{1010} - r_{1011}^2 = 0,$$

298 while for the other leaf labelings

$$r_{0011}r_{1001}r_{1010} - r_{1011}^2 > 0.$$

299   Thus, we need only show that CDMs with leaf labelings $(o, (a, (b, c)))$ and $(o, (a, (c, b)))$
300 are distinguishable. To do this, we show that the intersection of the parameter spaces of
301 the two CDMs is the empty set. Letting y_i be the parameters corresponding to leaf labeling
302 $(o, (a, (b, c)))$ and z_i corresponding to $(o, (a, (c, b)))$, we equate the equations for each element

23

303 of the two phylogenetic tensors and solve for the z_i parameters,

$$
\begin{cases}
y_4 y_5 y_6 y_8 = & z_4 z_5 z_6 z_8, \\[4pt]
y_9 \left(1 - y_8\right) + y_2 y_3 y_4 y_6 y_8 = & z_8 \left(z_7 \left(1 - z_6\right) + z_2 z_3 z_5 z_6\right), \\[4pt]
y_8 \left(y_7 \left(1 - y_6\right) + y_2 y_3 y_5 y_6\right) = & z_9 \left(1 - z_8\right) + z_2 z_3 z_4 z_6 z_8, \\[4pt]
y_2 y_3 y_4 y_5 y_6 y_8 = & z_2 z_3 z_4 z_5 z_6 z_8, \\[4pt]
y_1 y_2 y_4 y_8 = & z_1 z_2 z_5 z_6, \\[4pt]
y_1 y_2 y_5 y_6 = & z_1 z_2 z_4 z_8, \\[4pt]
y_1 y_2 y_4 y_5 y_6 y_8 = & z_1 z_2 z_4 z_5 z_6 z_8, \\[4pt]
y_1 y_3 y_6 y_8 = & z_1 z_3 z_6 z_8, \\[4pt]
y_1 y_2 y_3 y_4 y_6 y_8 = & z_1 z_2 z_3 z_5 z_6 z_8, \\[4pt]
y_1 y_2 y_3 y_5 y_6 y_8 = & z_1 z_2 z_3 z_4 z_6 z_8, \\[4pt]
y_1 \left(y_4 y_8 \left(y_2 y_7 \left(1 - y_6\right) + y_3 y_5 y_6\right)\right. & z_1 \left(z_4 z_8 \left(z_2 z_7 \left(1 - z_6\right) + z_3 z_5 z_6\right)\right. \\
\left.+ y_2 y_5 y_6 y_9 \left(1 - y_8\right)\right) = & \left.+ z_2 z_5 z_6 z_9 \left(1 - z_8\right)\right), \\[4pt]
y_1 y_2 y_3 y_4 y_5 y_6 y_8 = & z_1 z_2 z_3 z_4 z_5 z_6 z_8.
\end{cases}
$$

304 Solving this system of equations — see S10.m2 (output file S11.txt) and the expressions
305 simplified in S8.nb (text version S9.txt) on https://github.com/jonathanmitchell88/CDMsSI
306 — we obtain

$$
z_1 z_2^2 z_3 z_4 z_6 z_8 \left(1 - z_6\right) \left(1 - z_8\right) \left(z_4 z_7 z_8 - z_5 z_9\right) = 0,
$$

307 which has no solutions since $z_i \in (0, 1)$ for all $i \in \{1, 2, \ldots, 8\}$ and the generating parameter
308 must be a generic point in the parameter space, i.e. $z_4 z_7 z_8 - z_5 z_9 \neq 0$. Thus, for CDM 5,
309 any two CDMs with different leaf labelings are distinguishable.

24

CDM 4

The proof is identical to that of CDM 5, but with the addition of $y_9 = z_9 = 1$. Again, see S8.nb (text version S9.txt) and S10.m2 (output file S11.txt). We obtain

$$z_1 z_2 z_3 z_4 z_5 z_6 z_8 \left(1 - z_6\right) \left(1 - z_7 z_8\right) = 0,$$

which again has no solutions. Thus, for CDM 4, any two CDMs with different leaf labelings are distinguishable.

CDM 3

See S8.nb (text version S9.txt) for proofs of the following claims.

For leaf labeling pairs $(o, (a, (b, c)))$ and $(o, (c, (b, a)))$, $(o, (a, (c, b)))$ and $(o, (b, (c, a)))$ and $(o, (b, (a, c)))$ and $(o, (c, (a, b)))$,

$$\begin{cases} \min\left(r_{0011} r_{1100}, r_{0101} r_{1010}, r_{0110} r_{1001}\right) = r_{0101} r_{1010}, \\ \min\left(r_{0011} r_{1100}, r_{0101} r_{1010}, r_{0110} r_{1001}\right) = r_{0110} r_{1001}, \\ \min\left(r_{0011} r_{1100}, r_{0101} r_{1010}, r_{0110} r_{1001}\right) = r_{0011} r_{1100}, \end{cases}$$

respectively, where each equation corresponds to a leaf labeling pair. Thus, any CDM from one pair is distinguishable from a CDM from another pair.

All that is left is to prove that CDMs from an arbitrary pair, for example, $(o, (a, (b, c)))$ and $(o, (c, (b, a)))$, are distinguishable. For leaf labeling $(o, (a, (b, c)))$, but not $(o, (c, (b, a)))$,

$$r_{0011} r_{1001} r_{1010} - r_{1011}^2 = 0.$$

For leaf labeling $(o, (c, (b, a)))$, but not $(o, (a, (b, c)))$,

$$r_{0011} r_{1001} r_{1010} - r_{1011}^2 > 0.$$

Thus for CDM 3, any two CDMs with different leaf labelings are distinguishable.

CDM 2

See S8.nb (text version S9.txt) for proofs of the following claims.

327 The constraints for CDM 2 include those described above for CDM 3. Thus for CDM 2,

328 any two CDMs with different leaf labelings are distinguishable.

CDM 1

330 See S8.nb (text version S9.txt) for proofs of the following claims.

331 For leaf labelings $(o, (a, (b, c)))$, $(o, (b, (a, c)))$ and $(o, (c, (a, b)))$,

$$
\begin{cases}
r_{0101}r_{1010} = r_{0110}r_{1001} < r_{0011}r_{1100}, \\
r_{0011}r_{1100} = r_{0110}r_{1001} < r_{0101}r_{1010}, \\
r_{0011}r_{1100} = r_{0101}r_{1010} < r_{0110}r_{1001},
\end{cases}
$$

332 respectively. Thus, for CDM 1 any two CDMs with different leaf labelings are distinguishable.

333 $\square$

5A Proof of Theorem 6

335 **Theorem 6** (Distance on the topology of an N-taxon principal tree) *Let $\mathcal{T}$ be a principal*

336 *tree, with outgroup o. Suppose $\mathcal{T}$ is given the rooted triple metrization. Then the distance*

337 $d_{\mathcal{T}}(x, y)$ *between leaf taxa x and y is*

$$
d_{\mathcal{T}}(x, y) = \begin{cases}
0 & \text{if } x=y, \\
2N - 2 & \text{if } x \neq y \text{ and one of } x = o,\ y = o, \\
2\,|R_{x,y}| + 2 & \text{otherwise,}
\end{cases}
$$

338 *where $R_{x,y}$ is the set of rooted 4-taxon principal trees displayed on $\mathcal{T}$ with outgroup o*

339 *displaying both x and y, where x and y are non-sisters.*

340 *Proof* Clearly, if $x = y$ then $d_{\mathcal{T}}(x, y) = 0$.

341 Next suppose $x \neq y$ and one of $x = o$, $y = o$. With no loss of generality, assume $y = o$.

342 Then

$$
d_{\mathcal{T}}(x, y) = d_{\mathcal{T}}(x, o) = d_{\mathcal{T}}(x, v) + d_{\mathcal{T}}(v, o),
$$

26

where v is the most recent common ancestor (MRCA) of x and o. Since $y = o$, v must be the root of $\mathcal{T}$. Then from the rooted triple metrization, by the same arguments as Rhodes (2019),

$$d_{\mathcal{T}}(x, v) = d_{\mathcal{T}}(v, o) = N - 1$$

and

$$d_{\mathcal{T}}(x, o) = 2N - 2.$$

Finally, suppose $x \neq y$ and $x, y \neq o$. Again suppose that v is the MRCA of x and y. Then again by the same arguments as Rhodes (2019),

$$d_{\mathcal{T}}(x, y) = 2k - 2,$$

where k is the number of leaf taxa descended from v.

For x and y to be non-sisters on a rooted 4-taxon principal tree displayed on $\mathcal{T}$ with outgroup o, we require the leaf taxon that is not x, y or o to be one of the $k - 2$ leaf taxa descended from v that is not x or y. Thus,

$$|R_{x,y}| = k - 2$$

and

$$d_{\mathcal{T}}(x, y) = 2|R_{x,y}| + 2.$$

$\square$

6A Inferring topologies of N-taxon principal trees

We prove that consistent inference of the topology of the N-taxon principal tree follows from consistent inference of the principal trees of the displayed 4-taxon CDMs. However, it is possible that a displayed 4-taxon CDM does not meet the assumptions of Section 3.2. Specifically, even if an N-taxon CDM meets the assumptions, some displayed 4-taxon CDMs may have sister convergence. By assuming that all convergence

27

parameters of the N-taxon CDM are sufficiently "small", then all convergence parameters of the displayed 4-taxon CDMs, including those of sister convergence groups of the displayed 4-taxon CDMs are "small". Then all topologies of the displayed 4-taxon principal trees are inferred consistently by Algorithm 1.

To prove this result, we first prove a proposition similar to Proposition 1.2 of Haughton (1988). Proposition 1.2 states that if the generating model is among the set of candidate models, the probability that the model selected by the BIC is the generating model converges to 1. Our adaptation relaxes Proposition 1.2, such that none of the candidate models are the generating model, but some candidate models are sufficiently "close" to the generating model. That is, the generating parameter is a "small" perturbation from a point in the parameter space of a candidate model. We then use our proposition to prove that all topologies of the displayed 4-taxon principal trees are inferred consistently by Algorithm 1.

For the following proposition, $f(X, \phi) = \exp(X\phi - b(\phi))$ is the density for a regular exponential family, m_1 and m_2 are the parameter spaces of two models, $\operatorname{int}\Theta$ is the interior of some topological space Θ, $\overline{m}_1$ and $\overline{m}_2$ are the Zariski closures of m_1 and m_2, respectively and $E_\theta X_i = \nabla b(\theta)$ is the expected value of random variable X_i given generating parameter θ. The function $g(\phi) = \nabla b(\theta)\phi - b(\phi)$ for $\phi \in \Theta$ attains its unique maximum at θ (Barndorff-Nielsen 1978).

Proposition 4A *Let m_1 and m_2 be two different models satisfying $m_1 \cap m_2 = \emptyset$. Then there exists some $\theta \in \operatorname{int}\Theta$, $\theta \notin \overline{m}_1, \theta \notin \overline{m}_2$, with a neighborhood $\mathfrak{N}$ of θ such that $\mathfrak{N} \cap m_1 = \emptyset$, $\mathfrak{N} \cap m_2 \neq \emptyset$ and*

$$\lim_{n \to \infty} P_\theta^n \left(\gamma(n, 1) < \gamma(n, 2)\right) = 1.$$

Proof The proof requires only a slight modification to the proof of Proposition 1.2 of Haughton (1988).

28

From Haughton (1988), since $\mathfrak{N} \cap m_1 = \emptyset$,

$$\sup_{\phi \in m_1 \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) + \epsilon \leq \nabla b\left(\theta\right)\theta - b\left(\theta\right) \tag{9A}$$

and asymptotically with probability 1,

$$\left| \sup_{\phi \in m_i \cap \Theta} \left(Y_n\phi - b\left(\phi\right)\right) - \sup_{\phi \in m_i \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) \right| < \frac{\epsilon}{4}, \tag{10A}$$

where $\epsilon > 0$.

Although $\mathfrak{N} \cap m_2 \neq \emptyset$, $g\left(\phi\right)$ attains its maximum at θ and $\theta \notin \overline{m}_2$. Thus, we can choose $\widetilde{\epsilon} > 0$ such that

$$\sup_{\phi \in m_2 \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) + \widetilde{\epsilon} = \nabla b\left(\theta\right)\theta - b\left(\theta\right). \tag{11A}$$

We consider the two possible signs of the argument of the absolute value in Inequality (10A). If

$$\sup_{\phi \in m_i \cap \Theta} \left(Y_n\phi - b\left(\phi\right)\right) - \sup_{\phi \in m_i \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) \geq 0,$$

then from Inequality 10A,

$$\sup_{\phi \in m_1 \cap \Theta} \left(Y_n\phi - b\left(\phi\right)\right) < \sup_{\phi \in m_1 \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) + \frac{\epsilon}{4}.$$

Similarly, if

$$\sup_{\phi \in m_i \cap \Theta} \left(Y_n\phi - b\left(\phi\right)\right) - \sup_{\phi \in m_i \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) < 0,$$

then

$$\sup_{\phi \in m_1 \cap \Theta} \left(Y_n\phi - b\left(\phi\right)\right) < \sup_{\phi \in m_1 \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right)$$

$$< \sup_{\phi \in m_1 \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) + \frac{\epsilon}{4}.$$

Thus, from Inequalities (9A) and (10A), asymptotically with probability 1,

$$\sup_{\phi \in m_1 \cap \Theta} \left(Y_n\phi - b\left(\phi\right)\right) < \sup_{\phi \in m_1 \cap \Theta} \nabla b\left(\theta\right)\phi - b\left(\phi\right) + \frac{\epsilon}{4} \leq \nabla b\left(\theta\right)\theta - b\left(\theta\right) - \frac{3\epsilon}{4}. \tag{12A}$$

29

By similar arguments, from Inequality (10A) and Equation (11A), asymptotically with probability 1,

$$\sup_{\phi \in m_2 \cap \Theta} (Y_n \phi - b(\phi)) > \sup_{\phi \in m_2 \cap \Theta} \nabla b(\theta) \phi - b(\phi) - \frac{\epsilon}{4} = \nabla b(\theta) \theta - b(\theta) - \widetilde{\epsilon} - \frac{\epsilon}{4}. \quad (13A)$$

By Inequalities (12A) and (13A),

$$
\begin{aligned}
\sup_{\phi \in m_1 \cap \Theta} (Y_n \phi - b(\phi)) &< \nabla b(\theta) \theta - b(\theta) - \frac{3\epsilon}{4} \\
&= \nabla b(\theta) \theta - b(\theta) - \widetilde{\epsilon} - \frac{\epsilon}{4} + \widetilde{\epsilon} - \frac{\epsilon}{2} \\
&< \sup_{\phi \in m_2 \cap \Theta} (Y_n \phi - b(\phi)) + \widetilde{\epsilon} - \frac{\epsilon}{2} \\
&= \sup_{\phi \in m_2 \cap \Theta} (Y_n \phi - b(\phi)) - \delta,
\end{aligned}
$$

where $\delta = \frac{\epsilon}{2} - \widetilde{\epsilon}$.

If it is possible to choose $\delta > 0$, then asymptotically with probability 1,

$$\sup_{\phi \in m_1 \cap \Theta} (Y_n \phi - b(\phi)) + \delta < \sup_{\phi \in m_2 \cap \Theta} (Y_n \phi - b(\phi)).$$

We are free to choose any $\theta \in \operatorname{int} \Theta$. Thus, we choose θ to be an arbitrarily small perturbation from some point in m_2. Then $\widetilde{\epsilon} > 0$ is arbitrarily small and $\delta > 0$. The remainder of the proof then follows from Haughton (1988).

$\square$

A convergence group on the generating N-taxon CDM may be a sister convergence group on some displayed 4-taxon CDMs and a non-sister convergence group on others. Thus, we must assume that all convergence parameters of the generating N-taxon CDM are "small" relative to the divergence parameters.

Next, we adapt Theorem 3 of Steel (1992) to prove that the N-taxon principal tree can be identified from the set of 4-taxon principal trees that include the outgroup.

30

410 **Theorem 5A** (Steel, 1992) *For a set of rooted triples R, $\langle R \rangle = \{T\}$ if and only if R is*
411 *consistent with T, and for each internal edge e of T there is a rooted triple in R which*
412 *distinguishes e.*

413   The consequence of Theorem 5A of Steel (1992) is that if all trees of a set of
414 (binary) rooted 3-taxon trees R are displayed on a (binary) rooted N-taxon tree T
415 and each internal edge of T is an internal edge of at least one tree in R, then T is
416 the only N-taxon tree that displays all the 3-taxon trees of R. In other words, the
417 N-taxon tree T can be identified from the set of 3-taxon trees R.

418   Steel (1992) note that an analogous theorem exists for unrooted quartets. Thus, the
419 *unrooted* principal tree of the N-taxon CDM can be identified from the set of $\binom{N-1}{3}$
420 topologies of *unrooted* 4-taxon principal trees that include the outgroup displayed on
421 the *unrooted* principal tree of the N-taxon CDM. The principal tree of the N-taxon
422 CDM is then rooted by the outgroup.

6A.1 Proof of Theorem 7

424 Finally, from Proposition 4A and Theorem 5A adapted to unrooted quartets that
425 include the outgroup, we can prove Theorem 7.

426 **Theorem 7** *Suppose CDM $\mathcal{N}$ has topology of principal tree $\mathcal{T}$. Suppose the BIC is used*
427 *for model selection in step 2 of Algorithm 1. Suppose $\widehat{\mathcal{T}}$ is the estimate of $\mathcal{T}$ inferred by*
428 *Algorithm 1. Then there exists some constant $c > 0$ such that if the largest convergence*
429 *parameter of $\mathcal{N}$ is less than c,*

$$\lim_{n \to \infty} \mathbb{P}\left(\widehat{\mathcal{T}} = \mathcal{T}\right) = 1.$$

31

₄₃₀ *Proof* Suppose $\mathcal{N}$ has a displayed 4-taxon CDM $\mathcal{N}_4$ with topology of principal tree $\mathcal{T}_4 =$
₄₃₁ $(o, (a, (b, c)))$. Then from the proof of Theorem 5, for $\mathcal{N}_4$,

$$r_{0011}r_{1001}r_{1010} - r_{1011}^2 = 0,$$

₄₃₂ while for some 4-taxon CDM with topology of principal tree $\mathcal{T}_4' \neq (o, (a, (b, c)))$,

$$r_{0011}r_{1001}r_{1010} - r_{1011}^2 > 0.$$

₄₃₃ Suppose m_1 corresponds to the union of parameter spaces for CDMs $1-5$ for the topology
₄₃₄ of principal tree $\mathcal{T}_4$. Suppose also that m_2 corresponds to the union of parameter spaces for
₄₃₅ CDMs $1-5$ for any 4-taxon topology of principal tree that is not $\mathcal{T}_4$. Then $m_1 \cap m_2 = \emptyset$
₄₃₆ — recall from Assumption 11 that the generating parameter is a generic point. Suppose
₄₃₇ $\theta \notin \overline{m}_1, \overline{m}_2$. Then if $c > 0$ is sufficiently small, Proposition 4A holds and m_1 is selected by
₄₃₈ the BIC asymptotically with probability 1.

₄₃₉ Next, we prove the claim that the set of inferred topologies of 4-taxon principal trees
₄₄₀ equals the set of topologies of the principal trees of the 4-taxon CDMs displayed on $\mathcal{N}$. Then
₄₄₁ from the adaptation of Theorem 5A to unrooted quartets, the topology of the principal tree of
₄₄₂ $\mathcal{N}$ is the only topology that displays all inferred 4-taxon principal trees. Thus, any consistent
₄₄₃ supertree inference method used in step 3 of Algorithm 1 infers the topology of the principal
₄₄₄ tree of $\mathcal{N}$ consistently and the proof is complete.

₄₄₅ All that is left to prove is the claim that the probability of the set of inferred 4-taxon
₄₄₆ principal trees equalling the set of topologies of principal trees of 4-taxon CDMs displayed
₄₄₇ on $\mathcal{N}$ converges to 1.

₄₄₈ Suppose A_i is the event where the topology of the ith 4-taxon principal tree is inferred
₄₄₉ incorrectly, given some arbitrary order. Then, by Proposition 4A, there exists some sample
₄₅₀ size n such that for $n' > n$, $\mathbb{P}(A_i) < \epsilon_i$ for some arbitrarily small $\epsilon_i > 0$. Then by Boole's
₄₅₁ inequality,

$$\mathbb{P}\left(\cup_{i=1}^{\binom{N-1}{3}} A_i\right) \leq \sum_{i=1}^{\binom{N-1}{3}} \mathbb{P}(A_i) < \sum_{i=1}^{\binom{N-1}{3}} \epsilon_i,$$

32

an arbitrarily small positive quantity. Thus, the set of topologies of the inferred 4-taxon principal trees of step 2 of Algorithm 1 equals the set of topologies of the principal trees of the 4-taxon CDMs displayed on $\mathcal{N}$ with probability converging to 1.

$\square$

7A Proof of Proposition 8

Proposition 8 *Let CDM $\mathcal{N}_i$ have topology of principal tree $\mathcal{T}_i$ and $t_{i,j}$ be the epoch length of epoch j. Let $\mathcal{N}_1$ be the tree, with $\mathcal{T}_1 = (o, (a, (b, c)))$, $t_{1,2} \to \infty$ and $t_{1,3} \to 0$. Let $\mathcal{N}_2$ and $\mathcal{N}_3$ be CDMs with $\mathcal{T}_2 = (o, (b, (a, c)))$ and $\mathcal{T}_3 = (o, (c, (a, b)))$, each with a non-sister convergence group $\{\{b\}, \{c\}\}$ in the tip epoch, with $t_{2,4}, t_{3,4} \to \infty$. Then the sets of possible phylogenetic tensors of $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$ converge to the same set.*

Proof Consider phylogenetic tensors for $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$ with arbitrary finite, positive substitution rates and epoch times. Let apostrophe superscripts denote parameters of $\mathcal{N}_2$.

33

For generic parameters on $\mathcal{N}_1$, after setting $x_6 = x_7 = x_8 = x_9 = x_{10} = x_{11} = 1$, Equations (7A) reduce to

$$
\begin{cases}
f_{0011} = x_4 x_5, \\[4pt]
f_{0101} = x_2 x_3 x_4, \\[4pt]
f_{0110} = x_2 x_3 x_5, \\[4pt]
f_{0111} = x_2 x_3 x_4 x_5, \\[4pt]
f_{1001} = x_1 x_2 x_4, \\[4pt]
f_{1010} = x_1 x_2 x_5, \\[4pt]
f_{1011} = x_1 x_2 x_4 x_5, \\[4pt]
f_{1100} = x_1 x_3, \\[4pt]
f_{1101} = x_1 x_2 x_3 x_4, \\[4pt]
f_{1110} = x_1 x_2 x_3 x_5, \\[4pt]
f_{1111} = x_1 x_3 x_4 x_5, \\[4pt]
\quad f_{\delta} = x_1 x_2 x_3 x_4 x_5.
\end{cases}
$$

Taking the limit as the epoch time of the second epoch diverges and the epoch time of

the tip epoch converges to 0 is equivalent to $x_1, x_2, x_3 \to 0$ and $x_4, x_5 \to 1$. Thus,

$$
\begin{cases}
f_{0011} \to 1, \\
f_{0101} \to 0, \\
f_{0110} \to 0, \\
f_{0111} \to 0, \\
f_{1001} \to 0, \\
f_{1010} \to 0, \\
f_{1011} \to 0, \\
f_{1100} \to 0, \\
f_{1101} \to 0, \\
f_{1110} \to 0, \\
f_{1111} \to 0, \\
\quad f_{\delta} \to 0.
\end{cases}
$$

35

For generic parameters on $\mathcal{N}_2$, after setting $x_7' = x_8' = x_9' = x_{10}' = x_{11}' = 1$ and permuting leaf labels, Equations (7A) reduce to

$$
\begin{cases}
f_{0011} = 1 - x_6' \left(1 - x_2' x_3' x_5'\right), \\
f_{0101} = x_4' x_5' x_6', \\
f_{0110} = x_2' x_3' x_4' x_6', \\
f_{0111} = x_2' x_3' x_4' x_5' x_6', \\
f_{1001} = x_1' x_2' x_5' x_6', \\
f_{1010} = x_1' x_3' x_6', \\
f_{1011} = x_1' x_2' x_3' x_5' x_6', \\
f_{1100} = x_1' x_2' x_4', \\
f_{1101} = x_1' x_2' x_4' x_5' x_6', \\
f_{1110} = x_1' x_2' x_3' x_4' x_6', \\
f_{1111} = x_1' x_4' \left(x_2' \left(1 - x_6'\right) + x_3' x_5' x_6'\right), \\
\quad f_\delta = x_1' x_2' x_3' x_4' x_5' x_6'.
\end{cases}
$$

470   Taking the limit as the epoch time of the tip epoch diverges is equivalent to $x'_1, x'_4, x'_6 \to 0$.
471 Thus,

$$
\begin{cases}
f_{0011} \to 1, \\
f_{0101} \to 0, \\
f_{0110} \to 0, \\
f_{0111} \to 0, \\
f_{1001} \to 0, \\
f_{1010} \to 0, \\
f_{1011} \to 0, \\
f_{1100} \to 0, \\
f_{1101} \to 0, \\
f_{1110} \to 0, \\
f_{1111} \to 0, \\
\quad f_\delta \to 0.
\end{cases}
$$

472   It is straightforward to show that Equations (7A) have the same limits for $\mathcal{N}_3$ and this
473 is omitted.

474   □

8A   Controlling overfitting the CDM

476 For Algorithm 2 that infers the convergence groups of the N-taxon CDM, we introduce
477 two constraints to avoid overfitting the CDM with too many convergence groups. The
478 constraints are used to determine whether convergence groups are inferred on the
479 CDM in a stepwise fashion.

480   The constraints ensure that only a small number of expected proportions of con-
481 verging quartets can exceed their observed values. Furthermore, convergence groups
482 can only be included if expected proportions exceed their observed values by small

37

amounts. Suppose $\boldsymbol{O}$ and $\boldsymbol{E}$ are the matrices of observed and expected proportions of converging quartets. Suppose $C = \{c_1, c_2\}$ is an arbitrary convergence group, with $c_1 \cup c_2 \subset [N]$. Then $a_1 = \max_{i \in c_1, j \in c_2} \left| [\boldsymbol{O}]_{ij} - [\boldsymbol{E}]_{ij} \right|$ and $a_2 = \frac{1}{(N-1)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_{ij}$, where $\delta_{ij} = 0$ if $[\boldsymbol{E}]_{ij} \leq [\boldsymbol{O}]_{ij}$ and 1 otherwise. We include the convergence group on the CDM only if $a_1 \leq A_1$ and $a_2 \leq A_2$, where $A_1, A_2 \in [0, 1]$. Note that the denominator of a_2 has $N - 1$ since the outgroup is not involved in convergence. ($[\boldsymbol{O}]_{iN} = [\boldsymbol{O}]_{Nj} = [\boldsymbol{E}]_{iN} = [\boldsymbol{E}]_{Nj} = 0$ for all $i, j \in \{1, 2, \ldots, N\}$.)

Further control of overfitting is achieved with a multiple comparisons correction, favoring 4-taxon trees over non-tree CDMs. For a given 4-taxon set that includes the outgroup taxon, the model selection criterion values are first converted into weights, for example, AIC or BIC (Burnham and Anderson 2004). These weights are a "tree weight" corresponding to the AIC or BIC of the tree and a "non-tree weight" corresponding to the AIC or BIC of the best fitting non-tree CDM. Tree weights could then be multiplied by some positive constant $b \geq 1$ to achieve further control for overfitting. A multiple comparisons correction, such as the Holm-Bonferroni method (Holm 1979), could then be applied to the tree weights over all 4-taxon sets that include the outgroup taxon, as if the weights were p-values. If the tree is "rejected", then the non-tree CDM with the lowest AIC or BIC is selected.

9A Proof of Proposition 9

Proposition 9 *For convergence group $C = \{c_1, c_2\}$ on CDM $\mathcal{N}$, let $a \in c_1$ and $b \in c_2$. Let v be the MRCA node of a and b, X_v be the set of leaf taxa descending from v and $X_C = c_1 \cup c_2$. Then the expected proportion of converging quartets for $\{a, b\}$ is*

$$\frac{|X_v \setminus X_C|}{N - 3} = \frac{|X_v| - |X_C|}{N - 3},$$

where $|X_v|$ and $|X_C|$ are the cardinalities of sets X_v and X_C.

38

Proof To determine the expected proportions of converging quartets, suppose taxa a and b are converging. Then convergence between these taxa can only be inferred on 4-taxon CDMs with topology of principal tree $(o, (a, (b, c)))$ or $(o, (b, (a, c)))$, for some arbitrary taxon c. With no loss of generality, we assume that the topology of the principal tree of some 4-taxon CDM is $(o, (a, (b, c)))$. To determine the expected proportions, we must determine the number of 4-taxon CDMs displayed on $\mathcal{N}$, displaying both a and b where they appear as non-sisters.

We start with the rooted tree $(o, (a, b))$ and append taxon c and include a convergence group C. One edge corresponding to the convergence group C must be ancestral to a, while the other must be ancestral to b. Thus, for C to be a non-sister convergence group, the remaining taxon c must be placed on an edge directly descended from v, corresponding to a speciation event before the epoch C is in. Thus, c could be any of the $|X_v \setminus X_C| = |X_v| - |X_C|$ taxa out of the $N - 3$ possible taxa that are not o, a or b. $\qquad\square$

10A Proof of Proposition 11

Proposition 11 *An arbitrary pair of distinct convergence groups on CDM $\mathcal{N}$ share no pair of converging leaf taxa.*

Proof Suppose C_1 and C_2 are two distinct convergence groups on $\mathcal{N}$. By Assumption 5 of Section 3.2, there can be at most one convergence group in each epoch. Thus, C_1 is either in an epoch before or after C_2. With no loss of generality, we assume that C_1 is in an epoch before C_2.

In order to share at least one pair of converging taxa, C_2 must be nested in C_1. However, by Assumption 9 of Section 3.2, there can be no convergence groups nested in other convergence groups.

$\qquad\square$

11A   Proof of Proposition 12

We assume that the topology of the principal tree of $\mathcal{N}$ is known. However, we note that if it is not known, from Theorem 7 it can be inferred consistently.

Proposition 12 *The set of all convergence groups on CDM $\mathcal{N}$ can be identified from the set of displayed 4-taxon CDMs after suppressing sister convergence groups.*

Proof The set of displayed 4-taxon CDMs after suppressing sister convergence groups defines a matrix of proportions of converging quartets. However, in general the set of all convergence groups on $\mathcal{N}$ cannot be identified from the matrix (see Figure 5). Instead, we can identify a *set* of possible sets of convergence groups on $\mathcal{N}$ that correspond to the matrix of proportions of converging quartets. Since the set of displayed 4-taxon CDMs after suppressing sister convergence groups is assumed known, for the remainder of the proof we can restrict to this set of sets of convergence groups. We must then prove that we can identify the specific set of all convergence groups of $\mathcal{N}$.

If $\mathcal{N}$ is a tree, then the set of displayed 4-taxon CDMs after suppressing sister convergence groups is a set of trees. Thus, by Corollary 10, the matrix of proportions of converging quartets is the zero matrix. Alternatively, if $\mathcal{N}$ is not a tree, then $\mathcal{N}$ must have at least one non-sister convergence group. Call one such non-sister convergence group $C = \{c_1, c_2\}$, with v the most recent common ancestral node of c_1 and c_2. Then by Proposition 9, the expected proportion of converging quartets for $a \in c_1$ and $b \in c_2$ is $\frac{|X_v| - |X_C|}{N-3}$, where X_v is the set of all taxa descending from v and $|X_C| = |c_1| + |c_2|$. By the definition of non-sister convergence groups, $|X_v| - |X_C| > 0$. Thus, the matrix of converging quartets is not the zero matrix. Thus, if $\mathcal{N}$ is a tree, the set of convergence groups can be identified from the set of displayed 4-taxon CDMs after suppressing sister convergence groups via the matrix of converging quartets.

For the remainder of the proof, we can assume that $\mathcal{N}$ is not a tree. Then the set of non-sister convergence groups defines a set S of 4-taxon CDMs displayed on $\mathcal{N}$ with non-sister convergence groups after suppressing sister convergence groups — note that 4-taxon CDMs

40

of S can have one or two non-sister convergence groups. Suppose similarly that S' is a set of 4-taxon CDMs defined by a set of non-sister convergence groups not on $\mathcal{N}$ but with the same matrix of proportions of converging quartets as the set of non-sister convergence groups on $\mathcal{N}$. We must prove that there exists some 4-taxon CDM in S that is not in S'. Then we can identify the set of convergence groups on $\mathcal{N}$ from the set of 4-taxon CDMs.

We prove that there is some 4-taxon CDM in S that is not in S'. We first consider an arbitrary 4-taxon CDM $\mathcal{N}_4$ in S. Consider arbitrary leaf taxon pair $\{a, b\}$, where $a \in c_1$ and $b \in c_2$. Furthermore, assume $c \in X_v \setminus X_C$. Then with no loss of generality, we can assume the topology of the principal tree of $\mathcal{N}_4$ is $(o, (b, (a, c)))$.

Suppose that $C' = \{c_1', c_2'\}$ is one such non-sister convergence group that defines S', with c_1', c_2', v', X_v' and $X_{C'}$ as in Proposition 9. Now consider 4-taxon CDM $\mathcal{N}_4'$, defined by C' and on leaf taxon set $\{o, a, b, c\}$, with topology of principal tree $(o, (b, (a, c)))$. Since we require a non-sister convergence group on $\mathcal{N}_4'$ where a and b are both converging, we must have either $a \in c_1'$ and $b \in c_2'$ or $a \in c_2'$ and $b \in c_1'$. With no loss of generality, we assume that $a \in c_1'$ and $b \in c_2'$. Then $c_1 \subseteq c_1'$ or $c_1 \supset c_1'$. Similarly, $c_2 \subseteq c_2'$ or $c_2 \supset c_2'$. Both v and v' are the MRCA of a and b. Thus $v' = v$.

Now assume that $X_{C'} = X_C$. Then $c_1' = c_1$ and $c_2' = c_2$ and in turn, $C' = C$. Thus, S' is defined by a set of convergence groups that includes C and the 4-taxon CDM is in S'. Thus, we can assume that $X_{C'} \neq X_C$ and we cannot have both $c_1' = c_1$ and $c_2' = c_2$. However, since the matrices of proportions of converging quartets must be the same for the two sets of convergence groups, we must have

$$\frac{|X_v| - |X_C|}{N - 3} = \frac{|X_{v'}| - |X_{C'}|}{N - 3},$$

which simplifies to $|X_C| = |X_{C'}|$, since $v' = v$. Thus, either $c_1 \subset c_1'$ and $c_2 \supset c_2'$ or $c_1 \supset c_1'$ and $c_2 \subset c_2'$. With no loss of generality, we assume that $c_1 \subset c_1'$ and $c_2 \supset c_2'$.

Then there exists some choice of c such that $c \in c_1' \setminus c_1$. For $c_2 \supset c_2'$, there must similarly be some taxon $d \in c_2 \setminus c_2'$. Thus, we are assuming that $N \geq 5$ — the outgroup and taxa a, b, c and d. Then $a, c \in c_1'$ and $b \in c_2'$. Thus, before suppressing sister convergence groups to form $\mathcal{N}_4'$, C' must correspond with a sister convergence group on the 4-taxon CDM on leaf taxa $\{o, a, b, c\}$ — see Figure 1A for a graphical depiction of C and C'. Then any other choice

41

583 of convergence group that defines S', say C'', must satisfy $c_1'' \supset c_1$ and the claim follows.

584 Finally, since we have assumed $N \geq 5$, we must also consider $N = 4$. For $N = 4$, it is clear

585 from the identifiability and distinguishability of all CDMs with no sister convergence that
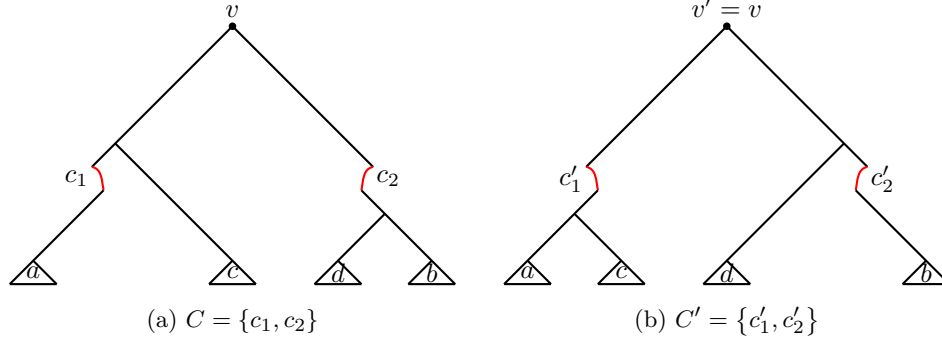
586 the claim holds.

587 $\square$

(a) $C = \{c_1, c_2\}$ (b) $C' = \{c_1', c_2'\}$

Fig. 1A Convergence groups C and C'. Labels c_1, c_2, c_1' and c_2' indicate the set of leaf taxa below that edge. Triangles are displayed CDMs. Labels inside triangles indicate one of possibly many taxa on leaves of those displayed CDMs. There may be more displayed CDMs not drawn below v that are not below either C or C'

12A Proof of Theorem 13

589 **Theorem 13** *Suppose CDM $\mathcal{N}$ has topology of principal tree $\mathcal{T}$ and convergence groups $\mathcal{G}$.*

590 *Suppose for all l, $\alpha_l = \beta_l$. Suppose for convergence group $\mathcal{C}_i = \{c_{1,i}, c_{2,i}\}$ that if $a \in c_{1,i} \cup c_{2,i}$,*

591 *then $a \notin c_{1,j} \cup c_{2,j}$ for any $j \neq i$. Suppose $\mathcal{T}$ is input into Algorithm 2, the BIC is used*

592 *for model selection in step 2, there are no multiple comparisons corrections and the tolerance*

593 *criterion is $u = 1$. Suppose $\widehat{\mathcal{G}}$ is the estimate of $\mathcal{G}$ inferred by Algorithm 2. Then there exists*

594 *some constant $c > 0$ such that if the largest convergence parameter of $\mathcal{N}$ is less than c,*

$$\lim_{n \to \infty} \mathbb{P}\left(\widehat{\mathcal{G}} = \mathcal{G}\right) = 1.$$

42

Proof We start by finding expressions for the transformed phylogenetic tensors for various 4-taxon CDMs with and without sister convergence. We prove that the CDMs with sister convergence are not distinguishable from the CDMs with the sister convergence groups suppressed. Thus, regardless of whether the 4-taxon CDMs have sister convergence groups or not, the non-sister convergence group is inferred consistently.

Since $\alpha_l = \beta_l$, $\gamma = 0$ and the transformed phylogenetic tensor for a 4-taxon CDM of Equation (6A) simplifies to

$$
\widehat{P} = \begin{bmatrix}
1 \\
0 \\
0 \\
r_{0011} \\
0 \\
r_{0101} \\
r_{0110} \\
0 \\
0 \\
r_{1001} \\
r_{1010} \\
0 \\
r_{1100} \\
0 \\
0 \\
r_{1111}
\end{bmatrix}.
$$

With no loss of generality, we assume the topology of the principal tree of an arbitrary 4-taxon CDM displayed on $\mathcal{N}$ is $(o\,(a,(b,c)))$. Then of the possible convergence groups on the 4-taxon CDM, the convergence group in the epoch closest to the root is the sister convergence group $C = \{\{a\}, \{b,c\}\}$. Thus, we first consider the distinguishability of two 4-taxon CDMs, one a tree and the other with this convergence group. For both CDMs we assume the tip epoch has epoch time 0. The first, which we call $\mathcal{N}_{4,1}$, is the tree $(o,(a,(b,c)))$. Since the tip

43

epoch has epoch time 0, taxa b and c are identical. The second CDM, which we call $\mathcal{N}_{4,2}$, has

a single convergence group, $C = \{\{a\}, \{b, c\}\}$, followed by a speciation event involving b and

c. Again, since the tip epoch has epoch time 0, taxa b and c are identical. See Figure 2A for

a graphical depiction of the two CDMs. Suppose $\mathcal{N}_{4,1}$ has parameters with no apostrophes

and $\mathcal{N}_{4,2}$ has parameters with apostrophes.

For $\mathcal{N}_{4,1}$ (see Mathematica file S12.nb (text version S13.txt) on https://github.com/

jonathanmitchell88/CDMsSI for a derivation),

$$\begin{cases} r_{0011} = & 1, \\ r_{0101} = & x_2 x_3, \\ r_{0110} = & x_2 x_3, \\ r_{1001} = & x_1 x_2, \\ r_{1010} = & x_1 x_2, \\ r_{1100} = & x_1 x_3, \\ r_{1111} = & x_1 x_3. \end{cases}$$

For $\mathcal{N}_{4,2}$ (see Mathematica file S12.nb (text version S13.txt) for a derivation),

$$\begin{cases} r_{0011} = & 1, \\ r_{0101} = & 1 - x_4' \left(1 - x_2' x_3'\right), \\ r_{0110} = & 1 - x_4' \left(1 - x_2' x_3'\right), \\ r_{1001} = & x_1' x_2' x_4', \\ r_{1010} = & x_1' x_2' x_4', \\ r_{1100} = & x_1' x_3' x_4', \\ r_{1111} = & x_1' x_3' x_4'. \end{cases}$$

616   For $\mathcal{N}_{4,1}$,

$$\begin{cases}
x_1 = & \sqrt{\dfrac{r_{1001}r_{1100}}{r_{0101}}}, \\[2mm]
x_2 = & \sqrt{\dfrac{r_{0101}r_{1001}}{r_{1100}}}, \\[2mm]
x_3 = & \sqrt{\dfrac{r_{0101}r_{1100}}{r_{1001}}}, \\[2mm]
r_{0101} = & r_{0110}, \\[2mm]
r_{1001} = & r_{1010}, \\[2mm]
r_{1100} = & r_{1111}.
\end{cases}$$

617   Since $x_1, x_2, x_3 \in (0,1)$, for $\mathcal{N}_{4,1}$,

$$\begin{cases}
r_{0101} = & r_{0110}, \\[2mm]
r_{1001} = & r_{1010}, \\[2mm]
r_{1100} = & r_{1111}, \\[2mm]
r_{0101}r_{1001} < & r_{1100}, \\[2mm]
r_{0101}r_{1100} < & r_{1001}, \\[2mm]
r_{1001}r_{1100} < & r_{0101}.
\end{cases}$$

618   Similarly, for $\mathcal{N}_{4,2}$,

$$\begin{cases}
r_{0101} = & r_{0110}, \\[2mm]
r_{1001} = & r_{1010}, \\[2mm]
r_{1100} = & r_{1111}.
\end{cases}$$

619   Since we are assuming that all convergence parameters of $\mathcal{N}$ are less than some constant
620    $c > 0$, we can assume that $x_4' = 1 - \epsilon$, where $\epsilon > 0$ is some small positive constant. Then for
621    $\mathcal{N}_{4,2}$,

$$\begin{cases}
r_{1100} - r_{0101}r_{1001} = & x_1'x_3'\left(1 - x_2'^2\right) + O\left(\epsilon\right), \\[2mm]
r_{1001} - r_{0101}r_{1100} = & x_1'x_2'\left(1 - x_3'^2\right) + O\left(\epsilon\right), \\[2mm]
r_{0101} - r_{1001}r_{1100} = & x_2'x_3'\left(1 - x_1'^2\right) + O\left(\epsilon\right).
\end{cases}$$

45

Since $c > 0$ can be chosen, there exists some choice of $\epsilon > 0$ sufficiently small such that for $\mathcal{N}_{4,2}$,

$$
\begin{cases}
r_{0101} = & r_{0110}, \\
r_{1001} = & r_{1010}, \\
r_{1100} = & r_{1111}, \\
r_{0101}r_{1001} < & r_{1100}, \\
r_{0101}r_{1100} < & r_{1001}, \\
r_{1001}r_{1100} < & r_{0101}.
\end{cases}
$$

Thus, $\mathcal{N}_{4,1}$ and $\mathcal{N}_{4,2}$ are not distinguishable for this choice of $c > 0$. Thus, any 4-taxon CDM with $\alpha_l = \beta_l$ and this sister convergence group is not distinguishable from the CDM that results from suppressing the sister convergence. Thus, to determine the transformed phylogenetic tensor of any 4-taxon CDM with $\alpha_l = \beta_l$, we can assume there is no sister convergence in this epoch.

The next closest epoch to the root that could have a convergence group is the epoch just after taxa b and c have diverged. Thus, we compare the tree $(o, (a, (b, c)))$, which we call $\mathcal{N}_{4,3}$, and the CDM with topology of principal tree $(o, (a, (b, c)))$ and sister convergence group $\{\{b\}, \{c\}\}$ in the tip epoch, which we call $\mathcal{N}_{4,4}$. See Figure 3A for a graphical depiction of the two CDMs. Again, suppose $\mathcal{N}_{4,3}$ has parameters with no apostrophes and $\mathcal{N}_{4,4}$ has parameters with apostrophes.

46

635   For $\mathcal{N}_{4,3}$ (see Mathematica file S12.nb (text version S13.txt) for a derivation),

$$
\begin{cases}
r_{0011} = & x_4 x_5, \\
r_{0101} = & x_2 x_3 x_4, \\
r_{0110} = & x_2 x_3 x_5, \\
r_{1001} = & x_1 x_2 x_4, \\
r_{1010} = & x_1 x_2 x_5, \\
r_{1100} = & x_1 x_3, \\
r_{1111} = & x_1 x_3 x_4 x_5.
\end{cases}
$$

636   For $\mathcal{N}_{4,4}$ (see Mathematica file S12.nb (text version S13.txt) for a derivation),

$$
\begin{cases}
r_{0011} = & 1 - x_6' \left(1 - x_4' x_5' \right), \\
r_{0101} = & x_2' x_3' x_4' x_6', \\
r_{0110} = & x_2' x_3' x_5' x_6', \\
r_{1001} = & x_1' x_2' x_4' x_6', \\
r_{1010} = & x_1' x_2' x_5' x_6', \\
r_{1100} = & x_1' x_3', \\
r_{1111} = & x_1' x_3' \left(1 - x_6' \left(x_4' x_5' \right) \right).
\end{cases}
$$

637   For $\mathcal{N}_{4,3}$,

$$
\begin{cases}
x_1 = & \sqrt{\dfrac{r_{1001} r_{1100}}{r_{0101}}}, \\
x_2 = & \sqrt{\dfrac{r_{0110} r_{1001}}{r_{0011} r_{1100}}}, \\
x_3 = & \sqrt{\dfrac{r_{0101} r_{1100}}{r_{1001}}}, \\
x_4 = & \sqrt{\dfrac{r_{0011} r_{0101}}{r_{0110}}}, \\
x_5 = & \sqrt{\dfrac{r_{0011} r_{0110}}{r_{0101}}}, \\
r_{0101} r_{1010} = & r_{0110} r_{1001}, \\
r_{0011} r_{1100} = & r_{1111}.
\end{cases}
$$

47

Since $x_1, x_2, x_3, x_4, x_5 \in (0, 1)$, for $\mathcal{N}_{4,3}$,

$$
\begin{cases}
r_{0101}r_{1010} = & r_{0110}r_{1001}, \\
r_{0011}r_{1100} = & r_{1111}, \\
r_{0011}r_{0101} < & r_{0110}, \\
r_{0011}r_{0110} < & r_{0101}, \\
r_{0101}r_{1100} < & r_{1001}, \\
r_{0110}r_{1001} < & r_{0011}r_{1100}, \\
r_{1001}r_{1100} < & r_{0101}.
\end{cases}
$$

Similarly, for $\mathcal{N}_{4,4}$,

$$
\begin{cases}
r_{0101}r_{1010} = & r_{0110}r_{1001}, \\
r_{0011}r_{1100} = & r_{1111}.
\end{cases}
$$

Since we are assuming that all convergence parameters of $\mathcal{N}$ are less than some constant $c > 0$, we can assume that $x_6' = 1 - \epsilon$, where $\epsilon > 0$ is some small positive constant. Then for $\mathcal{N}_{4,4}$,

$$
\begin{cases}
r_{0110} - r_{0011}r_{0101} = & x_2'x_3'x_5'\left(1 - x_4'^2\right) + O\left(\epsilon\right), \\
r_{0101} - r_{0011}r_{0110} = & x_2'x_3'x_4'\left(1 - x_5'^2\right) + O\left(\epsilon\right), \\
r_{1001} - r_{0101}r_{1100} = & x_1'x_2'x_4'\left(1 - x_3'^2\right) + O\left(\epsilon\right), \\
r_{0011}r_{1100} - r_{0110}r_{1001} = & x_1'x_3'x_4'x_5'\left(1 - x_2'^2\right) + O\left(\epsilon\right), \\
r_{0101} - r_{1001}r_{1100} = & x_2'x_3'x_4'\left(1 - x_1'^2\right) + O\left(\epsilon\right).
\end{cases}
$$

48

643   Since $c > 0$ can be chosen, there exists some choice of $\epsilon > 0$ sufficiently small such that

644   for $\mathcal{N}_{4,4}$,

$$
\begin{cases}
r_{0101}r_{1010} = & r_{0110}r_{1001}, \\[2mm]
r_{0011}r_{1100} = & r_{1111}, \\[2mm]
r_{0011}r_{0101} < & r_{0110}, \\[2mm]
r_{0011}r_{0110} < & r_{0101}, \\[2mm]
r_{0101}r_{1100} < & r_{1001}, \\[2mm]
r_{0110}r_{1001} < & r_{0011}r_{1100}, \\[2mm]
r_{1001}r_{1100} < & r_{0101}.
\end{cases}
$$

645   Thus, $\mathcal{N}_{4,3}$ and $\mathcal{N}_{4,4}$ are not distinguishable for this choice of $c > 0$. Thus, any 4-taxon

646   CDM with $\alpha_l = \beta_l$ and this sister convergence group is not distinguishable from the CDM

647   that results from suppressing the sister convergence. Thus, to determine the transformed

648   phylogenetic tensor of any 4-taxon CDM with $\alpha_l = \beta_l$, we can again assume there is no sister

649   convergence in this epoch.

650   By the assumption that no leaf taxa belong to more than one convergence group, there

651   can be no more than one convergence group on any arbitrary 4-taxon CDM displayed on

652    $\mathcal{N}$. Thus, taking into consideration $\mathcal{N}_{4,1}$ and $\mathcal{N}_{4,2}$ not being distinguishable and $\mathcal{N}_{4,3}$ and

653    $\mathcal{N}_{4,4}$ not being distinguishable, we can conclude that any arbitrary 4-taxon CDM displayed

654   on $\mathcal{N}$ is not distinguishable from the 4-taxon CDM that results from suppressing any sister

655   convergence group, which is one of CDM $1 - 3$ of Figure 3.

₆₅₆ Next, we establish that CDM 3 is identifiable under these assumptions. For this CDM,
₆₅₇ which we call $\mathcal{N}_{4,5}$ (see Mathematica file S12.nb (text version S13.txt)),

$$
\begin{cases}
r_{0011} = & x_4 x_5 x_6 x_7, \\[4pt]
r_{0101} = & x_2 x_3 x_4 x_6 x_8, \\[4pt]
r_{0110} = & x_7 x_8 \left(1 - x_6 \left(1 - x_2 x_3 x_5\right)\right), \\[4pt]
r_{1001} = & x_1 x_2 x_4, \\[4pt]
r_{1010} = & x_1 x_2 x_5 x_6 x_7, \\[4pt]
r_{1100} = & x_1 x_3 x_6 x_8, \\[4pt]
r_{1111} = & x_1 x_4 x_7 x_8 \left(x_2 \left(1 - x_6\right) + x_3 x_5 x_6\right).
\end{cases}
$$

₆₅₈ In terms of the set of parameters $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9\}$ of Section 3A.1,

$$
\begin{cases}
r_{0011} = & y_4 y_5 y_6, \\[4pt]
r_{0101} = & y_2 y_3 y_4 y_6, \\[4pt]
r_{0110} = & y_7 \left(1 - y_6\right) + y_2 y_3 y_5 y_6, \\[4pt]
r_{1001} = & y_1 y_2 y_4, \\[4pt]
r_{1010} = & y_1 y_2 y_5 y_6, \\[4pt]
r_{1100} = & y_1 y_3 y_6, \\[4pt]
r_{1111} = & y_1 \left(y_2 y_4 y_7 \left(1 - y_6\right) + y_3 y_5 y_5 y_6\right).
\end{cases}
$$

₆₅₉ In S14.m2 (output file S15.txt) on https://github.com/jonathanmitchell88/CDMsSI, we
₆₆₀ see that the set of parameters $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$ is identifiable. It follows that CDMs
₆₆₁ 1 and 2 are also identifiable.

₆₆₂ Thus, using similar arguments to those of the proof of Theorem 7, with probability
₆₆₃ converging to 1, step 2 of Algorithm 2 infers all the 4-taxon CDMs with the outgroup that
₆₆₄ are displayed on $\mathcal{N}$ after suppressing sister convergence groups.

₆₆₅ If $\mathcal{N}$ is a tree, then $s = 0$ in step 4 of Algorithm 2, the algorithm terminates and the
₆₆₆ tree is returned. If $\mathcal{N}$ is not a tree, since $u = 1$, a potential convergence group on $\mathcal{N}$ is only
₆₆₇ considered if, for all pairs of converging taxa in the convergence group, the inferred 4-taxon

50

CDMs with that pair of taxa as non-sisters all have the pair converging. Thus, asymptotically with probability 1, only convergence groups on $\mathcal{N}$ can be on the inferred N-taxon CDM. If not all convergence groups of $\mathcal{N}$ have been included on the inferred CDM, then there are some elements of $\boldsymbol{O}$ that are non-zero corresponding to elements of $\boldsymbol{E}$ that are zero. These elements correspond to the pairs of converging taxa in convergence groups of $\mathcal{N}$ that are not yet on the inferred CDM. Including these convergence groups on the inferred CDM makes these elements of $\boldsymbol{E}$ equal to the corresponding elements of $\boldsymbol{O}$, decreasing the sum of squared differences. Once all convergence groups of $\mathcal{N}$ have been appended to the inferred CDM, $\boldsymbol{O} = \boldsymbol{E}$. Thus, no more convergence groups can be appended to the inferred CDM to decrease the sum of squared differences and the algorithm terminates.

$\square$
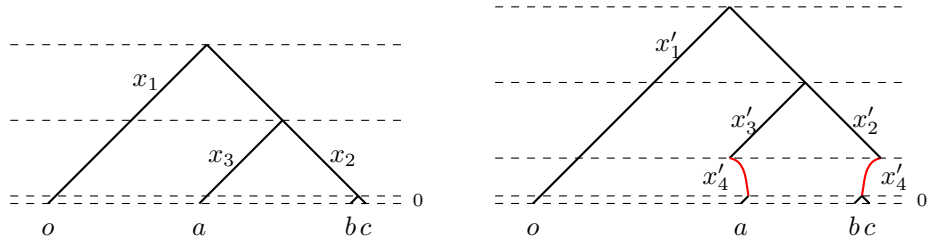

(a) $\mathcal{N}_{4,1}$ — x_1 represents the entire outgroup edge when unrooted.

(b) $\mathcal{N}_{4,2}$ — x_1' represents the entire outgroup edge when unrooted.

Fig. 2A Two CDMs that are not distinguishable under the assumptions of Theorem 13

13A Inferring convergence group orders on N-taxon CDMs

The next algorithms infer partial orders on the convergence groups and determine whether or not there is a convergence group in the tip epoch. CDMs 4 and 5 have two convergence groups and thus provide power to determine convergence group orders. Whether or not there is a convergence group in the tip epoch can also be determined
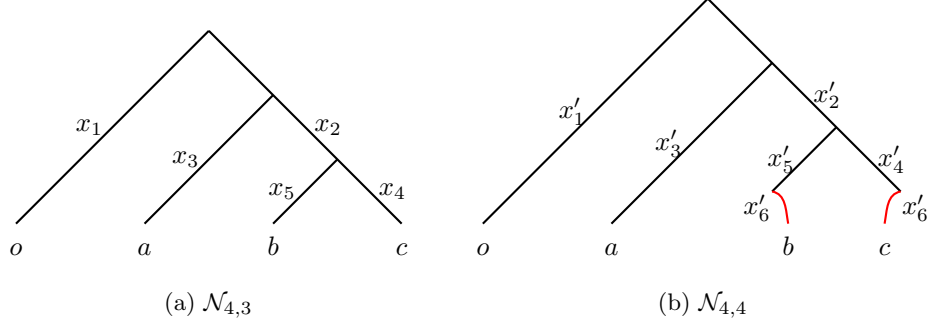
51

(a) $\mathcal{N}_{4,3}$ (b) $\mathcal{N}_{4,4}$

Fig. 3A Two CDMs that are not distinguishable under the assumptions of Theorem 13

from the inferred 4-taxon CDMs. For example, CDM 2 and CDM 3 differ by CDM 2 having its convergence group in the tip epoch versus CDM 3 having its convergence group in the epoch before the tip epoch.

Suppose an inferred 4-taxon CDM has two non-sister convergence groups. One of the edges of the 4-taxon principal tree corresponds to a converging taxon in both convergence groups. The order of these convergence groups may not be determined by the matrix of edge partial orders from Algorithm 2. If the convergence group order is not determined, we determine which order is best supported by selecting CDMs from those with the appropriate convergence groups with a model selection procedure.

For convergence groups C_i and C_j, whose partial order has not been determined, we tally proportions of 4-taxon CDMs best supported by the two orders to obtain a matrix of "observed" convergence group order ratios. Convergence group orders are resolved in a stepwise fashion by minimizing the sum of squared differences between matrices of observed and "expected" partial convergence group order ratios. The matrix of inferred convergence group order ratios is updated after each convergence group order is inferred. We discard 4-taxon CDMs with convergence group orders not consistent with the matrix of inferred convergence group orders.

Suppose an arbitrary convergence group is $C_i = \{c_{1,i}, c_{2,i}\}$. On the N-taxon CDM, if $|c_{1,i}| > 1$ and/or $|c_{2,i}| > 1$ or C_i is in an epoch before another convergence group,

52

$_{704}$ then C_i cannot be in the tip epoch. For other convergence groups, whether they are

$_{705}$ in the tip epoch or not must be inferred.

$_{706}$ For each 4-taxon CDM with a fixed leaf labeling with a possible convergence group

$_{707}$ in the tip epoch, we determine which CDM is best supported among the two CDMs,

$_{708}$ for example, CDM 2 versus CDM 3 or CDM 4 versus CDM 5. For C_i, we tally the 4-

$_{709}$ taxon CDMs displaying the given convergence group with and without the convergence

$_{710}$ group in the tip epoch.

$_{711}$ If C_i corresponds to a greater proportion of 4-taxon CDMs with the convergence

$_{712}$ group in the tip epoch than any other convergence group and the proportion is greater

$_{713}$ than some cutoff, for example, half, then we infer that C_i is in the tip epoch. We

$_{714}$ retain only one possible CDM for each 4-taxon set after the convergence group order

$_{715}$ has been assigned and it has been determined which, if any, convergence group is in

$_{716}$ the tip epoch.

$_{717}$ Note that some convergence group orders may still be undefined. Suppose two

$_{718}$ convergence groups do not have an order defined by the edge partial order of the

$_{719}$ principal tree or the orders of other convergence groups. Suppose both convergence

$_{720}$ groups are only ever present on 4-taxon CDMs where one convergence group is a sister

$_{721}$ convergence group. Then there will be no information to resolve the order of these two

$_{722}$ convergence groups. We leave these convergence group orders unresolved. Thus, we

$_{723}$ have a *partial* order on the convergence groups. Algorithms 1A and 2A for inferring

$_{724}$ convergence group orders and any convergence group in the tip epoch then follow.

$_{725}$ We do not prove consistency of inference of the convergence group partial orders

$_{726}$ from Algorithm 1A. This is because Theorem 13 assumes that no leaf taxa belong to

$_{727}$ more than one convergence group. Thus, all 4-taxon CDMs displayed on $\mathcal{N}$ have at

$_{728}$ most one non-sister convergence group and there are no convergence group orders to

$_{729}$ infer. Furthermore, we do not prove consistency of inference of the convergence groups

$_{730}$ in the tip epoch.

53

Algorithm 1A Convergence group order inference

Input: N-taxon CDM $\widehat{\mathcal{N}}$ comprising N-taxon topology of principal tree $\widehat{\mathcal{T}}$ and list of convergence groups $\widehat{\mathcal{G}}$, as well as $\binom{N-1}{3} \times 27$ matrix of model selection criterion values $\boldsymbol{M}$ and matrix of partial edge orders $\boldsymbol{P}$.

1. Initialize empty list of inferred 4-taxon CDMs L_Q. Initialize $k \times k$ matrix of observed convergence group orders $\boldsymbol{O}$ as zero matrix, where k is length of list $\widehat{\mathcal{G}}$. Initialize $k \times k$ matrix $\boldsymbol{E}$ of expected convergence group orders as convergence group orders defined by $\boldsymbol{P}$, with $[\boldsymbol{E}]_{ij} = 1$ if convergence group i before j and 0 otherwise.
2. For each 4-taxon set that includes outgroup o, with model selection criterion, select CDM from those displayed on $\widehat{\mathcal{N}}$ and permitted by $\boldsymbol{E}$ and append to L_Q.
3. For all i, j, compute $[\boldsymbol{O}]_{ij}$ as proportion of inferred 4-taxon CDMs displaying convergence groups i and j, where i is before j.
4. Compute initial sum of squared differences between elements of $\boldsymbol{O}$ and $\boldsymbol{E}$, $s = \sum_{i=1}^{k} \sum_{j=1}^{k} \left([\boldsymbol{O}]_{ij} - [\boldsymbol{E}]_{ij} \right)^2$.
5. Assign new order between two convergence groups that minimizes s.
6. Update $\boldsymbol{E}$ and s to reflect newly inferred convergence group order. Suppose new order is convergence group x before y. Then all convergence groups above x are also above y and all convergence groups below y are also below x. If no pairs of convergence groups left to assign orders to, terminate algorithm.
7. Return to Step 5.

Output: N-taxon CDM $\widehat{\mathcal{N}}$ comprising N-taxon topology principal tree $\widehat{\mathcal{T}}$ and list of convergence groups $\widehat{\mathcal{G}}$, as well as $\binom{N-1}{3} \times 27$ matrix of model selection criterion values $\boldsymbol{M}$, matrix of partial edge orders $\boldsymbol{P}$ and matrix of expected convergence group orders $\boldsymbol{E}$.

However, if all inferred 4-taxon CDMs that include the outgroup are the 4-taxon CDMs displayed on the generating N-taxon CDM after suppressing sister convergence groups, then it is straightforward to prove that Algorithm 1A correctly infers all orders of convergence groups of the generating N-taxon CDM that can be determined from the displayed 4-taxon CDMs. Furthermore, it is also straightforward to prove that Algorithm 2A correctly infers which, if any, convergence group of the generating N-taxon CDM is in the tip epoch.

14A Proof of Proposition 14

Proposition 14 *All edge lengths of the principal tree of each of CDM $1-5$ are identifiable.*

Algorithm 2A Inference of convergence groups in tip epochs

Input: N-taxon CDM $\widehat{\mathcal{N}}$ comprising N-taxon topology principal tree $\widehat{\mathcal{T}}$ and list of convergence groups $\widehat{\mathcal{G}}$, as well as $\binom{N-1}{3} \times 27$ matrix of model selection criterion values $\boldsymbol{M}$, matrix of partial edge orders $\boldsymbol{P}$, matrix of expected convergence group orders $\boldsymbol{E}$ and tolerance $\tau \in [0, 1]$.

1. Initialize empty list of inferred 4-taxon CDMs L_Q. Initialize vector $\boldsymbol{D}$ of length k of convergence groups in tip epoch as zero vector, where k is length of list $\widehat{\mathcal{G}}$.
2. For each 4-taxon set that includes outgroup o, select CDM from those displayed on $\widehat{\mathcal{N}}$ and permitted by $\boldsymbol{E}$ with model selection criterion and append to L_Q.
3. For all i, if convergence group $C_i = \{c_{1,i}, c_{2,i}\}$ satisfies $|c_{1,i}| = |c_{2,i}| = 1$ and is not before any other convergence group of $\widehat{\mathcal{N}}$, compute $[\boldsymbol{D}]_i$ as proportion of inferred 4-taxon CDMs with C_i in tip epoch.
4. If $\max_{i \in \{1,2,\dots,k\}} [\boldsymbol{D}]_i = [\boldsymbol{D}]_j$ and $D_j > \tau$, set $[\boldsymbol{D}]_j = 1$.

Output: N-taxon CDM $\widehat{\mathcal{N}}$ comprising N-taxon topology principal tree $\widehat{\mathcal{T}}$ and list of convergence groups $\widehat{\mathcal{G}}$, as well as $\binom{N-1}{3} \times 27$ matrix of model selection criterion values $\boldsymbol{M}$, matrix of partial edge orders $\boldsymbol{P}$, matrix of expected convergence group orders $\boldsymbol{E}$ and vector of convergence groups in tip epoch $\boldsymbol{D}$.

---

740   *Proof* Using the parameterization of Section 3A.1, for CDM 5, with principal tree

741    $(o, (a, (b, c)))$, the sums of edge lengths between leaf taxa are

$$
\begin{cases}
d_{o,a} = & l_1 + l_3 + l_6 + l_8 + l_9 + l_{11} = -\log\left(x_1 x_3 x_6 x_8 x_9 x_{11}\right) = -\log\left(y_1 y_3 y_6 y_8\right), \\[2mm]
d_{o,b} = & l_1 + l_2 + l_5 + l_6 + l_7 = -\log\left(x_1 x_2 x_5 x_6 x_7\right) = -\log\left(y_1 y_2 y_5 y_6\right), \\[2mm]
d_{o,c} = & l_1 + l_2 + l_4 + l_9 + l_{10} = -\log\left(x_1 x_2 x_4 x_9 x_{10}\right) = -\log\left(y_1 y_2 y_4 y_8\right), \\[2mm]
d_{a,b} = & l_2 + l_3 + l_5 + 2l_6 + l_7 + l_8 + l_9 + l_{11} = -\log\left(x_2 x_3 x_5 x_6^2 x_7 x_8 x_9 x_{11}\right) \\[2mm]
 = & -\log\left(y_2 y_3 y_5 y_6^2 y_8\right), \\[2mm]
d_{a,c} = & l_2 + l_3 + l_4 + l_6 + l_8 + 2l_9 + l_{10} + l_{11} = -\log\left(x_2 x_3 x_4 x_6 x_8 x_9^2 x_{10} x_{11}\right) \\[2mm]
 = & -\log\left(y_2 y_3 y_4 y_6 y_8^2\right), \\[2mm]
d_{b,c} = & l_4 + l_5 + l_6 + l_7 + l_9 + l_{10} = -\log\left(x_4 x_5 x_6 x_7 x_9 x_{10}\right) = -\log\left(y_4 y_5 y_6 y_8\right).
\end{cases}
$$

742   From Equations (8A), the set $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9\}$ is identifiable. Thus, the set

743    $\{d_{o,a}, d_{o,b}, d_{o,c}, d_{a,b}, d_{a,c}, d_{b,c}\}$ is also identifiable for CDM 5. Solving for the lengths of the

744 edges of the principal tree,

$$
\begin{cases}
l_o = & \frac{1}{2}\left(d_{o,a} + d_{o,b} - d_{a,b}\right), \\[2mm]
l_a = & \frac{1}{2}\left(d_{o,a} - d_{o,b} + d_{a,b}\right), \\[2mm]
l_b = & \frac{1}{2}\left(d_{a,b} - d_{a,c} + d_{b,c}\right), \\[2mm]
l_c = & \frac{1}{2}\left(-d_{a,b} + d_{a,c} + d_{b,c}\right), \\[2mm]
l_{bc} = & \frac{1}{2}\left(-d_{o,a} + d_{o,b} + d_{a,c} - d_{b,c}\right),
\end{cases}
$$

745 where l_o is the sum of divergence parameters along the two edges of the principal tree whose

746 parent node is the root, l_a, l_b and l_c are the sums of divergence and possibly convergence

747 parameters along the terminal edges whose descendent leaf taxa are a, b and c respectively

748 and l_{bc} is the sum of divergence parameters along the edge whose descendent leaf taxa are b

749 and c.

750   It follows that all edge lengths are also identifiable for CDMs $1-4$ since expressions for

751 the sums of edge lengths are the same, except that some $y_i = 1$.

752   □

753 15A   Proof of Proposition 15

754 **Proposition 15** *All convergence parameters of each of CDM $2-5$ are identifiable.*

755 *Proof* On CDM 5, parameters $y_6 = x_6$ and $y_8 = x_9$ are identifiable. Thus, the convergence

756 parameters $l_6 = a_6 + b_6 = -\log(y_6)$ and $l_9 = a_9 + b_9 = -\log(y_8)$ are identifiable. Thus, for

757 all other CDMs with these convergence parameters, they are also identifiable. □

758 16A   Proof of Proposition 16

759 **Proposition 16** *The root parameter $\gamma = [\mathbf{\Pi}]_0 - [\mathbf{\Pi}]_1$, where $[\mathbf{\Pi}]_0$ and $[\mathbf{\Pi}]_1$ are the*

760 *probabilities of states 0 and 1 at the root, respectively, is identifiable on each of CDM $1-5$.*

56

Proof From Equation (6A) for the phylogenetic tensor of CDM 5, $q_{0001} = q_{0010} = q_{0100} =$

$q_{1000} = \gamma$. Thus, γ is identifiable for CDM 5. Since all other CDMs are nested in CDM 5 and

none correspond to generic values of γ — instead they correspond to some generic values of

x_i or y_i — γ is also identifiable for CDMs $1 - 4$. $\hfill\square$

17A Proof of Theorem 17

Theorem 17 *Suppose CDM $\mathcal{N}$ has topology of principal tree $\mathcal{T}$, convergence groups $\mathcal{G}$, prin-*

cipal tree edge lengths $\boldsymbol{l}$, root parameter γ and convergence parameters $\boldsymbol{v}$. Suppose $\mathcal{T}$, $\mathcal{G}$,

convergence group partial orders and tip epoch convergence groups of $\mathcal{N}$ are input into Algo-

rithm 3. Suppose in step 4 of Algorithm 3 only 4-taxon sets for which 4-taxon CDMs displayed

on $\mathcal{N}$ have no sister convergence are considered. Suppose that for each convergence group of

$\mathcal{G}$ — say $C_a = \{c_{1,a}, c_{2,a}\}$ — there is at least one 4-taxon CDM displayed on $\mathcal{N}$ with no

sister convergence where $x \in c_{1,a}$, $y \in c_{2,a}$ are non-sister leaf taxa on the displayed CDM.

Suppose further that matrix $\boldsymbol{X}$ in step 6 of Algorithm 3 has rank $2N - 3$. Suppose $\widehat{\boldsymbol{l}}$, $\widehat{\gamma}$ and

$\widehat{\boldsymbol{v}}$ are the estimates of $\boldsymbol{l}$, γ and $\boldsymbol{v}$, respectively, inferred by Algorithm 3. Then for any $\epsilon_{>}0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\widehat{\boldsymbol{l}} - \boldsymbol{l}\right| > \epsilon\right) = 0, \quad \lim_{n\to\infty} \mathbb{P}\left(|\widehat{\gamma} - \gamma| > \epsilon\right) = 0, \quad \lim_{n\to\infty} \mathbb{P}\left(|\widehat{\boldsymbol{v}} - \boldsymbol{v}| > \epsilon\right) = 0,$$

where $\left|\widehat{\boldsymbol{l}} - \boldsymbol{l}\right|$ and $|\widehat{\boldsymbol{v}} - \boldsymbol{v}|$ involve l^1 norms.

Proof In step 4 of Algorithm 3, only 4-taxon sets that include the outgroup for which 4-taxon

CDMs displayed on $\mathcal{N}$ have no sister convergence are considered. Thus, all such 4-taxon

CDMs displayed on $\mathcal{N}$ are CDM $1 - 5$. Since some 4-taxon sets may not be considered, we

cannot yet assume that all parameters are identifiable. However, for a given 4-taxon set that

is considered, from the proof of Proposition 14, all sums of edge lengths between leaf taxa in

the 4-taxon set are identifiable. From Propositions 15 and 16, all convergence parameters on

the 4-taxon CDM displayed on $\mathcal{N}$ and the root parameter γ are also identifiable.

Thus, for the given 4-taxon set, the estimates of sums of edge lengths between taxa formed

from the sums of maximum likelihood estimates of parameters converge in probability to the

sums of edge lengths between taxa for $\mathcal{N}$. Likewise, the maximum likelihood estimates of the

convergence parameters converge in probability to the convergence parameters on $\mathcal{N}$ and the maximum likelihood estimate of γ also converges in probability to γ. Thus, it follows that when averaging over all 4-taxon sets that are considered, the estimates of the sums of edge lengths between taxa converge in probability to the values for $\mathcal{N}$.

Now, since the matrix $\boldsymbol{X}$ has rank $2N - 3$, $\boldsymbol{X}^T\boldsymbol{X}$ is invertible. It follows that $\widehat{\boldsymbol{l}}$ also converges in probability to $\boldsymbol{l}$ in step 7 of Algorithm 3. By assumption, for each convergence group of $\mathcal{G}$ there is at least one 4-taxon CDM displayed on $\mathcal{N}$ where two converging taxa of the convergence group are non-sister taxa and there is no sister convergence. Thus, each convergence parameter of $\widehat{\boldsymbol{v}}$ is estimated at least once. Thus, $\widehat{\boldsymbol{v}}$ converges in probability to $\boldsymbol{v}$. Finally, since γ is fixed across all 4-taxon CDMs displayed on $\mathcal{N}$, to be consistently estimated it only needs to be estimated for one 4-taxon CDM displayed on $\mathcal{N}$. Thus, $\widehat{\gamma}$ converges in probability to γ.

$\square$

References

Barndorff-Nielsen O (1978) Information and exponential families in statistical theory. Wiley, New York

Burnham K, Anderson D (2004) Model selection and multi-model inference: A Practical information-theoretic approach, 2nd edn. Springer-Verlag New York

Haughton DM (1988) On the choice of a model to fit data from an exponential family. The Annals of Statistics pp 342–355. https://doi.org/10.1214/aos/1176350709

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics pp 65–70

Rhodes JA (2019) Topological metrizations of trees, and new quartet methods of tree inference. IEEE/ACM Transactions on Computational Biology and Bioinformatics 17(6):2107–2118. https://doi.org/10.1109/tcbb.2019.2917204

Steel M (1992) The complexity of reconstructing trees from qualitative characters and subtrees. Journal of Classification 9:91–116. https://doi.org/10.1007/bf02618470

Sumner JG, Fernández-Sánchez J, Jarvis PD (2012a) Lie markov models. Journal of Theoretical Biology 298:16–31. https://doi.org/10.1016/j.jtbi.2011.12.017

Sumner JG, Holland B, Jarvis P (2012b) The algebra of the general markov model on phylogenetic trees and networks. Bulletin of Mathematical Biology 74(4):858–880. https://doi.org/10.1007/s11538-011-9691-z