Version dated: July 7, 2014

SUPPLEMENTARY MATERIAL TO "How much information is needed to infer reticulate evolutionary histories?"

Katharina T. Huber[1], Leo van Iersel[2], Vincent Moulton[1] and Taoyang Wu[1]

[1] *School of Computing Sciences, University of East Anglia, Norwich, United Kingdom*

[2] *CWI, Amsterdam, Netherlands*

Corresponding author: Vincent Moulton, School of Computing Sciences, University of East Anglia, Norwich, United Kingdom; E-mail: vincent.moulton@cmp.uea.ac.uk.

Tree-equivalence theorems

This section uses the terminology and notation in "A fundamental obstruction to constructing phylogenetic networks" by K. T. Huber, L. van Iersel, V. Moulton, and T. Wu. We shall prove that the networks $\mathcal{N}_1$ and $\mathcal{N}_2$ both display the same set of phylogenetic trees (see Theorem 2 below). We also prove that this is the case for networks $\mathcal{H}_1$ and $\mathcal{H}_2$ (see Theorem 6 below).

We assume again that X is a finite non-empty set, that n is a positive integer and that $i \in \{1, 2\}$. We begin by presenting some definitions. Following van Iersel et al. (2010), we say that a phylogenetic tree $\mathcal{T}$ on X is *displayed* by a phylogenetic network $\mathcal{N}$ on X if there exists a subdigraph $\mathcal{T}'$ of $\mathcal{N}$ so that $\mathcal{T}$ can be obtained from $\mathcal{T}'$ by suppressing all degenerate vertices. The set of all phylogenetic trees on X displayed by a phylogenetic network $\mathcal{N}$ on X is denoted by $\mathscr{T}(\mathcal{N})$. We say that two phylogenetic networks $\mathcal{N}$ and $\mathcal{N}'$

on X are *tree-equivalent* if for every tree $\mathscr{T} \in \mathscr{T}(\mathcal{N})$ there exists a tree $\mathscr{T}' \in \mathscr{T}(\mathcal{N}')$ such that $\mathscr{T}$ and $\mathscr{T}'$ are equivalent and vice versa. Note that some examples of tree-equivalent networks that are not equivalent are presented by Willson (2011).

To show that $\mathcal{N}_1$ and $\mathcal{N}_2$ are tree-equivalent we shall use the following result concerning binary sequences. We say that a subset A of $\{1, 2, \dots, n\}$ is *dominated* by a sequence $w \in \mathcal{B}_n$ if $A \subseteq supp(w)$ holds. More generally, if $A_1, \dots A_k$ are subsets of $\{1, 2, \dots, n\}$ and there exist sequences $w_j \in \mathcal{B}_n$, $1 \leq j \leq k$, such that, for all $1 \leq j \leq k$, A_j is dominated by w_j then we say that the list $(A_1, \dots, A_k)$ is *dominated* by the list $(w_1, \dots, w_k)$ or that $(w_1, \dots, w_k)$ *dominates* $(A_1, \dots, A_k)$. Now, suppose X has size at least two. For convenience, we denote any partition $\{A_1, \dots, A_k\}$, $k \geq 2$, of X into k pairwise disjoint sets by $A_1|A_2|\dots|A_k$, where the order in which we list the sets A_j, $2 \leq j \leq k$, is not important.

Lemma 1. *Suppose that $n \geq 3$ and that $A_1|\dots|A_k$ is a partition of $\{1, \dots, n\}$, $k \geq 2$. Then, there exist distinct sequences $w_1, \dots, w_k$ in $\mathcal{B}_n^i$ such that $(A_1, \dots, A_k)$ is dominated by $(w_1, \dots, w_k)$.*

Proof. Since the proof for $i = 2$ is similar, we restrict our attention to showing that the lemma holds for $i = 1$. We proceed by induction on n. Suppose first that $n = 3$. If $k = 2$ then, by symmetry, we may assume that the partition of $Y = \{1, 2, 3\}$ under consideration is $\{1, 2\}|\{3\}$. Clearly, the list $(\{1, 2\}, \{3\})$ is dominated by $(111, 001)$, all of whose sequences are distinct. If $k = 3$, then the sole partition of Y is $\{1\}|\{2\}|\{3\}$ and $(\{1\}, \{2\}, \{3\})$ is dominated by $(100, 010, 001)$, all of whose sequences are again distinct.

Now assume that $n > 3$, and that the result holds for $n - 1$. Fix an arbitrary partition $\pi = A_1|\dots|A_k$ of $\{1, \dots, n\}$ where $k \geq 2$. By relabeling the elements of π if necessary, we may assume that $n \in A_k$. Consider the map $\kappa : \mathcal{B}_{n-1}^1 \to \mathcal{B}_n^1$ that maps each sequence w in $\mathcal{B}_{n-1}^1$ to the sequence in $\mathcal{B}_n^1$ obtained from w by appending "0" to the right, that is, $[\kappa(w)]_j = [w]_j$ for $1 \leq j \leq n-1$ and $[\kappa(w)]_n = 0$. Note that if a set A (as a subset of $\{1, \ldots, n-1\}$) is dominated by a sequence w in $\mathcal{B}^1_{n-1}$, then A (as a subset of $\{1, \ldots, n\}$) is also dominated by $\kappa(w)$.

The remainder of the proof consists of two cases that depend on the cardinality of A_k:

Case (i): Suppose $|A_k| = 1$. Then $A_k = \{n\}$. Assume first that $k = 2$. Then $\pi = \{1, \ldots, n-1\}|\{n\}$. Let $w_1 \in \mathcal{B}^1_n$ with $\{1, \ldots, n-1\} \subseteq supp(w_1)$ holding. Then $(\{1, \ldots, n-1\}, \{n\})$ is dominated by $(w_1, w_{n,n})$.

Now assume that $k \geq 3$. Then $\pi = A_1|\ldots|A_{k-1}|\{n\}$. Since $= A_1|\ldots|A_{k-1}$ is a partition of $\{1, \ldots, n-1\}$ the induction assumption implies that there exist distinct sequences in $w'_j \in \mathcal{B}^1_{n-1}$, $1 \leq j \leq k-1$, such that $(w'_1, \ldots, w'_{k-1})$ dominates $(A_1, \ldots, A_{k-1})$. Since $\kappa(w'_j) \neq \kappa(w'_l)$ for all $j, l \in \{1, \ldots, k-1\}$ distinct and $\kappa(w'_j) \neq w_{n,n}$ for all $j \in \{1, \ldots, k-1\}$ it follows that $(\kappa(w'_1), \ldots, \kappa(w'_{k-1}), w_{n,n})$ dominates $(A_1, \ldots, A_{k-1}, \{n\})$.

Case (ii): Suppose $|A_k| \geq 2$. Then $\pi' = A_1|\ldots|A_{k-1}|A_k - \{n\}$ is a partition of $\{1, \ldots, n-1\}$. By induction assumption, there exist distinct sequences $w'_j \in \mathcal{B}^1_{n-1}$, $1 \leq j \leq k$, such that $(A_1, \ldots, A_k - \{n\})$ is dominated by $(w'_1, \ldots, w'_k)$. Since $|A_k| < n$ there exists a sequence $w \in \mathcal{B}^1_n$ that dominates A_k. In particular, $n \in A_k$ implies $[w]_n = 1$. Hence, for all $1 \leq i \leq k$, the sequences $\kappa(w'_i)$ and w are distinct. Thus, $(\kappa(w'_1), \ldots, \kappa(w'_{k-1}), w)$ dominates $(A_1, \ldots, A_k)$. $\square$

Theorem 2. *The networks $\mathcal{N}_1$ and $\mathcal{N}_2$ are tree-equivalent.*

Proof. We shall show that for any tree $\mathscr{T}_1 \in \mathscr{T}(\mathcal{N}_1)$ there exists a tree $\mathscr{T}_2 \in \mathscr{T}(\mathcal{N}_2)$ such that $\mathscr{T}_1$ and $\mathscr{T}_2$ are equivalent. The claimed equivalence then follows by symmetry.

Let $\mathscr{T}_1 \in \mathscr{T}(\mathcal{N}_1)$. Note that if $\mathscr{T}_1$ is the star tree on X then $\mathscr{T}_1$ is clearly displayed by $\mathcal{N}_2$. Thus, we may assume that $\mathscr{T}_1$ is not the star tree on X. Let $T'_1 = (V'_1, E'_1)$ be the tree with leaf set X obtained from $\mathcal{D}_1$ by deleting a set of arcs and removing the resulting isolated vertices such that T_1' is a subdivision of $\mathscr{T}_1$. Put $\{w_1^1, \dots, w_l^1\} = V_1' \cap \mathcal{B}_n^1$ and note that $l \geq 2$ as $\mathscr{T}_1$ is not the star tree on X. For all $1 \leq j \leq l$, let $A_j \subseteq X$ comprise of all leaves $x \in X$ for which the path in T_1' from the root $\rho(T_1')$ of T_1' to x contains w_j^1. Clearly, $A_1|\dots|A_l$ is a partition of X. By Lemma 1 it follows that there exist l distinct sequences $w_1^2, \dots, w_l^2$ in $\mathcal{B}_n^2$ such that, for all $1 \leq j \leq l$, A_j is dominated by w_j^2. Let $T_2' = (V_2', E_2')$ be the tree with leaf set X, vertex set $V_2' = \{\rho_2, w_1^2, \dots, w_l^2\} \cup X \cup Y$ where ρ_2 is the root of $\mathcal{D}_2$ and arc set E_2' comprising (i) for all $1 \leq j \leq l$ of the arc (ρ_2, w_j^2), (ii) for all $1 \leq j \leq n$ of the arcs (y_j, x_j), and (iii) for all $1 \leq t \leq n$ and all $1 \leq j \leq n$ of the arcs (w_j^2, y_t) precisely if $x_t \in A_j$. Since $A_j \subseteq supp(w_j^2)$ holds for $1 \leq j \leq l$, it is easy to see that T_2' can be obtained from $\mathcal{D}_2$ by deleting a set of arcs and removing all resulting isolated vertices.

Now, let $\kappa : V_1' \to V_2'$ denote the map that maps, for all $1 \leq j \leq l$, w_j^1 to w_j^2 and is the identity on $V_1' - \{w_1^1, \dots, w_l^1\}$. Since for all $1 \leq j \leq l$ and all $1 \leq t \leq n$, we have that $(w_j^1, y_t) \in E_1'$ if and only if $x_t \in A_j$ if and only if $(w_j^2, y_t) \in E_2'$ it follows that κ induces an isomorphism between the trees T_1' and T_2' that is the identity on X. Consequently the phylogenetic tree on X obtained from T_2' by suppressing all degenerate vertices is contained in $\mathscr{T}(\mathcal{N}_2)$ and equivalent with $\mathscr{T}_1$. $\square$

To prove that $\mathcal{H}_1$ and $\mathcal{H}_2$ are tree-equivalent, we require some further definitions and results. Ideally, we would have liked to deduce this fact from the previous theorem, but we have not been able to find such a proof. Suppose that G is a rooted DAG and v is a vertex of G. We denote the set of leaves of G for which v is an ancestor by $\mathcal{L}_G(v)$ and simply write $\mathcal{L}(v)$ if the graph we are referring to is clear from the context. In addition, if u is a vertex of G such that u is an ancestor of v then we write $v \preceq_G u$ (we also write $v \prec_G u$ if $v \neq u$) where we will omit the subscript if the rooted DAG we are referring to is clear form the context. Note that $\preceq_G$ is a partial ordering on $V(G)$.

Lemma 3. *If $\mathcal{A} \subseteq \mathcal{A}_n$ is a complete set such that the tree $\mathcal{P}[\mathcal{A}]$ is binary and*

$L(\mathcal{P}[\mathcal{A}]) \subseteq \mathcal{B}_n^i$ then $\mathcal{P}[\mathcal{A}]$ is displayed by $\mathcal{P}_n|_{\mathcal{B}_n^i}$.

Proof. Since the proof for $i = 2$ is similar, we restrict our attention to showing that the lemma holds for $i = 1$. Let $\mathcal{A} \subseteq \mathcal{A}_n$ denote a complete set such that the tree $\mathcal{P}[\mathcal{A}]$ is binary and $L(\mathcal{P}[\mathcal{A}]) \subseteq \mathcal{B}_n^i$. We perform induction on $m = |L(\mathcal{P}[\mathcal{A}])|$. That the lemma holds in case $m \in \{1, 2\}$ is straightforward to verify. So assume that $m > 2$ and that the lemma holds for all subsets $\mathcal{A}' \subseteq \mathcal{A}_n$ with $|L(\mathcal{P}[\mathcal{A}'])| < m$ that satisfy the assumptions of the lemma.

Let $w_1, w_2 \in \mathcal{A}$ denote two distinct sequences that share a common parent w in $\mathcal{P}[\mathcal{A}]$, let w' denote the parent of w in $\mathcal{P}[\mathcal{A}]$, and let $w_3 \in \mathcal{A}$ be a sequence contained in $\mathcal{L}(w')$ but not in $\mathcal{L}(w)$. Then w is not a prefix of w_3. Put $\mathcal{A}' = \mathcal{A} - \{w_1, w\}$. Then $\mathcal{A}'$ is a complete subset of $\mathcal{A}_n$, the tree $\mathcal{P}[\mathcal{A}']$ is binary, $L(\mathcal{P}[\mathcal{A}']) \subseteq \mathcal{B}_n^1$, and $|L(\mathcal{P}[\mathcal{A}'])| = m - 1$. By induction hypothesis, it follows that $\mathcal{P}[\mathcal{A}']$ is displayed by $\mathcal{P}_n|_{\mathcal{B}_n^1}$. Let $\widetilde{w} \in \mathcal{A}_n$ denote the common prefix of w_1 and w_2 of maximal length. Then w is a prefix of $\widetilde{w}$. Moreover, since w is not a prefix of w_3, $\widetilde{w}$ is not a prefix of w_3.

Since $\mathcal{P}[\mathcal{A}']$ is displayed by $\mathcal{P}_n|_{\mathcal{B}_n^1}$, there exists a tree T' with leaf set $L(\mathcal{P}[\mathcal{A}'])$ obtained from $\mathcal{P}_n|_{\mathcal{B}_n^1}$ via a series of vertex and arc deletions so that when also suppressing all degenerate vertices in T', we obtain a phylogenetic tree that is equivalent to $\mathcal{P}[\mathcal{A}']$. Let u be the lowest non-degenerate vertex in T' such that $w_1 \in \mathcal{L}(u)$, that is, $u \preceq v$ holds for all non-degenerate interior vertices v in T' with $w_1 \in \mathcal{L}(v)$. Note that, by construction, u is a common prefix of w_1 and w_3. Consequently, u is a degenerate vertex in T' with $w_1 \prec \widetilde{w} \prec u$. Let T be the tree obtained from T' by adding the path in $\mathcal{P}_n|_{\mathcal{B}_n^1}$ from $\widetilde{w}$ to w_2. Note that T can be obtained from $\mathcal{P}_n|_{\mathcal{B}_n^1}$ via a series of vertex and arc deletions. Moreover, when suppressing all degenerate vertices in T, we obtain a phylogenetic tree that is equivalent to $\mathcal{P}[\mathcal{A}]$. Therefore, $\mathcal{P}[\mathcal{A}]$ is also displayed by $\mathcal{P}_n|_{\mathcal{B}_n^1}$. This completes the induction step and hence the proof of the lemma. □

We now present some results which will help us ensure that the tree used to replace the "top layer" vertices in obtaining $\mathcal{D}_{1,2}$ from $\mathcal{D}_{1,1}$ is mapped correctly into the tree used to replace the same layer of vertices in obtaining $\mathcal{D}_{2,2}$ from $\mathcal{D}_{2,1}$.

Let $\mathcal{C}^n$ denote the phylogenetic tree on X obtained from the rooted caterpillar $\mathcal{C}_{\mathbf{1_n}}$ by relabeling each of its leaves j by the corresponding element $x_j \in X$. In addition, let $\mathcal{T}_1 \in \mathscr{T}(\mathcal{H}_1)$ be such that $\mathcal{T}_1$ and $\mathcal{C}^n$ are not equivalent and let T_1 denote the tree obtained from $\mathcal{D}_{1,3}$ by deleting a set of arcs and removing all resulting isolated vertices such that T_1 is a subdivision of $\mathcal{T}_1$. Put

$$W = \{w_1^1, \dots, w_m^1\} := V(T_1) \cap \mathcal{B}_n^1.$$

Then $m \geq 2$ must hold. Indeed, assume for contradiction that $m = 1$. Let w denote the unique vertex in W. Then $supp(w) = \{1, \dots, n\}$ as $\mathcal{T}_1$ has leaf set X. Let $\mathcal{Z}_w$ denote the rooted tree associated to w that was obtained from the tree $\mathcal{C}_w$ in the construction of the graph $\mathcal{D}_{2,3}$. Attaching to each leaf l in $\mathcal{Z}_w$ the directed path from l to the unique element $x_l \in X$ below l results in a phylogenetic tree $\mathcal{Z}'_w$ on X that is equivalent with $\mathcal{T}_1$. Since $X = L(\mathcal{C}^n)$, the construction of $\mathcal{Z}_w$ implies that $\mathcal{Z}'_w$ and $\mathcal{C}^n$ are also equivalent. Thus, $\mathcal{T}_1$ and $\mathcal{C}^n$ must be equivalent which is impossible. Consequently, $m \geq 2$ must hold as required.

Let T'_1 denote the tree obtained by suppressing all degenerate vertices in the restriction $T_1|_{V(T_1)\cap\mathcal{A}_n}$ Then $L(T'_1) = W$ and T'_1 is equivalent with $\mathcal{P}_n|_W$. Now, for each vertex v in T'_1, let

$$\mathbb{I}(v) := \{1 \leq j \leq n \, : \, x_j \text{ is below } v \text{ in } T_1\}.$$

Furthermore, for each vertex v in T'_1 that is not a leaf, denote by $\mathfrak{l}(v)$ that one of its two children for which the smallest element contained in $\mathbb{I}(\mathfrak{l}(v))$ is smaller than that of $\mathbb{I}(\mathfrak{r}(v))$, where $\mathfrak{r}(v)$ denotes the other child of v. Clearly, $\mathfrak{l}(v)$ (and therefore also $\mathfrak{r}(v)$) are uniquely determined and we call $\mathfrak{l}(v)$ and $\mathfrak{r}(v)$ the *left child* and *right child* of v, respectively.

Let $u_1, \dots, u_q \in V(T'_1)$, $q \geq 1$, denote a sequence of non-leaf vertices of T'_1 defined as follows. Put $u_1 = \rho(T_1')$ and, for all $1 \leq j \leq t$, assume that u_j has already been defined. If both children of u_t are leaves of T_1' or $|\mathbb{I}(\mathfrak{l}(u_t))| > 1$ holds then we stop and put $q = t$. Otherwise we put $u_{t+1} = \mathfrak{r}(u_t)$ and continue as before with t replaced by $t+1$. Note that in the case that both children of u_q are leaves of T_1' either $|\mathbb{I}(\mathfrak{l}(u_q))| > 1$ or $|\mathbb{I}(\mathfrak{l}(u_q))| = 1$ might hold. Also note that since T_1' is a finite tree there are only finitely many such vertices u_t.

Lemma 4. *(i) We have $j \in \mathbb{I}(u_j)$ for all $j \in \{1, \ldots, q\}$. Moreover, j is the smallest element in $\mathbb{I}(u_j)$.*

(ii) $|\mathbb{I}(u_q)| > 1$.

(iii) For all $j \in \{1, \ldots, q\}$, the length $l(u_j)$ of u_j is at least $j - 1$.

(iv) the q-th letter $[\mathfrak{l}(u_q)]_q$ of the left child $\mathfrak{l}(u_q)$ of u_q is 1.

Proof. (i) The recursive nature of the definition of the vertices u_j combined with the definition of the set $\mathbb{I}(\mathfrak{l}(u_j))$ implies that $j \in \mathbb{I}(\mathfrak{l}(u_j))$ holds for all $1 \leq j \leq q$. That j is the smallest element in $\mathbb{I}(\mathfrak{l}(u_j))$ is clear.

(ii) It suffices to consider the case that the construction of the list $u_1, \ldots, u_q$ terminated by finding that both children of u_q are leaves of T_1'. We claim that $|\mathbb{I}(\mathfrak{l}(u_q))| > 1$ must hold. Assume for contradiction that $|\mathbb{I}(\mathfrak{l}(u_q))| = 1$ and that both children of u_q are leaves of T_1'. Using the ordering of the elements of W induced by their indices, we obtain that T_1' must be equivalent with $\mathcal{C}_W$. Since the leaf set of $\mathcal{T}$ is X it follows that $\mathcal{T}_1$ and $\mathcal{C}^n$ are equivalent; a contradiction.

(iii) This follows from the fact that for all $1 \leq j \leq q-1$, the maximal precursor of u_{j+1} in $V(T_1')$ is u_j and u_1 is the empty sequence as u_1 is the root of T_1' and so has length zero.

(iv) This follows from Lemma 4(i) as it implies that $q \in supp(\mathfrak{l}(u_q))$. $\square$

We next present a technical result that lies at the heart of the proof that $\mathcal{H}_1$ and $\mathcal{H}_2$ are tree-equivalent. To establish it note that the termination criterion for the list of the vertices $u_1, \ldots, u_q$ implies that $\{j\} = \mathbb{I}(\mathfrak{l}(u_j))$ holds for all $j \in \{1, \ldots, q-1\}$. Hence, $\mathfrak{l}(u_j)$ is a leaf of T_1' for all such j. Without loss of generality we may assume for the remainder that $w_j^1 = \mathfrak{l}(u_j)$ holds for all $j \in \{1, \ldots, q-1\}$.

Theorem 5. *Setting $\mathcal{A} = V(T_1')$, there exists a map $\tau : \mathcal{A} \to \mathcal{A}_n$ such that the following holds*

(i) $\tau(w_j^1) \in \mathcal{B}_n^2$, *for all* $j \in \{1, \ldots, m\}$.

(ii) The trees $T_2' := \mathcal{P}[\tau(\mathcal{A})]$ *and* T_1' *are isomorphic and the underlying bijection maps, for all* $j \in \{1, \ldots, m\}$, *the sequence* w_j^1 *to* $\tau(w_j^1)$.

Proof. Put "$\prec$" = "$\prec_T$". Note first that in view of Assertions (i) and (ii) in Lemma 4 we may choose some $s \in \mathbb{I}(\mathfrak{l}(u_q))$ such that $s > q$. Let $t \in \mathbb{I}(\mathfrak{r}(u_q))$ denote the smallest element in $\mathbb{I}(\mathfrak{r}(u_q))$. Then $t > q$ must hold too. We establish the theorem by distinguishing between the cases that (a) $[\mathfrak{r}(u_q)]_q = 0$ and that (b) $[\mathfrak{r}(u_q)]_q = 1$.

Case (a): Set

$$\tau : \mathcal{A} \to \mathcal{A}_n : \ v \mapsto \begin{cases} \psi_t(v) & \text{if } v \preceq \mathfrak{l}(u_q) \\ \psi_s(v) & \text{otherwise.} \end{cases}$$

Then since $\tau(w_j^1)$ and w_j^1 differ in precisely one letter and $w_j^1 \in \mathcal{B}_n^1$ holds for all $1 \leq j \leq m$, we must have $\tau(w_j^1) \in \mathcal{B}_n^2$, for all $j \in \{1, \ldots, m\}$. Thus, (i) holds in this case.

To see (ii) for this case, we show first that τ preserves the precursor relationships between the sequences in $\mathcal{A}$. Put differently, we show that for any two distinct sequences $w, w' \in \mathcal{A}$ we have that if w is a precursor of w' in $\mathcal{A}$ then $\tau(w)$ is a precursor of $\tau(w')$ in $\tau(\mathcal{A})$. Our arguments are based on a detailed case analysis. Observe first that Lemma 4(i) and (iii) imply that $l(u_q) = q - 1$. Combined with the choice of s and t it follows for all

$1 \leq j \leq q$ that $l(u_j) \leq l(u_q) = q-1 < z$ for $z \in \{t, s\}$. This yields, $\tau(u_j) = \psi_s(u_j) = u_j$ for all such j. Let w and w' denote two distinct sequences in $\mathcal{A}$. If $w = u_j$ and $w' = u_{j+1}$ for some $1 \leq j \leq q-1$ then w is the maximal precursor of w' in $\mathcal{A}$. Hence, $\tau(w) = u_j$ is the maximal precursor of $\tau(w') = u_{j+1}$ in $\tau(\mathcal{A})$. If there exists some $1 \leq j \leq q-1$ such that $w = u_j$ and $w' = w_j^1$ then w is the maximal precursor of w' in $\mathcal{A}$. Hence, $\tau(w) = w$ is the maximal precursor of $\tau(w') = w_j^2$ in $\tau(\mathcal{A})$. Now assume that w and w' are such that neither one of the above two cases applies. Then either (α) $w \preceq \mathfrak{l}(u_q)$ and $w' \preceq \mathfrak{r}(u_q)$, or (β) $w, w' \preceq \mathfrak{l}(u_q)$, or (γ) $w, w' \preceq \mathfrak{r}(u_q)$ holds.

If Case (α) holds, we claim that neither w is precursor of w' nor w' is a precursor of w. Indeed, since $t > q$ and $\mathfrak{l}(u_q)$ is a precursor of w, we have $[\tau(w)]_q = [\psi_t(w)]_q = [w]_q = [\mathfrak{l}(u_q)]_q = 1$ and since $\mathfrak{r}(u_q)$ is a precursor of w', we also have $[\tau(w')]_q = [\psi_s(w)]_q = [w']_q = [\mathfrak{r}(u_q)]_q = 0$. Hence, neither w is a precursor of w' nor w' is a precursor of w, as claimed. If either Case (β) or Case (γ) holds then it is straightforward to see that w is a precursor of w' if and only if $\tau(w)$ is a precursor of $\tau(w')$. Since $\tau(u_q)$ is the common precursor of $\{\tau(\mathfrak{l}(u_q)), \tau(\mathfrak{r}(u_q))\}$ in $\tau(\mathcal{A})$ of maximal length, it follows that τ preserves the precursor relationships between the elements in $\mathcal{A}$, as required. Setting $T_2' = \mathcal{P}[\tau(\mathcal{A})]$ implies (ii) in this case.

Case (b): For $1 \leq j \leq q-1$, let w_j^2 be the sequence in $\mathcal{B}_n^2$ with $supp(w_j^2) = \{j, n\}$. For all $1 \leq j \leq q$, put $u_j' = \mathbf{0_{j-1}}$. Note that u_1' is the empty sequence in $\mathcal{B}_0$. Let α denote the precursor of u_{q-1} of length $l(\alpha) = q-1$ which exists by Lemma 4(iii). If $l(\alpha) \geq 1$ then put $supp(\alpha) = \{\alpha_1, \ldots, \alpha_r\}$ where $r = |\,supp(\alpha)|$. Also, put $\kappa = 1$ if $|\,supp(\alpha)|$ is odd and 0 if $|\,supp(\alpha)|$ is even or $supp(\alpha) = \emptyset$. Let $\varphi : \mathcal{A} \to \mathcal{A}$ denote the map defined by setting, for all $v \in \mathcal{A}$, $\varphi(v) = \psi_{\alpha_1} \circ \psi_{\alpha_2} \circ \cdots \circ \psi_{\alpha_r}(v)$ if $supp(\alpha) \neq \emptyset$ and $\varphi(v) = v$ otherwise. Note that for all $v \in \mathcal{A}$ with $v \preceq u_q$ holding, we have $supp(\varphi(v)) \cap \{1, \ldots, q-1\} = \emptyset$ as α is also a precursor of $\varphi(v)$ and so $[\varphi(v)]_j = 0$ holds for all $1 \leq j \leq r$. Let $\tau : \mathcal{A} \to \mathcal{A}_n$ denote the map

$$\tau : \mathcal{A} \to \mathcal{A}_n : \quad v \mapsto \begin{cases} w_j^2 & \text{if } v = w_j^1 \text{ for some } 1 \leq j \leq q-1, \\ u_j' & \text{if } v = u_j \text{ for some } 1 \leq j \leq q, \\ \varphi \circ \psi_t(v) & \text{if } \kappa = 0 \text{ and } v \preceq \mathfrak{l}(u_q), \\ \varphi \circ \psi_q(v) & \text{if } \kappa = 0 \text{ and } v \preceq \mathfrak{r}(u_q), \\ \varphi(v) & \text{if } \kappa = 1 \text{ and } v \preceq \mathfrak{l}(u_q), \text{ and} \\ \varphi \circ \psi_q \circ \psi_s(v) & \text{if } \kappa = 1 \text{ and } v \preceq \mathfrak{r}(u_q), \end{cases}$$

and for all $q \leq j \leq m$, put $w_j^2 = \tau(w_j^1)$. Then $w_j^2 \in \mathcal{B}_n^2$ clearly holds for all $1 \leq j \leq q-1$ and $w_j^2 \in \mathcal{B}_n^2$ holds for all $q \leq j \leq m$ because $w_j^1 \in \mathcal{B}_n^1$ and τ "flips" an odd number of positions in w_j^1. Thus (i) holds in this case, too.

To see (ii), we next show that τ preserves the precursor relationships in $\mathcal{A}$ (in the same sense as in the case that $[\mathfrak{r}(u_q)]_q = 0$). Note that, by construction, the maximal precursor of w_j^1 in $\mathcal{A}$ is u_j' for all $1 \leq j \leq q-1$ and the maximal precursor of u_{j+1}' in $\mathcal{A}$ is u_j', for all $1 \leq j \leq q-1$. Let w and w' denote two distinct sequences in $\mathcal{A}$. If $w \preceq \mathfrak{l}(u_q)$ and $w' \preceq \mathfrak{r}(u_q)$ then $[\tau(w)]_q = 1$ and $[\tau(w')]_q = 0$. Hence, neither $\tau(w)$ is precursor of $\tau(w)'$ nor $\tau(w)'$ is a precursor of $\tau(w)$. Setting $T_2' = \mathcal{P}[\tau(\mathcal{A})]$ and using similar arguments as in the corresponding analysis for Case (a) implies that (ii) also holds in this case.

$\square$

Theorem 6. *The networks $\mathcal{H}_1$ and $\mathcal{H}_2$ are tree-equivalent.*

Proof. Let $\mathcal{T}_1 \in \mathscr{T}(\mathcal{H}_1)$. We show that $\mathcal{T}_1$ is also displayed by $\mathcal{H}_1$, from which the theorem immediately follows by symmetry. We consider the cases (i) $\mathcal{T}_1$ and $\mathcal{C}^n$ are equivalent, and (ii) they are not.

Case (i): Assume first that n is even. Then $\mathbf{1_n} \in \mathcal{B}_n^2$ and so, by the construction of the graph $\mathcal{D}_{2,3}$, the tree $\mathcal{C}^n$ is displayed by $\mathcal{H}_2$. Since $\mathcal{T}_1$ and $\mathcal{C}^n$ are assumed to be equivalent, it follows that $\mathcal{T}_1$ is displayed by $\mathcal{H}_2$, as required.

So, suppose n is odd. Then the sequences $\tilde{w}, \overline{w} \in \mathcal{B}_n$ with $supp(\tilde{w}) = \{1, n\}$ and $supp(\overline{w}) = \{2, 3, \ldots, n\}$ are clearly contained in $\mathcal{B}_n^2$. Let $\mathcal{Z}_{\overline{w}}$ and $\mathcal{Z}_{\tilde{w}}$ denote the rooted trees associated to $\overline{w}$ and $\tilde{w}$, respectively, in the construction of the graph $\mathcal{D}_{2,3}$. We now construct a phylogenetic tree $\mathcal{T}$ on X from $\mathcal{Z}_{\overline{w}}$ and $\mathcal{Z}_{\tilde{w}}$.

First construct $\mathcal{C}_{\mathbf{1_n}}$ by joining the roots of $\mathcal{Z}_{\overline{w}}$ and $\mathcal{Z}_{\tilde{w}}$ respectively to the common precursor of $\{\overline{w}, \tilde{w}\}$ in $\mathcal{B}_n^2$ of maximal length and then deleting the arc $(\tilde{w}, n)$. Next transform $\mathcal{C}_{\mathbf{1_n}}$ into a phylogenetic tree on X by attaching to each leaf l of $\mathcal{C}_{\mathbf{1_n}}$ the directed path from l to the unique leaf in $\mathcal{H}_2$ below l and then suppressing all degenerate vertices. The resulting tree is $\mathcal{T}$.

Now, since by construction $\mathcal{T}$ is equivalent with $\mathcal{C}^n$ and, by assumption, $\mathcal{C}^n$ is equivalent with $\mathcal{T}_1$ it follows that $\mathcal{T}_1$ is displayed by $\mathcal{H}_2$, as required.

Case (ii): Let τ and T_2' be as specified in Theorem 5. Note that T_2' is a binary phylogenetic tree on $\{\tau(w_1^1), \ldots, \tau(w_m^1)\}$. For all $1 \leq j \leq m$, put $A_j = \mathbb{I}(w_j^1)$ and let $T(A_j)$ be the tree obtained from the rooted caterpillar $\mathcal{C}_{A_j}$ by replacing each of its leaves l with the corresponding element x_l in X.

Now, let $\mathcal{T}_2$ be the phylogenetic tree on X obtained by replacing each leaf $\tau(w_j^1)$ of T_2' with $T(A_j)$. Since T_2' is a binary tree with $V(T_2') \subseteq \mathcal{A}_n$ and $L(T_2') \subseteq \mathcal{B}_n^2$, Lemma 3 implies that T_2' is displayed by $\mathcal{P}_n|_{\mathcal{B}_n^2}$. Since $\mathcal{T}_2$ can also be obtained from $\mathcal{D}_2$ by deleting a set of arcs and removing resulting isolated vertices, it follows that $\mathcal{T}_2$ is displayed by $\mathcal{H}_2$. Since $\mathcal{T}_1$ can be obtained from T_1' by replacing each of its leaves w_j^1 with $T(A_j)$, it follows that $\mathcal{T}_1$ and $\mathcal{T}_2$ must be equivalent. Hence, $\mathcal{T}_1$ is displayed by $\mathcal{H}_2$, as required. $\square$

*

References

van Iersel, L., C. Semple, and M. Steel. 2010. Locating a tree in a phylogenetic network. Information Processing Letters 110:1037–1043.

Willson, S. 2011. Regular networks can be uniquely constructed from their trees. IEEE/ACM Transactions on Computational Biology and Bioinformatics 8(3):785–796.

Additional Figures

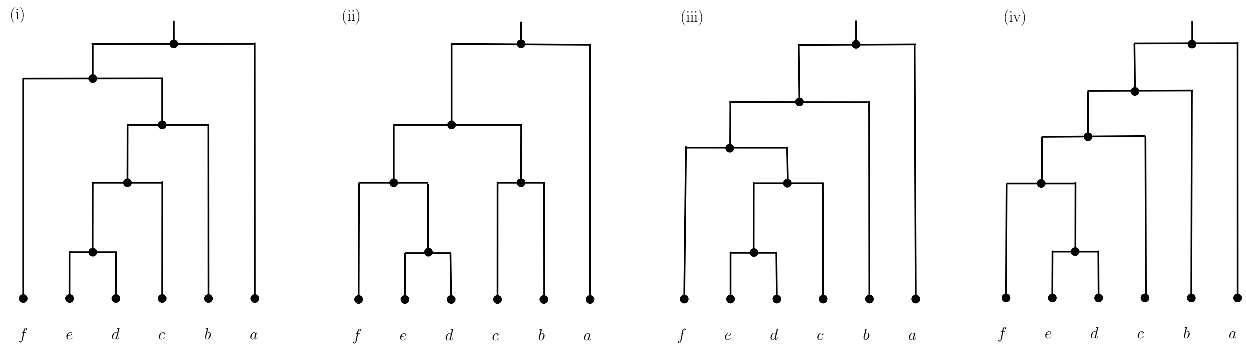

Figure 1: The four evolutionary trees displayed by both networks (i) and (ii) in Fig. 1 of the main text.

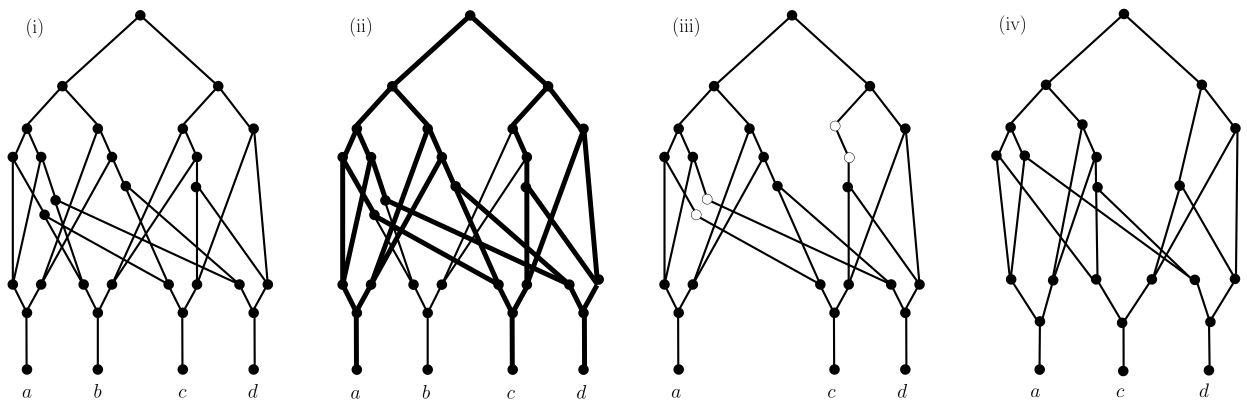

Figure 2: An illustration of obtaining a trinet from a network. (i) The network N on $X = \{a, b, c, d\}$ pictured in Fig. 2(i) of the main text. (ii) The network N with the paths from the root to leaves a, c, d highlighted in bold. (iii) The network on $\{a, c, d\}$ formed by the edges of the highlighted paths in (ii), where degenerate vertices are represented by unfilled circles. (iv) The trinet on Y induced by N, which is obtained from network (iii) by suppressing all degenerate vertices.

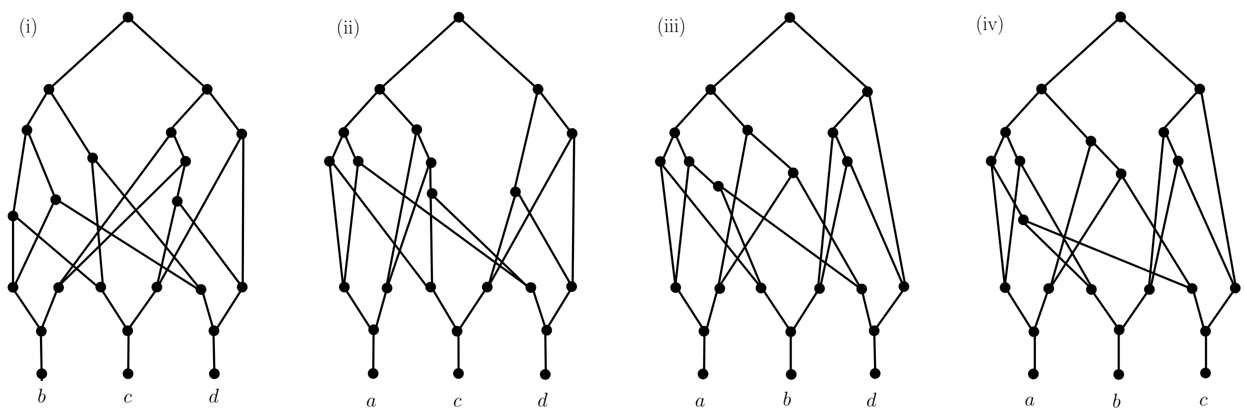

Figure 3: The four trinets induced by both networks (i) and (ii) in Fig. 2 of the main text.