

1 Online appendix for “Convergence-divergence  
2 models: Generalized pylogenetic trees modeling  
3 gene flow over time” for peer review in the  
4 Bulletin of Mathematical Biology

5 Jonathan D. Mitchell<sup>1,2\*</sup> and Barbara R. Holland<sup>1,2</sup>

6 <sup>1</sup>School of Natural Sciences (Mathematics), University of Tasmania,  
7 Hobart, TAS, Australia.

8 <sup>2</sup>ARC Centre of Excellence for Plant Success in Nature and  
9 Agriculture, University of Tasmania, Hobart, TAS, Australia.

10 \*Corresponding author(s). E-mail(s): [jonathanmitchell88@gmail.com](mailto:jonathanmitchell88@gmail.com);

11 Sections, algorithms, definitions, theorems, propositions, figures and expressions intro-  
12 duced in the main text are labeled numerically. Any introduced in the online appendix  
13 are labeled with a number followed by “A”.

## 14 1A Limiting behavior of converging taxa

### 15 1A.1 Proof of Proposition 1

16 Before proving the claim, we introduce some notation. Using similar notation to  
 17 [Sumner et al. \(2012b\)](#), for some arbitrary integer  $l \geq 1$  and  $X \in \{L_\alpha, L_\beta\}$ ,

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

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

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

20 Then we define

$$\begin{aligned} \mathcal{L}_\alpha^{[l]} &= \sum_{B \subseteq [l]} L_\alpha^{(B)} = (L_\alpha + I)^{\otimes l}, \\ \mathcal{L}_\beta^{[l]} &= \sum_{B \subseteq [l]} L_\beta^{(B)} = (L_\beta + I)^{\otimes l}, \end{aligned}$$

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

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

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

24 where

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

25 Then

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

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

27 The proof is split into four parts. We determine the rate matrix for an arbitrary  
 28 epoch in each part of the proof. 1) Instead of having  $N$  taxa, we assume that  $\mathcal{N}$  has  
 29  $l$  taxa, where  $l \in \{1, 2, \dots, N\}$ , all in the same convergence-divergence group in some  
 30 arbitrary epoch. 2) We assume that  $\mathcal{N}$  has  $N$  taxa, with the first  $l$  — according to the  
 31 indices  $i_1 i_2 \dots i_N$  and  $j_1 j_2 \dots j_N$  — in the same convergence-divergence group in the  
 32 epoch. 3) We determine the rate matrix corresponding to an arbitrary convergence-  
 33 divergence group with  $l$  taxa by permuting the taxon labels, which corresponds to  
 34 permuting the indices. 4) The rate matrix for the epoch is determined by summing the  
 35 rate matrices corresponding to all convergence-divergence groups in arbitrary epoch.

36 *Proof* 1) Suppose  $\mathcal{N}$  has only  $l$  taxa, where  $l \in \{1, 2, \dots, N\}$ , all in the same convergence-  
 37 divergence group in some epoch. Then the rate matrix for the epoch is

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

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

39 We first prove that

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

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

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

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

44 and in turn,

$$\tilde{Q}^{[1]} = \alpha (L_\alpha + I) + \beta (L_\beta + I) = \begin{bmatrix} \beta & \beta \\ \alpha & \alpha \end{bmatrix}.$$

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

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

46 We establish that if the claim is true, then

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

47 Since

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

48 it follows that

$$\tilde{Q}^{[m+1]} = \alpha (L_\alpha + I) \otimes \mathcal{L}_\alpha^{[m]} + \beta (L_\beta + I) \otimes \mathcal{L}_\beta^{[m]}$$

49 and thus the claim is established.

50 Then since

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

51 it follows that

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

52 Note that  $\sum_{i=1}^{2^l} \tilde{Q}_{ij}^{[l]} = \alpha + \beta$ . Thus,  $\sum_{i=1}^{2^l} Q_{ij}^{[l]} = 0$ . Summarizing,

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

53 From here onward, we let  $*$  represent the negative of the sum of all non-diagonal elements  
54 of column  $j$  of the rate matrix.

55 2) Suppose that  $\mathcal{N}$  has  $N$  taxa, with the first  $l$  — according to the indices  $i_1 i_2 \dots i_N$  and  
56  $j_1 j_2 \dots j_N$  — in a convergence-divergence group. Assume that this is the only convergence-  
57 divergence group in the epoch. That is, in the epoch the last  $N - l$  taxa are not in any  
58 convergence-divergence group — they will be assigned to convergence-divergence groups at

59 later steps in the proof. Then it follows directly from 1) that the rate matrix has elements

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

60 3) Consider a permutation in  $\mathfrak{S}_N$  of the leaf taxa. An arbitrary such permutation cor-  
 61 responds to an arbitrary re-ordering of the leaf taxa. We allow the permutation to act on  
 62  $V^{\otimes N}$  (the tensor product space of Sumner et al. (2012b)). For some arbitrary convergence-  
 63 divergence group  $C_r \in \mathcal{C}$  involving  $l$  taxa, there exists a permutation  $\sigma \in \mathfrak{S}_N$  such that  
 64  $\sigma(Q^{[l]} \otimes I^{\otimes N-l}) = Q^{[C_r]}$ , where  $Q^{[C_r]}$  is the rate matrix for the convergence-divergence  
 65 group  $C_r$  and all other leaf taxa belonging to no convergence-divergence group. Then by  
 66 Sumner et al. (2012b),

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

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

69 Then it follows directly from 2) that

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

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

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

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

72

□

## 73 1A.2 Proof of Theorem 2

74 For the proof, we assume an arbitrary phylogenetic tensor exists at the beginning of  
 75 the tip epoch. We then consider the effect of convergence-divergence groups in the tip  
 76 epoch on the phylogenetic tensor. As in the proof of Proposition 1A.1, the proof is  
 77 split into several parts. 1) We assume  $\mathcal{N}$  has  $N$  taxa, but  $\mathcal{C}$  has only one convergence-  
 78 divergence group  $C_a$  involving the first  $l$  taxa in the tip epoch. We determine the  
 79 elements of the transition matrix  $M^{[C_a]} = \exp(Q^{[l]}_t) \otimes I^{N-l}$ . 2) We determine the phy-  
 80 logenetic tensor in the limit as the epoch time of the tip epoch diverges. 3) We assume  
 81  $\mathcal{C}$  has  $k \geq 1$  convergence-divergence groups,  $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$ , and determine the  
 82 phylogenetic tensor in the limit as the epoch time of the tip epoch diverges.

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

$$\tilde{P} = \exp(Q^{[C_a]}_t) \cdot P',$$

87 where

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

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

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

89 Then

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

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

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

91 Focusing on  $(Q^{[l]})^2$ ,

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

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

93 By [Sumner et al. \(2012b\)](#),

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



94 It follows that

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

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

96 Returning to the Taylor series,

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

97 Now recall from the proof of Proposition 1A.1 that

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

98 Thus, for columns to sum to zero,

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

99

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

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

100

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

$$\lim_{t \rightarrow \infty} M_{ij}^{[l]} = \begin{cases} \frac{\alpha}{\alpha+\beta} & \text{if } \prod_{a=1}^l i_a = 1, \\ \frac{\beta}{\alpha+\beta} & \text{if } \prod_{a=1}^l (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

101

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

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

102

In summary, in the limit, the only rows of  $M^{[C_a]}$  with non-zero elements have the first

103

$l$  indices being either all 0 or all 1. Then in the limit, the only non-zero elements of  $\tilde{P}$  also

104

have the first  $l$  indices being either all 0 or all 1.

105

2) We let the substitution rates for convergence group  $C_k$  be  $\alpha_k, \beta_k > 0$  and recognize

106

that since  $\frac{\alpha_k}{\beta_k} = \frac{\alpha}{\beta}$ ,  $\frac{\alpha_k}{\alpha_k + \beta_k} = \frac{\alpha}{\alpha + \beta}$  and  $\frac{\beta_k}{\alpha_k + \beta_k} = \frac{\beta}{\alpha + \beta}$ . Then using the same arguments as in

107 Proposition 1A.1,  $\sigma(M^{[C_a]}) = M^{[C_k]}$  and

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

108 and

$$\lim_{t \rightarrow \infty} M_{ij}^{[C_k]} \cdot P' = \begin{cases} c_i^{[C_k]} > 0 & \text{if } \prod_{a \in C_k} i_a = 1 \text{ or } \prod_{a \in C_k} (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

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

110 3)  $\mathcal{N}$  has a tip epoch with epoch time  $t$  and set of convergence groups  $\mathcal{C} =$   
 111  $\{C_1, C_2, \dots, C_k\}$ . Then the phylogenetic tensor  $P$  representing the probabilities of combina-  
 112 tions of states at the leaves of the principal tree can be expressed as

$$\begin{aligned} P &= \exp(Q^{[\mathcal{C}]_t}) \cdot P' \\ &= \prod_{r=1}^k M^{[C_r]} \cdot P'. \end{aligned}$$

113 We prove that all elements of  $P$  converge to 0 except those where, for each  $C_r \in \mathcal{C}$ , all  
 114 taxa in  $C_r$  are in the same state.

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

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

116 We prove that

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

117 We prove this claim by induction on the  $k$  convergence groups. Note that in 2)  
 118 we have already proven the claim for the first convergence-divergence group  $C_k$  applied

119 to  $P'$ . Thus, all that remains it to prove that given the claim is true for  $P^{k-v+2} =$   
 120  $M^{[C_v]} \dots M^{[C_{v+1}]} \dots M^{[C_k]} \cdot P'$  for some  $v \in \{2, \dots, k\}$ , it must be true for  $P^{k-v+3} =$   
 121  $M^{[C_{v-1}]} \dots M^{[C_v]} \dots M^{[C_k]} \cdot P'$ .

122 We assume that

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

123 Then in the limit as  $t \rightarrow \infty$ , all elements of  $P^{k-v+2}$  are 0 except those where for all  
 124  $C_r \in \mathcal{C}$ ,  $r \in \{v, v+1, \dots, k\}$ , all taxa in  $C_r$  are in the same state.

125 For  $P_i^{k-v+3}$  to be non-zero, there must exist some index  $s$ , such that

$$M_{is}^{[C_{v-1}]} > 0 \tag{1A}$$

126 and

$$P_s^{k-v+2} > 0. \tag{2A}$$

127 For Equation (1A) to be true,

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

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

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

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

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

130 Constraint (5A) can be simplified to

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

131 In summary,

$$\lim_{t \rightarrow \infty} P_i = \begin{cases} c_i^{\left[\bigcup_{r=1}^k C_r\right]} > 0 & \text{if for all } r \in \{1, 2, \dots, k\}, \\ & \prod_{a \in C_r} i_a = 1 \text{ or } \prod_{a \in C_r} (1 - i_a) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

132

□

## 133 2A Parameter identifiability

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

137 Recall that rates and epoch times cannot be identified individually — only their  
 138 products can be identified. Differences in parameters between contiguous diverging  
 139 sections of an edge that span multiple epochs also cannot be identified. Instead,  
 140 an “average” over the sections can be identified. For example, suppose rate matrix  
 141  $Q_1$  applies over epoch time  $t_1$  to a diverging section of an edge immediately before  
 142 an event and rate matrix  $Q_2$  applies over epoch time  $t_2$  to a diverging section of  
 143 the edge immediately after the event. Then for the 2-state general Markov model,  
 144  $\exp(\hat{Q}(t_1 + t_2)) = \exp(Q_2 t_2) \exp(Q_1 t_1)$ , where  $\hat{Q}$  is again a rate matrix from the 2-  
 145 state general Markov model. Thus, we apply rate matrix  $\hat{Q}$  to both diverging sections  
 146 of the edge.

147 This lack of identifiability result follows from the 2-state general Markov model  
 148 forming a Lie algebra, sufficient for multiplicative closure of the model class ([Sumner  
 149 et al. 2012a](#)). Suppose  $\alpha_1$  and  $\beta_1$  and  $\alpha_2$  and  $\beta_2$  correspond with rate matrices  $Q_1$  and  
 150  $Q_2$ , respectively. Then by Definition 5,  $\frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2}$ . It is straightforward to show that  
 151 if  $\hat{\alpha}$  and  $\hat{\beta}$  are associated with  $\hat{Q}$ , then  $\frac{\hat{\alpha}}{\hat{\beta}} = \frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2}$ . Thus, the product of the two  
 152 transition matrices is replaced by a single “average” transition matrix.

153 All parameters except for those corresponding to the root distribution are of the  
 154 form  $l_i = a_i + b_i = \alpha_i t_i + \beta_i t_i = a_i \left(1 + \frac{b_i}{a_i}\right)$ , where  $i$  and  $j$  are arbitrary parameter  
 155 indices,  $a_i = \alpha_i t_i$  and  $b_i = \beta_i t_i$ . (Note that these parameters are scalars, whereas  
 156  $\exp\left(\widehat{Q}(t_1 + t_2)\right)$  is a matrix.) Since  $\frac{\alpha_i}{\beta_i} = \frac{\alpha_j}{\beta_j}$ , it follows that  $\frac{a_i}{b_i} = \frac{a_j}{b_j}$ .

157 Contiguous diverging sections of an edge — not separated by a converging section  
 158 — each have a single associated parameter  $l_i$ . Furthermore, each convergence group  
 159 has an associated parameter  $l_j$ , in common for all converging sections of edges in the  
 160 convergence group. In addition to parameters describing the convergence groups and  
 161 contiguous diverging sections, there is a parameter  $\gamma = \pi_0 - \pi_1 = \frac{-a_i + b_i}{a_i + b_i}$  describing  
 162 the difference in probabilities of states 0 and 1 on the root taxon.

163 To form the set of parameters of a CDM, we consider a particular unique set of  
 164 diverging and converging sections. Since differences in parameters between contiguous  
 165 diverging sections cannot be identified, the diverging sections we consider are those  
 166 sections on the principal tree between a node or converging section and another node  
 167 or converging section. Furthermore, since the exact root location on the outgroup edge  
 168 is not identifiable, we consider one diverging section to be the entire outgroup edge  
 169 when the principal tree of the CDM is unrooted. The converging sections correspond  
 170 to individual epochs where there is a convergence group. Converging sections cor-  
 171 respond to convergence parameters and diverging sections correspond to divergence  
 172 parameters.

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

### 179 **3A Identifiability of 4-taxon CDMs**

180 [Sumner et al. \(2012b\)](#) formally describe phylogenetic epoch models in their Defini-  
181 tion 6.1 and introduce notation to compute the phylogenetic tensors. We use the same  
182 notation for our CDMs.

183 For each 4-taxon CDM, the phylogenetic tensor  $P$  is transformed into the  
184 Hadamard basis  $\hat{P}$  by multiplying by  $H_{16} = H_2 \otimes H_2 \otimes H_2 \otimes H_2$ , where  $\otimes$  is the  
185 Kronecker product and

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

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

$$\begin{aligned}
 \hat{P} = H_{16} \cdot P = & \begin{bmatrix} q_{0000} \\ q_{0001} \\ q_{0010} \\ q_{0011} \\ q_{0100} \\ q_{0101} \\ q_{0110} \\ q_{0111} \\ q_{1000} \\ q_{1001} \\ q_{1010} \\ q_{1011} \\ q_{1100} \\ q_{1101} \\ q_{1110} \\ q_{1111} \end{bmatrix} \\
 = & \begin{bmatrix} 1 \\ \gamma \\ \gamma \\ \gamma^2 + (1 - \gamma^2) r_{0011} \\ \gamma \\ \gamma^2 + (1 - \gamma^2) r_{0101} \\ \gamma^2 + (1 - \gamma^2) r_{0110} \\ \gamma (\gamma^2 + (1 - \gamma^2) (r_{0011} + r_{0101} + r_{0110} - 2r_{0111})) \\ \gamma \\ \gamma^2 + (1 - \gamma^2) r_{1001} \\ \gamma^2 + (1 - \gamma^2) r_{1010} \\ \gamma (\gamma^2 + (1 - \gamma^2) (r_{0011} + r_{1001} + r_{1010} - 2r_{1011})) \\ \gamma^2 + (1 - \gamma^2) r_{1100} \\ \gamma (\gamma^2 + (1 - \gamma^2) (r_{0101} + r_{1001} + r_{1100} - 2r_{1101})) \\ \gamma (\gamma^2 + (1 - \gamma^2) (r_{0110} + r_{1010} + r_{1100} - 2r_{1110})) \\ \gamma^2 (\gamma^2 + (1 - \gamma^2) (r_{0011} + r_{0101} + r_{0110} + r_{1001} + r_{1010} + r_{1100} \\ - 2(r_{0111} + r_{1011} + r_{1101} + r_{1110} - 2\delta))) + (1 - \gamma^2)^2 r_{1111} \end{bmatrix}.
 \end{aligned} \tag{6A}$$



187 See Mathematica file S2.nb (text version S3.txt) on [https://github.com/](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
 188 [jonathanmitchell88/CDMs/tree/main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation) for a derivation of  
 189 Equation (6A) and equations for  $r_{0011}$ ,  $r_{0101}$ ,  $\dots$ ,  $r_{1111}$  and  $\delta$  for CDM 5. CDMs 1 – 4  
 190 are all nested in CDM 5. Thus, their phylogenetic tensors are also in the form of  
 191 Equation (6A).

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

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

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

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

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

208 does not help us to determine any invariants involving multiple variables  $r_{0011}, r_{0101},$   
 209  $\dots, r_{1111}, \delta.$ )

210 In the Macaulay2 file S4.m2 (output file S5.txt) on [https://github.](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
 211 [com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation) we derive the  
 212 (reduced) Gröbner basis for this ideal for a particular monomial ordering. Below we  
 213 outline how this Gröbner basis is computed.

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

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

216 The functions  $f_{0011} = f_{0011}(x_1, x_2, \dots, x_l), f_{0101} = f_{0101}(x_1, x_2, \dots, x_l), \dots,$   
 217  $f_{1111} = f_{1111}(x_1, x_2, \dots, x_l)$  and  $f_{\delta} = f_{\delta}(x_1, x_2, \dots, x_l)$  depend on the CDM in  
 218 question, e.g. CDM 5.

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

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

225 Note that  $q_{1111}$  is a function of both  $r_{1111}$  and  $\delta$ , the only element of  $\hat{P}$  that is  
 226 a function of either. Thus, the maximum cardinality set can include at most one of  
 227  $r_{1111}$  and  $\delta$ . In S4.m2 we find that when eliminating  $r_{1111}$  there are no generators that  
 228 involve  $\delta$ . Thus,  $\delta$  is another independent variable of the system, but not  $r_{1111}$ . Thus,  
 229 we have found the (reduced) Gröbner basis on our desired set of variables.

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

### 240 **3A.1 Proof of Proposition 3**

241 See S4.m2 (output file S5.txt) and S6.m2 (output file S7.txt) on [https://github.](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
 242 [com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation) for the com-  
 243 putations of the (reduced) Gröbner bases of the ideals in this proof.

244 *Proof* In S4.m2 (output file S5.txt), we see that there are 9 elements of  
 245  $\{r_{0011}, r_{0101}, \dots, r_{1110}, \delta\}$  that are free to vary. However, CDM 5 has 11 parameters exclud-  
 246 ing  $\gamma$ . Thus, this set of parameters is not identifiable. However, recall in Section 2A that  
 247 taking some products of  $x_i$  parameters may be required to obtain a set of identifiable  
 248 parameters. Since there are 9 elements of  $\{r_{0011}, r_{0101}, \dots, r_{1110}, \delta\}$  that are free to vary, we  
 249 desire a set of 9 parameters.

250 In S2.nb (text version S3.txt), we express  $f_{0011}, f_{0101}, \dots, f_{1111}, \delta$  in terms of the set of  
 251 parameters  $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9\}$ . Precisely,

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

252 In S6.m2 (output file S7.txt), we see that this set of parameters is identifiable. We note  
 253 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  
 254  $y_i \in (0, 1)$  for all  $i \in \{1, 2, \dots, 9\}$ . In S2.nb (text version S3.txt), we see that the solutions to

the system are

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

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.

□

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.

270 *Proof* By Proposition 4, if two CDMs have parameter spaces with different dimensions, then  
 271 they are distinguishable from each other. CDMs 1, 2, 3, 4 and 5 have parameter space  
 272 dimensions 6, 7, 8, 9 and 10 respectively, corresponding to the numbers of free parameters.

273 All that is left to prove is that any two CDMs that differ only in their leaf labelings are  
 274 distinguishable. The notation that follows is consistent with that of Section 3A. Recall that  
 275  $y_i \in (0, 1)$  for all  $i \in \{1, 2, \dots, 9\}$ .

## 276 CDM 5

277 See S8.nb (text version S9.txt) on [https://github.com/jonathanmitchell88/CDMs/tree/](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
 278 [main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation) for proofs of the following claims.

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

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

280 while for the other leaf labelings

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

281 We need only show that CDMs with these two leaf labelings are distinguishable. To do  
 282 this, we show that the intersection of the parameter spaces of the two CDMs is the empty set.  
 283 Letting  $y_i$  be the parameters corresponding to leaf labeling  $(o, (a, (b, c)))$  and  $z_i$  corresponding  
 284 to  $(o, (a, (c, b)))$ , we equate the equations for each element of the two phylogenetic tensors

285 and solve for the  $z_i$  parameters,

$$\left\{ \begin{array}{l} y_4 y_5 y_6 y_8 = z_4 z_5 z_6 z_8, \\ y_9 (1 - y_8) + y_2 y_3 y_4 y_6 y_8 = z_8 (z_7 (1 - z_6) + z_2 z_3 z_5 z_6), \\ y_8 (y_7 (1 - y_6) + y_2 y_3 y_5 y_6) = z_9 (1 - z_8) + z_2 z_3 z_4 z_6 z_8, \\ y_2 y_3 y_4 y_5 y_6 y_8 = z_2 z_3 z_4 z_5 z_6 z_8, \\ y_1 y_2 y_4 y_8 = z_1 z_2 z_5 z_6, \\ y_1 y_2 y_5 y_6 = z_1 z_2 z_4 z_8, \\ y_1 y_2 y_4 y_5 y_6 y_8 = z_1 z_2 z_4 z_5 z_6 z_8, \\ y_1 y_3 y_6 y_8 = z_1 z_3 z_6 z_8, \\ y_1 y_2 y_3 y_4 y_6 y_8 = z_1 z_2 z_3 z_5 z_6 z_8, \\ y_1 y_2 y_3 y_5 y_6 y_8 = z_1 z_2 z_3 z_4 z_6 z_8, \\ y_1 (y_4 y_8 (y_2 y_7 (1 - y_6) + y_3 y_5 y_6) \\ \quad + y_2 y_5 y_6 y_9 (1 - y_8)) = z_1 (z_4 z_8 (z_2 z_7 (1 - z_6) + z_3 z_5 z_6) \\ \quad + z_2 z_5 z_6 z_9 (1 - z_8)), \\ y_1 y_2 y_3 y_4 y_5 y_6 y_8 = z_1 z_2 z_3 z_4 z_5 z_6 z_8. \end{array} \right.$$

286 Solving this system of equations — see S10.m2 (output file S11.txt) and the expressions  
 287 simplified in S8.nb (text version S9.txt) on [https://github.com/jonathanmitchell88/CDMs/](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
 288 [tree/main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation) — we obtain

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

289 which has no solutions since  $z_i \in (0, 1)$  for all  $i \in \{1, 2, \dots, 8\}$  and the generating parameter  
 290 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,  
 291 any two CDMs with different leaf labelings are distinguishable.

## 292 CDM 4

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

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

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

## 297 CDM 3

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

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

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

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

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

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

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

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

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

## 307 CDM 2

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



309 The constraints for CDM 2 include those described above for CDM 3. Thus for CDM 2,  
 310 any two CDMs with different leaf labelings are distinguishable.

## 311 CDM 1

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

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

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

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

315 □

## 316 5A Proof of Theorem 6

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

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

319 Then

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

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

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

322 and

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

323 Finally, suppose  $x \neq y$  and  $x, y \neq o$ . Suppose now that  $v$  is the MRCA of  $x$  and  $y$ . Then  
 324 again by the same arguments as [Rhodes \(2019\)](#),

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

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

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

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

329 and

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

330

□

## 331 6A Inferring topologies of $N$ -taxon principal trees

332 We prove that consistent inference of the topology of the  $N$ -taxon principal tree fol-  
333 lows from consistent inference of the principal trees of the displayed 4-taxon CDMs.  
334 However, it is possible that a displayed 4-taxon CDM does not meet the assumptions  
335 of Section 3.2. Specifically, even if an  $N$ -taxon CDM meets the assumptions, some dis-  
336 played 4-taxon CDMs may have sister convergence. By assuming that all convergence  
337 parameters of the  $N$ -taxon CDM are sufficiently “small”, then all convergence param-  
338 eters of the displayed 4-taxon CDMs, including those of sister convergence groups of  
339 the displayed 4-taxon CDMs are “small”. Then all topologies of the displayed 4-taxon  
340 principal trees are inferred consistently by Algorithm 1.

341 To prove this result, we first prove a proposition similar to Proposition 1.2 of  
342 [Haughton \(1988\)](#), assuming the conditions of [Haughton \(1988\)](#) are satisfied. The  
343 proposition of [Haughton \(1988\)](#) states that if the generating model is among the set  
344 of candidate models, the probability that the model selected by the BIC is the gener-  
345 ating 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 spaces 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,  $\text{int } \Theta$  is the interior of some topological space  $\Theta$ ,  $\bar{m}_1$  and  $\bar{m}_2$  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  $\theta$ .

**Proposition 1A** *Let  $m_1$  and  $m_2$  be two different models satisfying  $m_1 \cap m_2 = \emptyset$ . Then there exists some  $\theta \in \text{int } \Theta$ ,  $\theta \notin \bar{m}_1, \theta \notin \bar{m}_2$ , with a neighborhood  $\mathfrak{N}$  of  $\theta$  such that  $\mathfrak{N} \cap m_1 = \emptyset$ ,  $\mathfrak{N} \cap m_2 \neq \emptyset$  and*

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

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

From Haughton (1988),

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

and asymptotically with probability 1,

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

where  $\epsilon > 0$ .

Furthermore, we can choose  $\tilde{\epsilon} > 0$  such that

$$\sup_{\phi \in m_2 \cap \Theta} \nabla b(\theta) \phi - b(\phi) + \tilde{\epsilon} = \nabla b(\theta) \theta - b(\theta). \quad (11A)$$

365 From Inequalities (9A) and (10A), asymptotically with probability 1,

$$\begin{aligned} \sup_{\phi \in m_1 \cap \Theta} (Y_n \phi - b(\phi)) &< \sup_{\phi \in m_1 \cap \Theta} \nabla b(\theta) \phi - b(\phi) + \frac{\epsilon}{4} \\ &\leq \nabla b(\theta) \theta - b(\theta) - \frac{3\epsilon}{4}. \end{aligned}$$

366 From Inequality (9A) and Equation (11A), asymptotically with probability 1,

$$\begin{aligned} \sup_{\phi \in m_2 \cap \Theta} (Y_n \phi - b(\phi)) &> \sup_{\phi \in m_2 \cap \Theta} \nabla b(\theta) \phi - b(\phi) - \frac{\epsilon}{4} \\ &= \nabla b(\theta) \theta - b(\theta) - \tilde{\epsilon} - \frac{\epsilon}{4}. \end{aligned}$$

367 If we choose  $\delta = \frac{\epsilon}{2} - \tilde{\epsilon} > 0$ , then asymptotically with probability 1,

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

368 There is always a choice of  $\tilde{\epsilon}$  such that  $\delta > 0$ . To see this, suppose  $\mathfrak{N}$  is a ball of radius  
369  $r > 0$ . Then  $r$  can be chosen to be arbitrarily small such that  $\tilde{\epsilon}$  is arbitrarily small. The proof  
370 then follows from Haughton (1988).

371 □

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

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

378 **Theorem 2A** (Steel, 1992) *For a set of rooted triples  $R$ ,  $\langle R \rangle = \{T\}$  if and only if  $R$  is*  
379 *consistent with  $T$ , and for each internal edge  $e$  of  $T$  there is a rooted triple in  $R$  which*  
380 *distinguishes  $e$ .*

381 The consequence of Theorem 2A of Steel (1992) is that if all trees of a set of  
 382 (binary) rooted 3-taxon trees  $R$  are displayed on a (binary) rooted  $N$ -taxon tree  $T$   
 383 and each internal edge of  $T$  is an internal edge of at least one tree in  $R$ , then  $T$  is  
 384 the only  $N$ -taxon tree that displays all the 3-taxon trees of  $R$ . In other words, the  
 385  $N$ -taxon tree  $T$  can be identified from the set of 3-taxon trees  $R$ .

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

### 391 6A.1 Proof of Theorem 7

392 Finally, from Proposition 1A and Theorem 2A adapted to unrooted quartets that  
 393 include the outgroup, we can prove Theorem 4.

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

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

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

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

397 Suppose  $m_1$  corresponds to the union of parameter spaces for CDMs 1–5 for the topology  
 398 of principal tree  $\mathcal{T}_4$ . Suppose also that  $m_2$  corresponds to the union of parameter spaces for  
 399 CDMs 1–5 for any 4-taxon topology of principal tree that is not  $\mathcal{T}_4$ . Then  $m_1 \cap m_2 = \emptyset$   
 400 — recall from Assumption 11 that the generating parameter is a generic point. Suppose  
 401  $\theta \notin \overline{m}_1, \overline{m}_2$ . Then if  $c > 0$  is sufficiently small there exists a neighborhood  $\mathfrak{N}$  of  $\theta$  such that  
 402  $\mathfrak{N} \cap m_1 \neq \emptyset$  and  $\mathfrak{N} \cap m_2 = \emptyset$  and Proposition 1A holds.

403 Next, we prove the claim that the set of inferred topologies of 4-taxon principal trees  
 404 equals the set of topologies of the principal trees of the 4-taxon CDMs displayed on  $\mathcal{N}$ . Then  
 405 from the adaptation of Theorem 2A to unrooted quartets, the topology of the principal tree of  
 406  $\mathcal{N}$  is the only topology that displays all inferred 4-taxon principal trees. Thus, any consistent  
 407 supertree inference method used in step 3 of Algorithm 1 infers the topology of the principal  
 408 tree of  $\mathcal{N}$  consistently and the proof is complete.

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

412 Suppose  $A_i$  is the event where the topology of the  $i^{th}$  4-taxon principal tree is inferred  
 413 incorrectly, given some arbitrary ordering. Then, by Proposition 1A, there exists some sample  
 414 size  $m$  such that for  $m' > m$ ,  $\mathbb{P}(A_i) < \epsilon_i$  for some arbitrarily small  $\epsilon_i > 0$ . Then by Boole's  
 415 inequality,

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

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

419 □

## 420 7A Proof of Proposition 8

421 *Proof* Consider phylogenetic tensors for  $\mathcal{N}_1$ ,  $\mathcal{N}_2$  and  $\mathcal{N}_2$  with arbitrary finite, positive sub-  
 422 stitution rates and epoch times. Let apostrophe superscripts denote parameters of  $\mathcal{N}_2$  and  
 423 double apostrophe superscripts denote parameters of  $\mathcal{N}_3$ .

424 For generic parameters on  $\mathcal{N}_1$ , after setting  $x_6 = x_7 = x_8 = x_9 = x_{10} = x_{11} = 1$ ,

425 Equations (7A) reduce to

$$\left\{ \begin{array}{l} f_{0011} = x_4 x_5, \\ f_{0101} = x_2 x_3 x_4, \\ f_{0110} = x_2 x_3 x_5, \\ f_{0111} = x_2 x_3 x_4 x_5, \\ f_{1001} = x_1 x_2 x_4, \\ f_{1010} = x_1 x_2 x_5, \\ f_{1011} = x_1 x_2 x_4 x_5, \\ f_{1100} = x_1 x_3, \\ f_{1101} = x_1 x_2 x_3 x_4, \\ f_{1110} = x_1 x_2 x_3 x_5, \\ f_{1111} = x_1 x_3 x_4 x_5, \\ f_{\delta} = x_1 x_2 x_3 x_4 x_5. \end{array} \right.$$

426 Taking the limit as the epoch time of the second epoch diverges and the epoch time of  
 427 the tip epoch converges to 0 is equivalent to  $x_1, x_2, x_3 \rightarrow 0$  and  $x_4, x_5 \rightarrow 1$ . Thus,

$$\left\{ \begin{array}{l} f_{0011} \rightarrow 1, \\ f_{0101} \rightarrow 0, \\ f_{0110} \rightarrow 0, \\ f_{0111} \rightarrow 0, \\ f_{1001} \rightarrow 0, \\ f_{1010} \rightarrow 0, \\ f_{1011} \rightarrow 0, \\ f_{1100} \rightarrow 0, \\ f_{1101} \rightarrow 0, \\ f_{1110} \rightarrow 0, \\ f_{1111} \rightarrow 0, \\ f_{\delta} \rightarrow 0. \end{array} \right.$$



428 For generic parameters on  $\mathcal{N}_2$ , after setting  $x'_7 = x'_8 = x'_9 = x'_{10} = x'_{11} = 1$  and permuting  
 429 leaf labels, Equations (7A) reduce to

$$\left\{ \begin{array}{l} f_{0011} = 1 - x'_6 (1 - x'_2 x'_3 x'_5), \\ f_{0101} = x'_4 x'_5 x'_6, \\ f_{0110} = x'_2 x'_3 x'_4 x'_6, \\ f_{0111} = x'_2 x'_3 x'_4 x'_5 x'_6, \\ f_{1001} = x'_1 x'_2 x'_5 x'_6, \\ f_{1010} = x'_1 x'_3 x'_6, \\ f_{1011} = x'_1 x'_2 x'_3 x'_5 x'_6, \\ f_{1100} = x'_1 x'_2 x'_4, \\ f_{1101} = x'_1 x'_2 x'_4 x'_5 x'_6, \\ f_{1110} = x'_1 x'_2 x'_3 x'_4 x'_6, \\ f_{1111} = x'_1 x'_4 (x'_2 (1 - x'_6) + x'_3 x'_5 x'_6), \\ f_\delta = x'_1 x'_2 x'_3 x'_4 x'_5 x'_6. \end{array} \right.$$

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

431 Thus,

$$\left\{ \begin{array}{l} f_{0011} \rightarrow 1, \\ f_{0101} \rightarrow 0, \\ f_{0110} \rightarrow 0, \\ f_{0111} \rightarrow 0, \\ f_{1001} \rightarrow 0, \\ f_{1010} \rightarrow 0, \\ f_{1011} \rightarrow 0, \\ f_{1100} \rightarrow 0, \\ f_{1101} \rightarrow 0, \\ f_{1110} \rightarrow 0, \\ f_{1111} \rightarrow 0, \\ f_{\delta} \rightarrow 0. \end{array} \right.$$

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

433

□

## 434 8A Controlling overfitting the CDM

435 For Algorithm 2 that infers the convergence groups of the  $N$ -taxon CDM, we introduce  
 436 two constraints to avoid overfitting the CDM with too many convergence groups. The  
 437 constraints are used to determine whether convergence groups are appended to the  
 438 CDM in a stepwise fashion.

439 The constraints ensure that only a small number of expected proportions of con-  
 440 verging quartets can exceed their observed values. Furthermore, convergence groups  
 441 can only be appended if expected proportions exceed their observed values by small  
 442 amounts. Suppose  $E$  and  $O$  are the matrices of expected and observed proportions

of converging quartets. Suppose  $C = \{c_1, c_2\}$  is an arbitrary convergence group, with  $c_1 \cup c_2 \subset [N]$ . Then  $a_1 = \max_{i \in c_1, j \in c_2} |E_{ij} - O_{ij}|$  and  $a_2 = \frac{1}{(N-1)^2} \sum_{i=1}^N \sum_{j=1}^N \delta_{ij}$ , where  $\delta_{ij} = 0$  if  $E_{ij} \leq O_{ij}$  and 1 otherwise. We append the convergence group to 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. ( $E_{iN} = E_{Nj} = O_{iN} = O_{Nj} = 0$  for all  $i, j \in \{1, 2, \dots, N\}$ .)

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, e.g. 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 of 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

*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

470 taxon  $c$  must be placed on an edge directly descended from  $v$ . Furthermore, for  $C$  to be a  
 471 non-sister convergence group, the speciation event corresponding to the node that  $c$  descends  
 472 from must be before the epoch  $C$  is in. Thus,  $c$  could be any of the  $|X_v \setminus X_C| = |X_v| - |X_C|$   
 473 taxa out of the  $N - 3$  possible taxa that are not  $o$ ,  $a$  or  $b$ .  $\square$

## 474 10A Proof of Proposition 11

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

479 In order to share at least one pair of converging taxa,  $C_2$  must be nested in  $C_1$ . How-  
 480 ever, by Assumption 9 of Section 3.2, there can be no convergence groups nested in other  
 481 convergence groups.  $\square$

## 483 11A Proof of Proposition 12

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

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

494 If  $\mathcal{N}$  is a tree, then the set of displayed 4-taxon CDMs after suppressing sister convergence  
 495 groups is a set of trees. Thus, 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 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 assume that every 4-taxon CDM in  $S$  is in  $S'$  and find a contradiction. 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  $c \in X_v \setminus X_C$ . The choice of  $c$  is not arbitrary and will be specified further later in the proof. 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. As in the choice of  $c$ ,  $C'$  is not chosen arbitrarily and will be specified later in the proof. 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  $C$ . 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

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

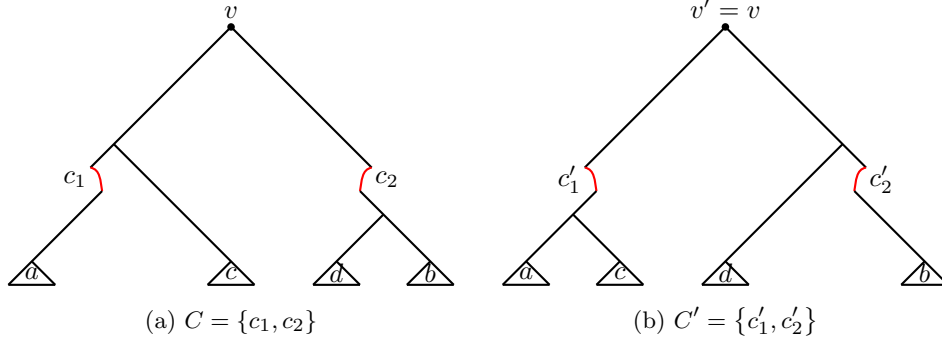
which simplifies to  $|X_C| = |X_{C'}|$ , since  $v' = v$ . Thus, either  $c_1 \subset c'_1$  and  $c_2 \supset c'_2$  or  $c_1 \supset c'_1$  and  $c_2 \subset c'_2$ . With no loss of generality, we assume that  $c_1 \subset c'_1$  and  $c_2 \supset c'_2$ .

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  $C'$ . Note that for  $|X_C| = |X_{C'}|$ , we must have the same number of leaf taxa on the subtree that  $c$  is on as the number of leaf taxa on the subtree that  $d$  is on. Then any other choice of convergence group that defines  $S'$ , say  $C''$ , must satisfy  $c'_1 \supset c_1$  and the claim follows. Finally, since we have assumed  $N \geq 5$ , we must also consider  $N = 4$ . For  $N = 4$ , it is clear from the identifiability and distinguishability of all CDMs with no sister convergence that the claim holds.

□

## 12A Proof of Theorem 13

*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, it is of no consequence whether the 4-taxon CDMs have sister convergence groups or not when it comes to inferring non-sister convergence groups.



**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 subtrees. Labels inside subtrees indicate one of possibly many taxa on leaves of those subtrees. There may be more subtrees not drawn below  $v$  that are not below either  $C$  or  $C'$ .

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

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

550 With no loss of generality, we assume the topology of the principal tree of an arbitrary 4-  
551 taxon CDM displayed on  $\mathcal{N}$  is  $(o(a, (b, c)))$ . Then of the possible convergence groups on the  
552 4-taxon CDM, the convergence group in the epoch closest to the root is the sister convergence  
553 group  $C = \{\{a\}, \{b, c\}\}$ . Thus, we first consider the distinguishability of two 4-taxon CDMs,  
554 one a tree and the other with this convergence group. For both CDMs we assume the tip  
555 epoch has epoch time 0. The first, which we call  $\mathcal{N}_{4,1}$ , is the tree  $(o(a, (b, c)))$ . Since the tip  
556 epoch has epoch time 0, taxa  $b$  and  $c$  are identical. The second CDM, which we call  $\mathcal{N}_{4,2}$ , has  
557 a single convergence group,  $C = \{\{a\}, \{b, c\}\}$ , followed by a speciation event involving  $b$  and  
558  $c$ . Again, since the tip epoch has epoch time 0, taxa  $b$  and  $c$  are identical. See Figure 2A for  
559 a graphical depiction of the two CDMs. Suppose  $\mathcal{N}_{4,1}$  has parameters with no apostrophes  
560 and  $\mathcal{N}_{4,2}$  has parameters with apostrophes.

561 For  $\mathcal{N}_{4,1}$  (see Mathematica file S12.nb (text version S13.txt) on [https://github.com/](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
562 [jonathanmitchell88/CDMs/tree/main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation) for a derivation),

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



563

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

$$\left\{ \begin{array}{l} r_{0011} = 1, \\ r_{0101} = 1 - x'_4 (1 - x'_2 x'_3), \\ r_{0110} = 1 - x'_4 (1 - x'_2 x'_3), \\ r_{1001} = x'_1 x'_2 x'_4, \\ r_{1010} = x'_1 x'_2 x'_4, \\ r_{1100} = x'_1 x'_3 x'_4, \\ r_{1111} = x'_1 x'_3 x'_4. \end{array} \right.$$

564

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

$$\left\{ \begin{array}{l} x_1 = \sqrt{\frac{r_{1001} r_{1100}}{r_{0101}}}, \\ x_2 = \sqrt{\frac{r_{0101} r_{1001}}{r_{1100}}}, \\ x_3 = \sqrt{\frac{r_{0101} r_{1100}}{r_{1001}}}, \\ r_{0101} = r_{0110}, \\ r_{1001} = r_{1010}, \\ r_{1100} = r_{1111}. \end{array} \right.$$

565

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

$$\left\{ \begin{array}{l} r_{0101} = r_{0110}, \\ r_{1001} = r_{1010}, \\ r_{1100} = r_{1111}, \\ r_{0101} r_{1001} < r_{1100}, \\ r_{0101} r_{1100} < r_{1001}, \\ r_{1001} r_{1100} < r_{0101}. \end{array} \right.$$

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

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

567 Since we are assuming that all convergence parameters of  $\mathcal{N}$  are less than some constant  
568  $c > 0$ , we can assume that  $x'_4 = 1 - \epsilon$ , where  $\epsilon > 0$  is some small positive constant. Then for  
569  $\mathcal{N}_{4,2}$ ,

$$\begin{cases} r_{1100} - r_{0101}r_{1001} = x'_1x'_3(1 - x_2'^2) + O(\epsilon), \\ r_{1001} - r_{0101}r_{1100} = x'_1x'_2(1 - x_3'^2) + O(\epsilon), \\ r_{0101} - r_{1001}r_{1100} = x'_2x'_3(1 - x_1'^2) + O(\epsilon). \end{cases}$$

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

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

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

576 The next closest epoch to the root that could have a convergence group is the epoch  
577 just after taxa  $b$  and  $c$  have diverged. Thus, we compare the tree  $(o, (a, (b, c)))$ , which we  
578 call  $\mathcal{N}_{4,3}$ , and the CDM with topology of principal tree  $(o, (a, (b, c)))$  and sister convergence  
579 group  $\{\{b\}, \{c\}\}$  in the tip epoch, which we call  $\mathcal{N}_{4,4}$ . See Figure for a graphical depiction

580 of the two CDMs. Again, suppose  $\mathcal{N}_{4,3}$  has parameters with no apostrophes and  $\mathcal{N}_{4,4}$  has  
 581 parameters with apostrophes.

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

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

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

$$\left\{ \begin{array}{lcl} r_{0011} & = & 1 - x'_6 (1 - x'_4 x'_5), \\ r_{0101} & = & x'_2 x'_3 x'_4 x'_6, \\ r_{0110} & = & x'_2 x'_3 x'_5 x'_6, \\ r_{1001} & = & x'_1 x'_2 x'_4 x'_6, \\ r_{1010} & = & x'_1 x'_2 x'_5 x'_6, \\ r_{1100} & = & x'_1 x'_3, \\ r_{1111} & = & x'_1 x'_3 (1 - x'_6 (x'_4 x'_5)). \end{array} \right.$$

584

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

$$\left\{ \begin{array}{l} x_1 = \sqrt{\frac{r_{1001}r_{1100}}{r_{0101}}}, \\ x_2 = \sqrt{\frac{r_{0110}r_{1001}}{r_{0011}r_{1100}}}, \\ x_3 = \sqrt{\frac{r_{0101}r_{1100}}{r_{1001}}}, \\ x_4 = \sqrt{\frac{r_{0011}r_{0101}}{r_{0110}}}, \\ x_5 = \sqrt{\frac{r_{0011}r_{0110}}{r_{0101}}}, \\ r_{0101}r_{1010} = r_{0110}r_{1001}, \\ r_{0011}r_{1100} = r_{1111}. \end{array} \right.$$

585

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

$$\left\{ \begin{array}{l} r_{0101}r_{1010} = r_{0110}r_{1001}, \\ r_{0011}r_{1100} = r_{1111}, \\ r_{0011}r_{0101} < r_{0110}, \\ r_{0011}r_{0110} < r_{0101}, \\ r_{0101}r_{1100} < r_{1001}, \\ r_{0110}r_{1001} < r_{0011}r_{1100}, \\ r_{1001}r_{1100} < r_{0101}. \end{array} \right.$$

586

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

$$\left\{ \begin{array}{l} r_{0101}r_{1010} = r_{0110}r_{1001}, \\ r_{0011}r_{1100} = r_{1111}. \end{array} \right.$$

587

Since we are assuming that all convergence parameters of  $\mathcal{N}$  are less than some constant

588

 $c > 0$ , we can assume that  $x'_6 = 1 - \epsilon$ , where  $\epsilon > 0$  is some small positive constant. Then for

589  $\mathcal{N}_{4,4}$ ,

$$\left\{ \begin{array}{l} r_{0110} - r_{0011}r_{0101} = x'_2x'_3x'_5(1 - x_4'^2) + O(\epsilon), \\ r_{0101} - r_{0011}r_{0110} = x'_2x'_3x'_4(1 - x_5'^2) + O(\epsilon), \\ r_{1001} - r_{0101}r_{1100} = x'_1x'_2x'_4(1 - x_3'^2) + O(\epsilon), \\ r_{0011}r_{1100} - r_{0110}r_{1001} = x'_1x'_3x'_4x'_5(1 - x_2'^2) + O(\epsilon), \\ r_{0101} - r_{1001}r_{1100} = x'_2x'_3x'_4(1 - x_1'^2) + O(\epsilon). \end{array} \right.$$

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

$$\left\{ \begin{array}{l} r_{0101}r_{1010} = r_{0110}r_{1001}, \\ r_{0011}r_{1100} = r_{1111}, \\ r_{0011}r_{0101} < r_{0110}, \\ r_{0011}r_{0110} < r_{0101}, \\ r_{0101}r_{1100} < r_{1001}, \\ r_{0110}r_{1001} < r_{0011}r_{1100}, \\ r_{1001}r_{1100} < r_{0101}. \end{array} \right.$$

592 Thus,  $\mathcal{N}_{4,3}$  and  $\mathcal{N}_{4,4}$  are not distinguishable. Thus, any 4-taxon CDM with  $\alpha_l = \beta_l$   
 593 and this sister convergence group is not distinguishable from the CDM that results from  
 594 suppressing the sister convergence. Thus, to determine the transformed phylogenetic tensor  
 595 of any 4-taxon CDM with  $\alpha_l = \beta_l$ , we can assume there is no sister convergence in this epoch.

596 By the assumption that no leaf taxa belong to more than one convergence group, there  
 597 can be no more than one convergence group on any arbitrary 4-taxon CDM displayed on  
 598  $\mathcal{N}$ . Thus, taking into consideration  $\mathcal{N}_{4,1}$  and  $\mathcal{N}_{4,2}$  not being distinguishable and  $\mathcal{N}_{4,3}$  and  
 599  $\mathcal{N}_{4,4}$  not being distinguishable, we can conclude that any arbitrary 4-taxon CDM displayed  
 600 on  $\mathcal{N}$  is not distinguishable from the 4-taxon CDM that results from suppressing any sister  
 601 convergence group, which is one of CDM 1 – 3 of Figure 3.

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

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

604 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,

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

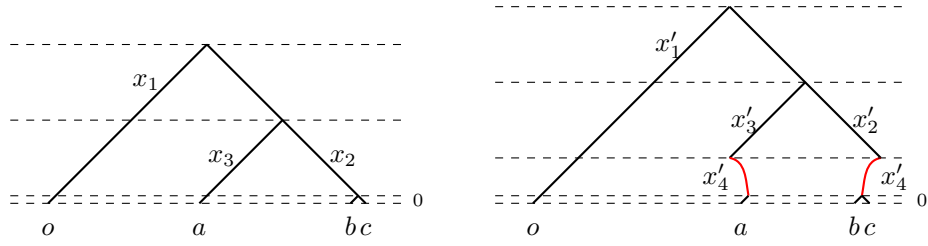
605 In S14.m2 (output file S15.txt) on [https://github.com/jonathanmitchell88/CDMs/tree/](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation)  
 606 [main/SupplementaryInformation](https://github.com/jonathanmitchell88/CDMs/tree/main/SupplementaryInformation), we see that the set of parameters  $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$   
 607 is identifiable. It follows that CDMs 1 and 2 are also identifiable.

608 Thus, using similar arguments to those of the proof of Theorem 3, with probability  
 609 converging to 1, step 2 of Algorithm 2 infers all the 4-taxon CDMs with the outgroup that  
 610 are displayed on  $\mathcal{N}$  after suppressing sister convergence groups.

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

614 with that pair of taxa as non-sisters all have the pair converging. Thus, only convergence  
615 groups on  $\mathcal{N}$  can be on the inferred  $N$ -taxon CDM. If not all convergence groups of  $\mathcal{N}$  have  
616 been appended to the inferred CDM, then there are some elements of  $O$  that are non-zero  
617 corresponding to elements of  $E$  that are zero, corresponding to the pairs of converging taxa in  
618 convergence groups of  $\mathcal{N}$  that are not yet on the inferred CDM. Including these convergence  
619 groups on the inferred CDM makes these elements of  $E$  equal to the corresponding elements  
620 of  $O$ , decreasing the sum of squared differences. Once all convergence groups of  $\mathcal{N}$  have been  
621 appended to the inferred CDM,  $O = E$ . Thus, no more convergence groups can be appended  
622 to the inferred CDM to decrease the sum of squared differences and the algorithm terminates.

□

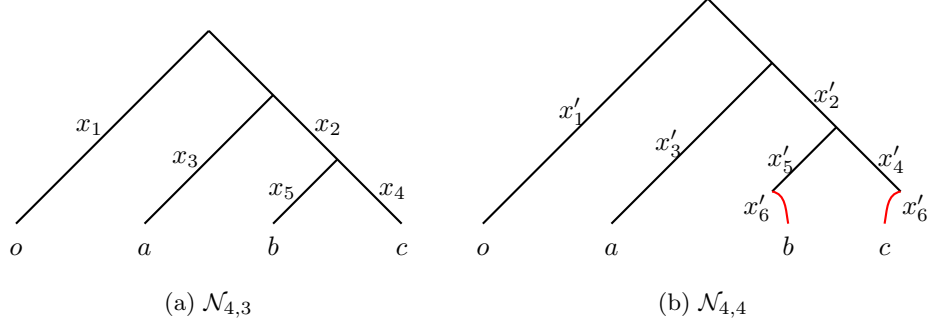


(a)  $\mathcal{N}_{4,1}$  —  $x_1$  represents the entire outgroup (b)  $\mathcal{N}_{4,2}$  —  $x'_1$  represents the entire outgroup edge when unrooted.

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

## 624 13A Inferring convergence group orders on $N$ -taxon 625 CDMs

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



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

630 from the inferred 4-taxon CDMs. For example, CDM 2 and CDM 3 differ by CDM 2  
 631 having a convergence group in the tip epoch and CDM 3 not having one.

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

638 For convergence groups  $C_i$  and  $C_j$ , whose partial order has not been determined,  
 639 we tally proportions of 4-taxon CDMs best supported by the two orders to obtain a  
 640 matrix of “observed” convergence group order ratios. Convergence group orders are  
 641 resolved one at a time by minimizing the sum of squared differences between matrices  
 642 of observed and “expected” partial convergence group order ratios. The matrix of  
 643 inferred convergence group order ratios is updated after each convergence group order  
 644 is inferred. We discard 4-taxon CDMs with convergence group orders not consistent  
 645 with the matrix of inferred convergence group orders.

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



648 then  $C_i$  cannot be in the tip epoch. For other convergence groups, whether they are  
649 in the tip epoch or not must be inferred.

650 For each 4-taxon CDM with possible a possible convergence group in the tip epoch,  
651 we determine which CDM is best supported among the two CDMs, e.g. CDM 2 versus  
652 CDM 3 or CDM 4 versus CDM 5, with a fixed leaf labelling. For a given convergence  
653 group, we tally the 4-taxon CDMs displaying the given convergence group with and  
654 without the convergence group in the tip epoch. If more than some cutoff, e.g. half,  
655 have the convergence group in the tip epoch, then we infer that the convergence group  
656 is in the tip epoch. For each 4-taxon set that includes the outgroup, we retain only one  
657 possible CDM after the convergence group ordering and determining whether there is  
658 a convergence group in the tip epoch or not.

659 Note that some convergence group orders may still be undefined. Suppose two  
660 convergence groups do not have an order defined by the edge partial order of the  
661 principal tree or the orders of other convergence groups. Suppose both convergence  
662 groups are only ever present on 4-taxon CDMs where one convergence group is a sister  
663 convergence group. Then there will be no information to resolve the order of these two  
664 convergence groups. We leave these convergence group orders unresolved. Thus, we  
665 have a *partial* order on the convergence groups. Algorithms 1A and 2A for inferring  
666 convergence group orders and whether or not there is a convergence group in the tip  
667 epoch then follow.

668 We do not prove consistency of inference of the convergence group partial orders  
669 from Algorithm 1A. This is because Theorem 13 assumes that no leaf taxa belong to  
670 more than one convergence group. Thus, all 4-taxon CDMs displayed on  $\mathcal{N}$  have at  
671 most one non-sister convergence group and there are no convergence group orders to  
672 infer. Furthermore, we do not prove consistency of inference of the convergence groups  
673 in the tip epoch.

---

**Algorithm 1A** Convergence group order inference

---

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

1. Initialize empty list of inferred 4-taxon CDMs  $L_Q$ . Initialize  $k \times k$  matrix of observed convergence group orders  $O$  as zero matrix, where  $k$  is length of list  $\hat{\mathcal{G}}$ . Initialize  $k \times k$  matrix  $E$  of expected convergence group orders as convergence group orders defined by  $P$ , with  $E_{ij} = 1$  if convergence group  $i$  is 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  $\hat{\mathcal{N}}$  and permitted by  $E$  and append to  $L_Q$ .
3. For all  $i, j$ , compute  $O_{ij}$  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  $O$  and  $E$ ,  $s = \sum_{i=1}^k \sum_{j=1}^k (O_{ij} - E_{ij})^2$ .
5. Assign new order between two convergence groups that minimizes sum of squared differences.
6. Update  $E$  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 there are no pairs of convergence groups left to assign orders to, terminate the algorithm.
7. Return to Step 5.

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

---

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

---

**Algorithm 2A** Inference of convergence groups in tip epochs

---

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

1. Initialize empty list of inferred 4-taxon CDMs  $L_Q$ . Initialize vector  $D$  of length  $k$  of convergence groups in tip epoch as zero vector, where  $k$  is length of list  $\hat{\mathcal{G}}$ .
2. For each 4-taxon set that includes outgroup  $o$ , select CDM from those displayed on  $\hat{\mathcal{N}}$  and permitted by  $E$  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  $\hat{\mathcal{N}}$ , compute  $D_i$  as proportion of inferred 4-taxon CDMs with  $C_i$  in the tip epoch.
4. If  $\max_{i \in \{1, 2, \dots, k\}} D_i = D_j$  and  $D_j > \tau$ , set  $D_j = 1$ .

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

---

## 14A Proof of Proposition 14

*Proof* Using the parameterization of Section 3A.1, for CDM 5, with principal tree  $(o, (a, (b, c)))$ , the sums of edge lengths between leaf taxa are

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

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

edges of the principal tree,

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

where  $l_o$  is the sum of divergence parameters along the two edges of the principal tree whose parent node is the root,  $l_a$ ,  $l_b$  and  $l_c$  are the sums of divergence and possibly convergence parameters along the terminal edges whose descendent leaf taxa are  $a$ ,  $b$  and  $c$  respectively and  $l_{bc}$  is the sum of divergence parameters along the edge whose descendent leaf taxa are  $b$  and  $c$ .

It follows that all edge lengths are also identifiable for CDMs 1 – 4 since expressions for the sums of edge lengths are the same, except that some  $y_i = 1$ .

□

## 15A Proof of Proposition 15

*Proof* On CDM 5, parameters  $y_6 = x_6$  and  $y_8 = x_9$  are identifiable. Thus, the convergence parameters  $l_6 = a_6 + b_6 = -\log(y_6)$  and  $l_9 = a_9 + b_9 = -\log(y_8)$  are identifiable. Thus, for all other CDMs with these convergence parameters, they are also identifiable.

□

## 16A Proof of Proposition 16

*Proof* From Equation (6A) for the phylogenetic tensor of CDM 5,  $q_{0001} = q_{0010} = q_{0100} = q_{1000} = \gamma$ . Thus,  $\gamma$  is identifiable for CDM 5. Since all other CDMs are nested in CDM 5 and none correspond to generic values of  $\gamma$  — instead they correspond to some generic values of  $x_i$  or  $y_i$  —  $\gamma$  is also identifiable for CDMs 1 – 4.

□

## 17A Proof of Theorem 17

*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  $\gamma$  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  $\gamma$  also converges in probability to  $\gamma$ . 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  $X$  has rank  $2N - 3$ ,  $X^T X$  is invertible. It follows that  $\hat{l}$  also converges in probability to  $l$  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  $\hat{v}$  is estimated at least once. Thus,  $\hat{v}$  converges in probability to  $v$ . Finally, since  $\gamma$  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,  $\hat{\gamma}$  converges in probability to  $\gamma$ .

□

## References

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

- 731 Haughton DM (1988) On the choice of a model to fit data from an exponential family.  
732 The Annals of Statistics pp 342–355. <https://doi.org/10.1214/aos/1176350709>
- 733 Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian  
734 Journal of Statistics pp 65–70
- 735 Rhodes JA (2019) Topological metrizations of trees, and new quartet methods of tree  
736 inference. IEEE/ACM Transactions on Computational Biology and Bioinformatics  
737 17(6):2107–2118. <https://doi.org/10.1109/tcbb.2019.2917204>
- 738 Steel M (1992) The complexity of reconstructing trees from qualitative characters and  
739 subtrees. Journal of classification 9:91–116. <https://doi.org/10.1007/bf02618470>
- 740 Sumner JG, Fernández-Sánchez J, Jarvis PD (2012a) Lie markov models. Journal of  
741 Theoretical Biology 298:16–31. <https://doi.org/10.1016/j.jtbi.2011.12.017>
- 742 Sumner JG, Holland B, Jarvis P (2012b) The algebra of the general markov model on  
743 phylogenetic trees and networks. Bulletin of Mathematical Biology 74(4):858–880.  
744 <https://doi.org/10.1007/s11538-011-9691-z>