UNICYCLIC NETWORKS: COMPATIBILITY AND ENUMERATION

CHARLES SEMPLE AND MIKE STEEL

Abstract. Graphs obtained from a binary leaf labelled ('phylogenetic') tree by adding an edge so as to introduce a cycle provide a useful representation of hybrid evolution in molecular evolutionary biology. This class of graphs (which we call 'unicyclic networks') also has some attractive combinatorial properties, which we present. We characterize when a set of binary phylogenetic trees is displayed by a unicyclic network in terms of tree rearrangement operations. This leads to a triple-wise compatibility theorem, and a simple, fast algorithm to determine 1-cycle compatibility. We also use generating function techniques to provide closed-form expressions that enumerate unicyclic networks with specified or unspecified cycle length, and we provide an extension to enumerate a class of multi-cyclic networks.

Date: 8 July 2005.

1991 *Mathematics Subject Classification.* 05C05; 92D15.

Key words and phrases. Phylogenetic tree, compatibility, circular orderings, generating function, galled-trees.

We thank the New Zealand Marsden Fund (UOC310) for supporting this research.

Corresponding author. Charles Semple.

1. Introduction

Although phylogenetic trees provide a useful representation of many evolutionary relationships, and have been well studied (see, for example, [4, 20]), there is increasing interest in using non-tree graphs to model reticulate evolution. Indeed during the last few years there has been a burst of activity in phylogenetic bioinformatics in developing methods to reconstruct and model reticulation—for example, see [1, 2, 6, 11, 7, 8, 9, 10, 15, 22, 23]. Reticulate evolution can be due to a variety of biological processes, including recombination, horizontal gene transfer, genome fusion, and the formation of hybrid species (as occurs in certain plant, insect and animal species) [14, 18]. The simplest type of non-tree graph are those that contain a single cycle, and it is this class that we study here.

This class has recently come to prominence in the description by Rivera and Lake [17] of a "ring of life" to better understand the origin of eukaryotes. These authors analysed ten complete genomes from prokaryotic and eukaryotic organisms, and found support for five conflicting trees – nevertheless, these five unrooted trees fitted perfectly into a network with a single cycle (for further details, see also [13]).

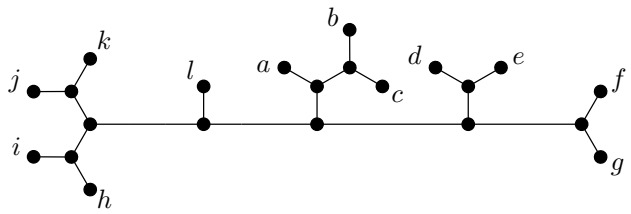

FIGURE 1. A binary phylogenetic tree.

To describe these types of single-cycle networks further, and to outline our results, we first introduce some definitions.

A *binary phylogenetic tree (on X)* is a tree $\mathcal{T}$ in which every interior vertex has degree three and whose leaf set is X. The set X is often referred to as the *label set* of $\mathcal{T}$ and its elements as *labels*. For example, a binary phylogenetic tree is shown in Fig. 1. Here $X = \{a, b, \ldots, l\}$. A *unicyclic network (on X)* is a graph $\mathcal{G}$ that has exactly one cycle (of length at least three), every interior vertex has degree three, and the set of degree-one vertices is X. Thus, by deleting a single edge of the cycle in $\mathcal{G}$ and suppressing the resulting degree-two vertices, we obtain a binary phylogenetic X-tree. Indeed, we say $\mathcal{G}$ *displays* a binary phylogenetic X-tree $\mathcal{T}$ if $\mathcal{T}$ can be obtained from $\mathcal{G}$ in this way. In general, let $\mathcal{P}$ be a collection of phylogenetic X-trees. Then $\mathcal{G}$ *displays* $\mathcal{P}$ if $\mathcal{G}$ displays each tree in $\mathcal{P}$, in which case we say that $\mathcal{P}$ is 1-*cycle compatible*. To illustrate these definitions, the unicyclic network shown in Fig. 2 displays the binary phylogenetic tree shown in Fig. 1.

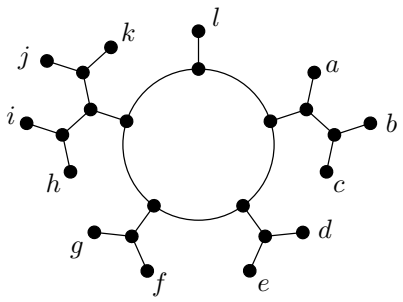

FIGURE 2. A unicyclic network.

Note that if $\mathcal{G}$ is a unicyclic network on X whose cycle has length k, then $\mathcal{G}$ displays exactly $k-2$ binary phylogenetic trees on X. In particular, if $k=3$, then $\mathcal{G}$ displays just one binary phylogenetic tree, namely the tree obtained from $\mathcal{G}$ by collapsing the cycle of length 3 to a single vertex. Although one could exclude cycles of length 3, we have found it convenient (particularly for enumerating galled-trees) to allow them.

Two unicyclic networks $\mathcal{G}$ and $\mathcal{G}'$ on X are *isomorphic* if there is a graph isomorphism from $\mathcal{G}$ to $\mathcal{G}'$ which when restricted to X is the identity map.

One of the main questions that motivates this study is the following: Given a collection $\mathcal{P}$ of binary phylogenetic trees on X when is $\mathcal{P}$ 1-cycle compatible? For $|\mathcal{P}|=2$, this question is closely related to tree rearrangement operations, and the number of possible unicyclic networks that display $\mathcal{P}$ is either 0, 1, or 3. When $\mathcal{P}$ has arbitrary size, the 1-cycle

compatibility question can be reduced to consideration of triples of trees from $\mathcal{P}$, allowing for a simple polynomial-time algorithm.

In this paper we also consider the enumeration of unicyclic networks, where the cycle length is either specified or left unspecified, and we use this to derive further enumerative results. We then provide an extension to count a certain class of networks ("galled-trees") where multiple cycles are allowed

1.1. **Biological relevance.** The modelling and analysis of reticulate evolution is currently a topical problem in systematic biology and bioinformatics. Most studies to date have dealt only with rooted trees as their input ([12, 16, 18]). Although certain processes (such as the formation of hybrid species) are normally viewed as requiring some time scale (i.e. at some time in the past two species exchanged genetic material), it is often desirable to have techniques for describing reticulate evolution when the input trees are unrooted. This is because most tree reconstruction methods (such as neighbour joining and maximum likelihood) output unrooted trees. We can formally ask whether conflicting unrooted trees (perhaps from different genes) can be reconciled by a single cycle, as in the study by Rivera and Lake [17]. We show that this unrooted compatibility question

has a concise mathematical and algorithmic description. In general, a unicyclic network may display more trees than those provided as input to the algorithm, however, these additional trees need not be regarded as having any particular biological significance. We also describe exact formulae for enumerating unicyclic networks and generalizations to allow several disjoint cycles. The underlying decomposition that leads to these formulae may in turn be useful for sampling uniformly from the set of such networks.

Of course, one may wish to consider more general and complex structures for modelling reticulate evolution than those considered in this paper – for example, by allowing multiple inter-twining cycles, or by allowing non-binary trees and networks. However our aim here is to provide an attractive mathematical foundation for a simple model of reticulate evolution, rather than an algorithmic analysis of a more complex scenario (for some approaches to the latter, see [6, 10]).

We end this section with some preliminaries that will be used throughout the paper.

1.2. **Preliminaries.** Throughout the paper, the notation and terminology follows [20]. An X-*split* is a partition of X into two non-empty sets. We denote the X-split whose blocks are A and B by $A|B$. Associated with

every phylogenetic X-tree $\mathcal{T}$ is a particular collection of X-splits. This collection consists of those X-splits $A|B$ that are induced by the components of the graph resulting from the deletion of a single edge e of $\mathcal{T}$. We say that the X-split $A|B$ *corresponds to* e and let $\Sigma(\mathcal{T})$ denote the set of X-splits that correspond to the edges of $\mathcal{T}$.

Let $\pi = (x_1, x_2, \ldots, x_n)$ be a cyclic permutation of X. For all $1 \leq i \leq j \leq n$, let $A_{ij} = \{x_k : i \leq k \leq j\}$ and let $\Sigma^\circ(\pi)$ denote the set

$$\Sigma^\circ(\pi) = \{A_{ij}|(X - A_{ij}) : 1 \leq i \leq j \leq n-1\}$$

of X-splits. Arranging the elements $x_1, x_2, \ldots, x_n$ clockwise in a circle in the plane, we may view $\Sigma^\circ(\pi)$ as the set of X-splits that can be obtained by separating these elements according to which side of a line segment in the plane they lie on. Consequently, $|\Sigma^\circ(\pi)| = \binom{n}{2}$. A collection Σ of X-splits is said to be *circular* if $\Sigma \subseteq \Sigma^\circ(\pi)$ for some cyclic permutation π of X. In case $\Sigma(\mathcal{T}) \subseteq \Sigma^\circ(\pi)$ for some phylogenetic X-tree $\mathcal{T}$, we say that π provides a *circular ordering* for $\mathcal{T}$. This last definition has an equivalent formulation as follows. Suppose we embed $\mathcal{T}$ in the plane, and trace around the outside of $\mathcal{T}$ beginning at some leaf $x \in X$ and eventually returning to x (in this way each edge of $\mathcal{T}$ is traversed exactly twice—once in each direction). The order in which the elements of X are met in this tracing induces a circular ordering for $\mathcal{T}$. The set of circular orderings for $\mathcal{T}$ is precisely the set of

orderings on X that are induced by tracing across all planar embeddings of $\mathcal{T}$. Similarly, we have an analogous notion of a *circular ordering* for a unicyclic network.

2. 1-Cycle Compatibility

In this section, we investigate the problem of determining precisely when a collection $\mathcal{P}$ of binary phylogenetic X-trees is 1-cycle compatible. This problem is motivated by the analysis in [17]. In the case $|\mathcal{P}| = 2$, this problem has an attractive solution in terms of tree rearrangements which we describe next. This solution will enable us to handle the case $|\mathcal{P}| \geq 3$ later in the section.

Let $\mathcal{T}$ be a binary phylogenetic X-tree and let $e = \{u, v\}$ be an edge of $\mathcal{T}$. Let $\mathcal{T}'$ be the binary phylogenetic X-tree that is obtained from $\mathcal{T}$ by deleting e, and then attaching the component C_v that contains v to the component C_u that contains u by adjoining a new edge f from C_v to C_u so that, once degree-two vertices are suppressed, the resulting tree is a binary phylogenetic X-tree. The two tree rearrangement operations that we now describe are restricted by how this new edge is adjoined. We begin with the least restrictive operation.

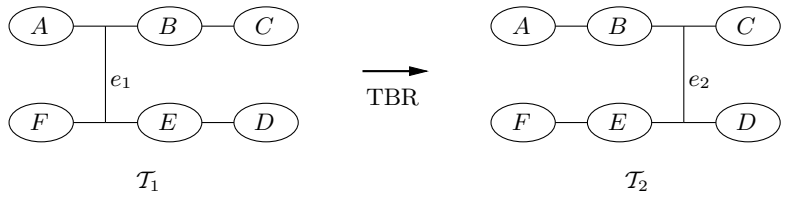

FIGURE 3. A schematic diagram of a TBR operation.

(i) We say that $\mathcal{T}'$ has been obtained from $\mathcal{T}$ by a *tree bisection and reconnection* (TBR) if there is no restriction on f.

(ii) We say that $\mathcal{T}'$ has been obtained from $\mathcal{T}$ by an (unrooted) *subtree prune and regraft* (SPR) if one end-vertex of f is v.

Observe that SPR is a special case of TBR. For further details of tree rearrangement operations, see [20].

The diagram shown in Fig. 3 is a schematic representation of a single TBR operation, where $\mathcal{T}_1$ and $\mathcal{T}_2$ are two binary phylogenetic X-trees. If B and E are both empty, then $\mathcal{T}_1$ is isomorphic to $\mathcal{T}_2$, and so the TBR operation is redundant. Furthermore, it is easily checked that the TBR operation is an SPR operation precisely if either $|A \cup B \cup C| = 1$ or $|D \cup E \cup F| = 1$, or one of B or E is empty. We will make use of this diagram in the next section and we may assume that, provided $|A \cup B \cup C|, |D \cup E \cup F| \geq 2$, we have $|A|, |C|, |D|, |F| \geq 1$.

Tree rearrangement operations play an important role in phylogenetics. One reason for this is that they each induce a metric on the collection of binary phylogenetic X-trees and thus enable one to quantify the "closeness" of any pair of such trees. In particular, let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two binary phylogenetic X-trees and let $\Theta \in \{\mathrm{SPR}, \mathrm{TBR}\}$. The Θ-*distance* between $\mathcal{T}_1$ and $\mathcal{T}_2$ is the minimum number of operations that is required to transform $\mathcal{T}_1$ into $\mathcal{T}_2$. We denote this distance by $d_\Theta(\mathcal{T}_1, \mathcal{T}_2)$. It is well-known that, for each Θ, one can always get from $\mathcal{T}_1$ to $\mathcal{T}_2$ by such a sequence of operations and $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) \leq d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) \leq 2d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2)$.

Theorem 2.1. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two distinct binary phylogenetic X-trees. Then there is a unicyclic network $\mathcal{G}$ on X that displays $\{\mathcal{T}_1, \mathcal{T}_2\}$ if and only if $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$. Moreover, in that case, there are unique edges e_1 and e_2 such that, up to suppressing degree-two vertices, $\mathcal{G}\backslash e_1$ is isomorphic to $\mathcal{T}_2$ and $\mathcal{G}\backslash e_2$ is isomorphic to $\mathcal{T}_1$.*

Proof. Suppose that there is a unicyclic network $\mathcal{G}$ on X that displays both $\mathcal{T}_1$ and $\mathcal{T}_2$. Then, as $\mathcal{T}_1$ and $\mathcal{T}_2$ are distinct, it follows by definition that there are two distinct edges e_1 and e_2 such that, up to suppressing degree-two vertices, $\mathcal{G}\backslash e_1$ and $\mathcal{G}\backslash e_2$ are isomorphic to $\mathcal{T}_1$ and $\mathcal{T}_2$. This implies that, for each i, $\mathcal{T}_i$ can be obtained from $\mathcal{G}\backslash\{e_1, e_2\}$ by adding e_i in the appropriate way. By the definition of TBR, we deduce that $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$.

Now suppose that $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$. Then, up to suppressing degree-two vertices, $\mathcal{T}_2$ can be obtained from $\mathcal{T}_1$ by deleting an edge e_1 say in $\mathcal{T}_1$, and then joining the resulting components by a new edge e_2 say. Now let $\mathcal{G}$ be the graph that is obtained from $\mathcal{T}_1$ by adding e_2 so that $\mathcal{G}\backslash e_1$ is isomorphic to $\mathcal{T}_2$. Since adding e_2 creates exactly one cycle, it follows that $\mathcal{G}$ is a unicyclic network on X. Moreover, up to suppressing degree-two vertices, $\mathcal{G}\backslash e_1$ and $\mathcal{G}\backslash e_2$ are isomorphic to $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. Thus $\mathcal{G}$ displays $\mathcal{T}_1$ and $\mathcal{T}_2$.

Lastly, suppose there is a unicyclic network $\mathcal{G}$ on X that displays $\mathcal{T}_1$ and $\mathcal{T}_2$. Since no two distinct edges f and f' of the cycle of $\mathcal{G}$ have the property that $\mathcal{G}\backslash f$ is isomorphic to $\mathcal{G}\backslash f'$, it follows that the choice of e_1 and e_2 is unique. This completes the proof of the theorem. □

Proposition 2.2. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two distinct binary phylogenetic X-trees. If $\{\mathcal{T}_1, \mathcal{T}_2\}$ is 1-cycle compatible, then $\Sigma(\mathcal{T}_1) \cup \Sigma(\mathcal{T}_2)$ is circular.*

Proof. Let $\mathcal{G}$ be a unicyclic network on X that displays $\mathcal{T}_1$ and $\mathcal{T}_2$. Let $x \in X$. Viewing $\mathcal{G}$ drawn in the plane with its leaves on the outside of the cycle, trace around the outside of $\mathcal{G}$ beginning at x, eventually returning to x. Let π be the cyclic permutation of X induced by the order in which the elements of X are met in this tracing. It is now easily checked that π

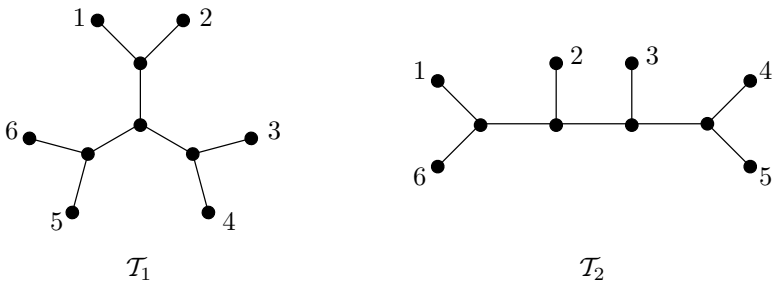

FIGURE 4. A counterexample to the converse of Proposition 2.2.

is a circular ordering for both $\mathcal{T}_1$ and $\mathcal{T}_2$, thus completing the proof of the proposition. □

We remark here that the converse of Proposition 2.2 does not hold. For a counterexample, consider the pair of trees $\{\mathcal{T}_1, \mathcal{T}_2\}$ in Fig. 4. Then, with $\pi = (1, 2, \ldots, 6)$, we have $\Sigma(\mathcal{T}_1) \cup \Sigma(\mathcal{T}_2) \subseteq \Sigma^\circ(\pi)$, and so $\Sigma(\mathcal{T}_1) \cup \Sigma(\mathcal{T}_2)$ is circular. However, $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) \geq 2$, and therefore, by Theorem 2.1, $\{\mathcal{T}_1, \mathcal{T}_2\}$ is not 1-cycle compatible.

We now consider the problem of determining precisely when an arbitrary collection of binary phylogenetic X-trees is 1-cycle compatible. To this end, we begin with the following proposition.

Proposition 2.3. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two binary phylogenetic trees on X, and suppose that $\{\mathcal{T}_1, \mathcal{T}_2\}$ is 1-cycle compatible. Then*

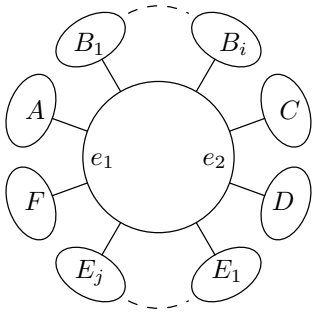

FIGURE 5. A schematic view of the unicyclic network described in (i) of Proposition 2.3.

(i) *If* $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) \neq 1$*, then there is exactly one unicyclic network on* X *that displays* $\mathcal{T}_1$ *and* $\mathcal{T}_2$*.*

(ii) *If* $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$ *and the pruned subtree consists of a single leaf, then there is exactly one unicyclic network on* X *that displays* $\mathcal{T}_1$ *and* $\mathcal{T}_2$*.*

(iii) *If* $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$ *and the pruned subtree has at least two leaves, then there are exactly three unicyclic networks on* X *that display* $\mathcal{T}_1$ *and* $\mathcal{T}_2$*.*

Proof. It follows by the definition of display that all unicyclic networks on X that display both $\mathcal{T}_1$ and $\mathcal{T}_2$ can be obtained by starting with $\mathcal{T}_1$ and adjoining a new edge e_2. The edge e_2 is added in such a way that $\mathcal{T}_2$ can be obtained from the resulting unicyclic network on X by deleting an edge e_1. By Theorem 2.1, there is exactly one choice for e_1. Thus to prove the proposition, it suffices to consider the possible ways by which e_2

can be added to $\mathcal{T}_1$. In establishing each of (i)—(iii), we make use of the schematic diagram of a TBR operation shown in Fig. 3. With regards to this diagram, it is clear that e_2 must join an edge of the minimal subtree of $\mathcal{T}_1$ that connects $A \cup B \cup C$ to an edge of the minimal subtree of $\mathcal{T}_1$ that connects $D \cup E \cup F$. Furthermore, as $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$ or $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$, we have $|X| \geq 4$.

First consider (i). Since $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$, but $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) \neq 1$, we may assume that $|A|, |B|, |C|, |D|, |E|, |F| \geq 1$ in Fig. 3. By noting that $A|(X - A), C|(X - C), D|(X - D), F|(X - F)$ are all X-splits of $\mathcal{T}_2$, this added edge cannot be joined to edges in any of the subtrees labelled A, C, D, and F. Furthermore, as $(A \cup B)|(X - (A \cup B))$ and $(E \cup F)|(X - (E \cup F))$ are both X-splits of $\mathcal{T}_2$, this added edge cannot be joined to edges in B or E. It now follows that there is exactly one way in which e_2 can be appropriately added to $\mathcal{T}_1$. Thus there is exactly one unicyclic network on X that displays both $\mathcal{T}_1$ and $\mathcal{T}_2$. This unicyclic network is schematically shown in Fig. 5, where $B_1, \ldots, B_i$ $(i \geq 1)$ are the subtrees of B attached to the path from e_1 to e_2, and $E_1, \ldots, E_j$ $(j \geq 1)$ are the subtrees of E attached to the path from e_2 to e_1.

Now consider (ii). Without loss of generality, we may assume that, in Fig. 3, $|A| = 1$, and B and C are both empty. Using an approach similar

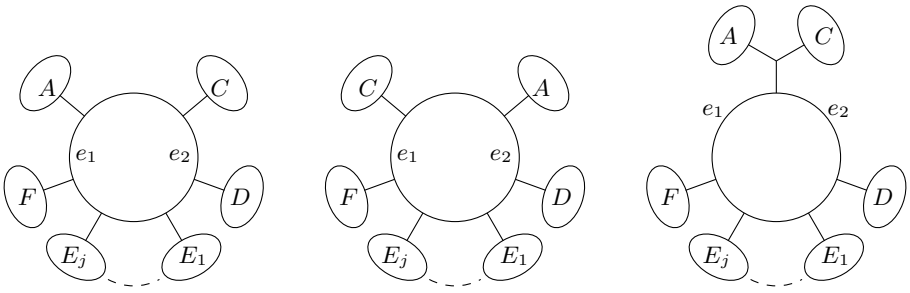

FIGURE 6. A schematic view of the unicyclic networks described in (iii) of Proposition 2.3.

to that in (i), it is easily seen that in this case there is also exactly one unicyclic network on X that displays both $\mathcal{T}_1$ and $\mathcal{T}_2$.

Lastly, consider (iii). In this case, as $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$ and the pruned subtree has at least two leaves, precisely one of B or E is empty, and $|A|, |C|, |D|, |F| \geq 1$. Without loss of generality, we may assume that B is empty, in which case E is non-empty. Again using the approach used in (i), we deduce, in this case, that there are exactly three unicyclic networks on X that display both $\mathcal{T}_1$ and $\mathcal{T}_2$. These three unicyclic networks are schematically shown in Fig. 6. This completes the proof of the proposition. □

Theorem 2.4. *Let $\mathcal{P}'$ be a collection of binary phylogenetic trees on X with $|\mathcal{P}'| \geq 3$. Then $\mathcal{P}'$ is 1-cycle compatible if and only if, for all subsets $\mathcal{P}$ of size three, $\mathcal{P}$ is 1-cycle compatible, in which case there is a unique unicyclic network on X that displays $\mathcal{P}'$.*

Proof. If there is a unicyclic network $\mathcal{G}$ on X that displays $\mathcal{P}'$, then every 3-element subset of $\mathcal{P}'$ is displayed by $\mathcal{G}$. This proves one direction of the theorem.

For the converse, suppose that $\mathcal{P}$ is 1-cycle compatible for every 3-element subset $\mathcal{P}$ of $\mathcal{P}'$. First assume that there is a pair $\mathcal{T}_1$ and $\mathcal{T}_2$ in $\mathcal{P}'$ such that either the assumptions of (i) or (ii) in the statement of Proposition 2.3 hold. In either case, it follows by Proposition 2.3 that there is exactly one unicyclic network, $\mathcal{G}$ say, on X that displays $\mathcal{T}_1$ and $\mathcal{T}_2$. Since $\mathcal{G}$ is unique and every 3-element subset of $\mathcal{P}'$ is 1-cycle compatible, we now deduce that, for each $i \in \{3, 4, \ldots, |\mathcal{P}'|\}$, there is exactly one unicyclic network that displays $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_i\}$ and that this unicyclic network is always $\mathcal{G}$. Hence, in this case, $\mathcal{P}'$ is 1-cycle compatible and there is a unique unicyclic network on X that displays $\mathcal{P}'$.

Now assume that, for every pair of trees in $\mathcal{P}'$, the assumptions of (iii) in Proposition 2.3 hold. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be a pair of trees in $\mathcal{P}'$. Then, by Proposition 2.3, there are exactly three unicyclic networks, $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ say, on X that display $\mathcal{T}_1$ and $\mathcal{T}_2$. Now consider $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_i\}$, where $\mathcal{T}_i \notin \{\mathcal{T}_1, \mathcal{T}_2\}$. By assumption, there is a unicyclic network on X that displays $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_i\}$. Moreover, this tree must be one of the three unicyclic networks that display $\mathcal{T}_1$ and $\mathcal{T}_2$. For each $j \in \{1, 2, 3\}$, it follows by

Theorem 2.1 that, up to degree-two vertices, there is a unique pair of edges in $\mathcal{G}_j$ such that the deletion of one results in $\mathcal{T}_1$ and the deletion of the other results in $\mathcal{T}_2$. By considering the remaining edges of the cycles of $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$, it is straightforward to deduce that the binary phylogenetic X-trees that result by deleting such an edge are distinct. This implies that there is exactly one unicyclic network on X that displays $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_i\}$. If, for all i, the unicyclic network displaying $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_i\}$ is the same, then $\mathcal{P}'$ is 1-cycle compatible and this unicyclic network on X is the only such network. Therefore assume that for some distinct i and j, the unicyclic network that displays $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_i\}$ is not isomorphic to the unicyclic network that displays $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_j\}$. We may also assume that the former network is $\mathcal{G}_1$ and the latter network is $\mathcal{G}_2$. By an argument similar to that used earlier in this paragraph, there is a unique unicyclic network that displays $\{\mathcal{T}_1, \mathcal{T}_i, \mathcal{T}_j\}$. Since $\mathcal{G}_1$ displays $\{\mathcal{T}_1, \mathcal{T}_i\}$, we deduce that it is $\mathcal{G}_1$. But $\mathcal{G}_1$ does not display $\mathcal{T}_j$; a contradiction. This completes the proof of the theorem. □

The sufficient part of the hypothesis in Theorem 2.4 is sharp in the sense that it is not sufficient for $\mathcal{P}'$ to be 1-cycle compatible if every subset of $\mathcal{P}'$ of size two is 1-cycle compatible. To see this, take $\mathcal{P}'$ to be the collection consisting of all three binary phylogenetic X-trees, where $|X| = 4$. Then it is easily checked that each of the three 2-element subsets of $\mathcal{P}'$ are 1-cycle compatible. However, the union of the X-splits of the trees in $\mathcal{P}'$ is not

circular and so, by the contrapositive of Proposition 2.2, $\mathcal{P}'$ is not 1-cycle compatible.

Theorem 2.1, Proposition 2.3, and Theorem 2.4 provide the basis and validity for the following polynomial-time algorithm for determining the 1-cycle compatibility of a collection of binary phylogenetic X-trees. We leave the formal details to the reader.

Algorithm: 1-CycleCompatibility$(\mathcal{P}, \mathcal{G})$

Input: A collection $\mathcal{P}$ of binary phylogenetic X-trees.

Output: A unicyclic network $\mathcal{G}$ on X that displays $\mathcal{P}$ or the statement $\mathcal{P}$ *is not* 1*-cycle compatible*.

1. Choose any two trees $\mathcal{T}_1$ and $\mathcal{T}_2$ in $\mathcal{P}$.

2. Decide whether or not $d_{\mathrm{TBR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$.

  (a) If no, then halt and return $\mathcal{P}$ *is not* 1*-cycle compatible*.

  (b) If yes, then construct a unicyclic network $\mathcal{G}$ on X that displays $\mathcal{T}_1$ and $\mathcal{T}_2$. In the case $d_{\mathrm{SPR}}(\mathcal{T}_1, \mathcal{T}_2) = 1$ and the pruned subtree has at least two leaves, construct all three unicyclic networks $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ on X that display $\mathcal{T}_1$ and $\mathcal{T}_2$.

3. Select another tree $\mathcal{T}_3 \in \mathcal{P}$.

(a) If exactly one unicyclic network is constructed in the previous step, then check to see whether or not $\mathcal{G}$ displays $\mathcal{T}_3$. If not, then halt and return $\mathcal{P}$ *is not* 1*-cycle compatible*.

(b) If three unicyclic networks are constructed in the previous step, then check to see whether or not $\mathcal{G}_1$, $\mathcal{G}_2$, or $\mathcal{G}_3$ displays $\mathcal{T}_3$. (At most one such tree has this property.) If not, then halt and return $\mathcal{P}$ *is not* 1*-cycle compatible*.

4. Let $\mathcal{G}$ denote the unicyclic network that displays $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$. For each $\mathcal{T}_i \in \mathcal{P} - \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$, check to see whether or not $\mathcal{G}$ displays $\mathcal{T}_i$. If not, then halt and return $\mathcal{P}$ *is not* 1*-cycle compatible*. Otherwise return $\mathcal{G}$.

3. Counting Unicyclic Networks

In this section, we use generating functions to derive the following exact expressions for the number of distinct unicyclic networks on a fixed set X.

Theorem 3.1. *Let X be a finite set of size $n \geq 3$.*

(i) *Let $c(n)$ denote the number of unicyclic networks on X. Then*

$$c(n) = (n-1)!2^{n-2} - \frac{(2n-2)!}{(n-1)!2^{n-1}}.$$

(ii) *For each $k \geq 3$, let $c(n,k)$ denote the number of unicyclic networks on X whose unique cycle is of length k. Then*

$$c(n,k) = \frac{(2n-k-1)!}{(n-k)!2^{n-k+1}}.$$

In proving Theorem 3.1, we make use of the following notation: for a power series $f(x)$, we let $[x^n]f(x)$ denote the coefficient of x^n in $f(x)$.

For $|X| \geq 2$, a *rooted binary phylogenetic X-tree* is a rooted tree whose root has degree two and every other interior vertex has degree three, and whose leaf set is X. If $|X| = 1$, then the tree consisting of a single-root vertex labelled by the element in X is a rooted binary phylogenetic X-tree. For all $n \geq 1$, let $r(n)$ denote the number of rooted binary phylogenetic trees on a set X of size n. For each $n \geq 2$, the number $r(n)$ is given by

$$r(n) = \frac{(2n-2)!}{(n-1)!2^{n-1}} = 1 \times 3 \times \cdots \times (2n-3), \tag{1}$$

a well-known result that dates back to 1870 [19].

For establishing Theorem 3.1, it will be convenient for us to consider one particular way in which $r(n)$ can be derived. Let

$$R(x) = \sum_{n \geq 1} r(n)\frac{x^n}{n!}$$

be the exponential generating function for $r(n)$. Now notice that if we delete the root of a binary phylogenetic tree that has $n \geq 2$ leaves, along

with its two incident edges, we obtain an unordered pair of rooted phylogenetic binary trees for which the numbers of labelled leaves in the resulting pair of trees sum to n. Since the labels can be distributed freely between these two trees, it follows that, for all $n \geq 2$,

$$r(n) = \frac{1}{2}\sum_{i=1}^{n-1} \binom{n}{i} r(i)r(n-i).$$

This expression for $r(n)$ translates into the more succinct equation

$$R(x) = \frac{1}{2}R(x)^2 + x. \tag{2}$$

The term "$+x$" in (2) accounts for the case where we have just a single isolated root vertex. If we regard (2) as a quadratic equation (in $R(x)$), and choose the root whose power series has non-negative coefficients, we get

$$R(x) = 1 - \sqrt{1-2x}. \tag{3}$$

Now, for all $n \geq 1$,

$$[x^n](1 - \sqrt{1-2x}) = \frac{(2n-2)!}{n!(n-1)!2^{n-1}}.$$

Therefore, as $r(n) = n![x^n]R(x)$, we obtain (1).

We now introduce two further exponential generating functions. Let

$$C(x) = \sum_{n\geq 3} c(n)\frac{x^n}{n!}$$

and, for all $k \geq 3$, let

$$C_k(x) = \sum_{n \geq 3} c(n,k) \frac{x^n}{n!}$$

denote the exponential generating functions for $c(n)$ and $c(n,k)$, respectively, where $n \geq 3$. Both these generating functions are closely related to $R(x)$. In particular,

$$c(n,k) = \frac{1}{2k} \sum_{(n_1,\ldots,n_k):n_1+\cdots+n_k=n} \frac{n!}{n_1! \cdots n_k!} \prod_{i=1}^{k} r(n_i). \tag{4}$$

To justify the right-hand side of (4), first note that the term

$$\frac{n!}{n_1! \cdots n_k!}$$

counts the number of k–tuples of sets of sizes $n_1, \ldots, n_k$ that form a partition of the set X (of size n), and the term $\prod_{i=1}^{k} r(n_i)$ is the number of choices of rooted binary phylogenetic trees that have specified leaf sets of sizes $n_1, \ldots, n_k$ where, for each i, $n_i \geq 1$. However, each unicyclic network with cycle length k generates exactly $2k$ such k-tuples of rooted binary phylogenetic trees, since we have k choices for which tree starts the cycle, and there are two directions that the cycle can be traversed. Equation 4 means that we may write $C_k(x)$ much more elegantly as

$$2C_k(x) = \frac{1}{k} R(x)^k. \tag{5}$$

Since $C(x) = \sum_{k\geq 3} C_k(x)$, it follows by (5) that the following relationship between $C(x)$ and $R(x)$ holds:

$$2C(x) = \frac{1}{3}R(x)^3 + \frac{1}{4}R(x)^4 + \cdots \tag{6}$$

Using the identity

$$-\log(1-t) = t + \frac{1}{2}t^2 + \frac{1}{3}t^3 + \cdots,$$

we can rewrite (6) as

$$C(x) = \frac{1}{2}\left(-R(x) - \frac{1}{2}R(x)^2 - \log(1-R(x))\right). \tag{7}$$

Replacing the term $\log(1-R(x))$ in (7) by $\log(\sqrt{1-2x})(=\frac{1}{2}\log(1-2x))$ as allowed by (3), and then the remaining term in (7), namely $-R(x)-\frac{1}{2}R(x)^2$, by $x-2R(x)$ as allowed by (2), we get

$$C(x) = \frac{1}{2}x - R(x) - \frac{1}{4}\log(1-2x).$$

The expression for $c(n)$ in the statement of Theorem 3.1 now follows by routine manipulation. This establishes part (i).

To prove part (ii), we first evaluate $[x^n]R(x)^k$. Notice that one can write $R(x) = x\phi(R(x))$ for the function $\phi(x) = (1-\frac{1}{2}x)^{-1}$. In such a situation, there is a convenient tool for extracting $[x^n]R(x)^k$ known as the *Lagrange inversion formula*. This formula (see [5] for details) states the following: Given two (formal) power series $\psi(x) = \sum_{i\geq 0} c_i\lambda^i$ where $c_0 \neq 0$

and $f(\lambda) = \sum_{i \geq 0} d_i \lambda^i$, there exists a unique power series $w(t)$ such that $w(t) = t\psi(w(t))$ and, for each $n > 0$,

$$[t^n] f(w(t)) = \frac{1}{n}[\lambda^{n-1}] f'(\lambda)\psi^n(\lambda),$$

where $f'(\lambda) = \sum_{i \geq 1} i d_i \lambda^{i-1}$ denotes the formal derivative of f. Applying this formula here (as was similarly applied in [3]), we obtain

$$[x^n]R(x)^k = \frac{1}{n}[\lambda^{n-1}]k\lambda^{k-1}\phi(\lambda)^n = \frac{k}{n}[\lambda^{n-k}](1-\frac{1}{2}\lambda)^{-n} = \frac{k}{n}\binom{2n-k-1}{n-k}2^{k-n}.$$

Therefore, by (5),

$$c(n,k) = n! \cdot \frac{1}{2k}[x^n]R(x)^k = \frac{(2n-k-1)!}{(n-k)!2^{n-k+1}}.$$

This establishes part (ii).

We end this section with the following consequence of Theorem 3.1 for which we recall the definition of a circular ordering of a unicyclic network from the introduction.

Corollary 3.2. *Let X be a finite set of size $n \geq 3$.*

(i) *Let $\mathcal{G}$ be a unicyclic network on X whose unique cycle has length k. Then the number of distinct circular orderings for $\mathcal{G}$ is 2^{n-k+1}.*

(ii) *Let π be a cyclic permutation of X. Then the number of unicyclic networks on X whose cycle has length k and for which π is a circular*

ordering is

$$\binom{2n-k-1}{n-1}$$

Proof. To prove (i), we first note that a binary phylogenetic tree with m leaves, where $m \geq 3$, has precisely 2^{m-2} circular orderings (see, for example, [20]). Now let $m_1, m_2, \ldots m_k$ denote the number of elements of X that appear (as leaves) on the k subtrees that are incident with the k vertices of the cycle in the unicyclic network $\mathcal{G}$. Then, as the cycle of $\mathcal{G}$ can be traversed in two directions, it is now straightforward to see that the number of circular orderings for $\mathcal{G}$ is

$$2\prod_{i=1}^{k} 2^{(m_i+1)-2} = 2^{n-k+1}.$$

Note that replacing m_i by m_i+1 in the exponent recognizes that the subtree that has m_i leaves from X can be viewed as a binary tree with m_i+1 leaves in total if we include the vertex on the cycle that the subtree attaches to. This establishes (i).

For the proof of (ii), let $c(n, k, \pi)$ denote the number of unicyclic networks on X whose unique cycles each have length k and for which π is a circular ordering. To evaluate $c(n, k, \pi)$, we will count the number of ordered pairs $(\mathcal{G}, \pi)$, where $\mathcal{G}$ is unicyclic network on X whose unique cycle has length k and π is a circular ordering for $\mathcal{G}$. We do this count in two ways. Firstly,

by Theorem 3.1(ii), there are

$$\frac{(2n-k-1)!}{(n-k)!2^{n-k+1}}$$

unicyclic networks whose unique cycle has length k. Furthermore, for each such network, there are precisely 2^{n-k+1} circular orderings, by part (i). Hence the number of ordered pairs $(\mathcal{G}, \pi)$ is

$$\frac{(2n-k-1)!}{(n-k)!}$$

Alternatively, we can calculate this number by noting that the number of cyclic permutations on X is $(n-1)!$ and, for every such cyclic permutation π, the number of unicyclic networks on X whose unique cycle has length k and for which π is a circular ordering is $c(n,k,\pi)$. Equating these two counts, we deduce (ii). □

4. Counting Galled-Trees

In this section, we extend Theorem 3.1(ii) to networks that contain k 'independent' cycles. The following definition is motivated by the terminology of [6] and [16] in the rooted digraph setting.

A (unrooted binary) *galled-tree (on X)* is a graph $\mathcal{G}$ that has the following properties:

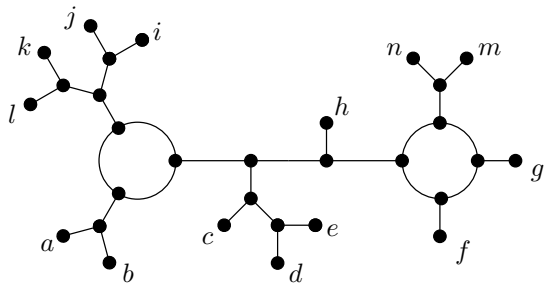

FIGURE 7. A galled-tree with two cycles.

(i) every vertex is in at most one cycle,

(ii) every non-leaf vertex has degree three, and

(iii) the set of degree-one vertices is X.

For example, Fig. 7 shows a galled-tree with two cycles.

The purpose of this section is to establish Theorem 4.1.

Theorem 4.1. *For a fixed finite set X of size n, let $g(n,k,m)$ denote the number of galled-trees on X containing k cycles and having a total of m edges across all the cycles. Then, for $n,m,k \geq 0$, we have*

$$g(n+2,k,m) = \frac{(2n-m+3k)!(m-2k-1)!2^{m-n-3k}}{(n-m+2k)!(m-3k)!(k-1)!k!}$$

if $3 \leq 3k \leq m \leq n+2k$ or $k=m=0$, and $g(n+2,k,m)=0$ otherwise.

Proof. First note that in order for $g(n+2,k,m)$ to be non-zero we must have that if $k=0$, then $m=0$. Furthermore, if $k>0$ then we require

that $m \geq 3k$ (since every cycle has at least three edges), and the inequality $m \leq n + 2k$ must also apply (by a simple counting argument).

Let

$$G = G(x, y, z) = \sum_{n,m,k \geq 0} g(n+1, k, m) \frac{x^n y^k z^m}{n!}.$$

Thus

$$G = x + \frac{1}{2!}x^2 + \frac{1}{2!}x^2yz^3 + \frac{3}{3!}x^3 + \frac{3}{3!}x^3yz^3 + \frac{3}{3!}x^3yz^4 + \frac{3}{3!}x^3y^2z^6 + \frac{15}{4!}x^4 + \cdots$$

Notice that

$$g(n+1, k, m) = n![x^n y^k z^m]G(x, y, z) \tag{8}$$

where $[x^n y^k z^m]G(x, y, z)$ denotes the coefficient of $x^n y^k z^m$ in G.

Given a galled-tree on X, we say that the rooted graph obtained by subdividing any edge of the network, and distinguishing the resulting degree 2 vertex as a root is a *rooted galled-tree network* on X. In this way, we we may regard $g(n+1, k, m)$ as counting the number of rooted galled-trees on X that have k cycle and m edges across all cycles (since there is a bijection between unrooted binary galled-trees on $X \cup \{\rho\}$ (where ρ is a label not in X), and rooted binary galled-trees on X).

This rooting leads to the following fundamental recursion for G:

$$G = x + \frac{1}{2}G^2 + \frac{1}{2}yz^3G^2(1 - zG)^{-1}. \tag{9}$$

(the term $\frac{1}{2}G^2$ counts the cases where the root of the rooted galled-tree does not lie on a cycle, while the term $\frac{1}{2}yz^3G^2(1-zG)^{-1}$ counts the other cases).

From (9) it follows that $G = x\phi(G, y, z)$ where

$$\phi(G, y, z) = (1 - \frac{1}{2}G(1 + \frac{yz^3}{(1 - zG)}))^{-1}.$$

Again applying the Lagrange inversion formula, this time to (9), we have

$$[x^n y^k z^m]G(x, y, z) = \frac{1}{n}[\lambda^{n-1} y^k z^m]\phi(\lambda, y, z)^n. \tag{10}$$

Now, applying the identity:

$$(1 - \theta)^{-n} = \sum_{i \geq 0} \binom{n + i - 1}{i} \theta^i \tag{11}$$

to $\theta = \frac{1}{2}\lambda(1 + \frac{yz^3}{(1-z\lambda)})$ we obtain $\phi(\lambda, y, z)^n = \sum_{i \geq 0} 2^{-i}\binom{n+i-1}{i}\lambda^i(1 + \frac{yz^3}{(1-z\lambda)})^i$. Thus,

(12)

$$[\lambda^{n-1} y^k z^m]\phi(\lambda, y, z)^n = \sum_{i \geq 0} 2^{-i}\binom{n + i - 1}{i}[\lambda^{n-i-1} y^k z^m](1 + \frac{yz^3}{(1 - z\lambda)})^i.$$

Now,

$$[y^k](1 + \frac{yz^3}{(1 - z\lambda)})^i = \binom{i}{k} z^{3k}(1 - z\lambda)^{-k},$$

and

$$[\lambda^{n-i-1} z^m]z^{3k}(1 - z\lambda)^{-k} = [\lambda^{n-i-1} z^{m-3k}](1 - z\lambda)^{-k}.$$

Furthermore, again invoking (11) we have

$$[\lambda^{n-i-1}z^{m-3k}](1-z\lambda)^{-k} = \begin{cases} \binom{m-2k-1}{m-3k}, & \text{if } n-i-1 = m-3k; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the only non-zero term in (12) occurs when $i = n-m+3k-1$, and for this value of i we have

$$[\lambda^{n-1}y^k z^m]\phi(\lambda, y, z)^n = 2^{-i}\binom{n+i-1}{i}\binom{i}{k}\binom{m-2k-1}{m-3k}.$$

Substituting this expression into (10) and (8), together with some routine algebra, we obtain the result described. □

Note that setting $k = 1$ gives the expression

$$g(n+2, 1, m) = \frac{(2n-m+3)!}{(n-m+2)!2^{n-m+3}}$$

in Theorem 3.1(ii).

5. Acknowledgments

We would like to thank the referees for several helpful suggestions.

References

[1] M. C. Baroni, "Hybrid phylogenies: a graph-based approach to represent reticulate evolution", PhD thesis, University of Canterbury, New Zealand, 2004

[2] M. C. Baroni, C. Semple, and M. Steel, "A framework for representing reticulate evolution", *Ann. Combin.*, vol. 8, pp. 391–408, 2004.

[3] M. Carter, M. D. Hendy, D. Penny, L. A. Székely, and N. C. Wormald, "On the distribution of lengths of evolutionary trees", *SIAM J. Discrete Math.*, vol. 3, pp. 38–47, 1990.

[4] J. Felsenstein, "Inferring Phylogenies", Sinauer Press, 2004.

[5] I. P. Goulden and D. M. Jackson, "Combinatorial Enumeration", John Wiley and Sons, New York, 1983.

[6] D. Gusfield, S. Eddhu, C. Langley. "Optimal, efficient reconstruction of phylogenetic networks with constrained recombination", *J. Bioinf. Comput. Biol.*, vol. 2, pp. 173–213, 2004.

[7] B. Holland, K. Huber, V. Moulton and P. J. Lockhart, "Using consensus networks to visualize contradictory evidence for species phylogeny", *Mol. Biol. Evol.*, vol. 21, pp. 1459–1461, 2004.

[8] D. H. Huson, T. Dezulian, T. Kloepper, and M. A. Steel, "Phylogenetic supernetworks from partial trees," *IEEE Trans. Comput. Biol. Bioinf.*, vol. 1, pp. 151–158, 2004

[9] D. H. Huson, T. Kloepper, P. J. Lockhart, M. A. Steel, "Reconstruction of reticulate networks from gene trees", *Proceedings of RECOMB 2005*, S. Miyano et al. (Eds.) LNBI 3500, pp. 233–249, Springer-Verlag Berlin Heidelberg.

[10] T.N.D. Huydn, J. Jansson, N.B. Nguyen and W.-K.Sung, 2005."Constructing a smallest refinining galled phylogenetic network," *Proceedings of RECOMB 2005*, S. Miyano et al. (Eds.) LNBI 3500, pp. 265–280, Springer-Verlag Berlin Heidelberg.

[11] J. Jansson and W-K. Sung, "The maximum agreement of two nested phylogenetic networks," *Proceedings of ISAAC 2004*, R. Fleischer and G. Trippen, eds., LNCS 3341, pp. 581–593, 2004.

[12] P. Legendre, "Biological applications of reticulate analysis", *J. Classification*, vol. 17, pp. 191–195, 2000.

[13] J. O. McInerney and M. Wilkinson, "New methods ring changes for the tree of life", *Trends Ecol. Evol.*, vol. 20, pp. 105–107, 2005.

[14] J. Mallet,"Hybridization as an invasion of the genome", *Trends Ecol. Evol.*. vol. 20, pp. 229–237, 2005.

[15] B. M. E. Moret, L. Nakhleh, T.Warnow, C.R. Linder, A.Tholse, A. Padolina, J. Sun and R. Timme, "Phylogenetic networks: modeling, reconstructibility, and accuracy", *IEEE/ACM Trans. Comput. Biol. Bioinf.* vol. 1, pp. 1–11, 2004.

[16] L. Nakhleh, T. Warnow, and C. Randal Linder, "Reconstructing reticulate evolution in species - theory and practice", *Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, pp. 337—346, 2004.

[17] M. C. Rivera and J. C. Lake, "The ring of life provides evidence for a genome fusion origin of eukaryotes", *Nature*, vol. 431, pp. 152–155, 2004.

[18] F. J. Rohlf, "Phylogenetic models and reticulations", *J. of Classification*, vol. 17, pp. 185–189, 2000.

[19] E. Schröder, "Vier combinatorische Probleme", *Zeitschrift für Mathematik und Physik*, vol. 15, pp. 361–376, 1870.

[20] C. Semple and M. Steel, "Phylogenetics", Oxford University Press, 2003.

[21] C. Semple and M. Steel, "Cyclic permutations and evolutionary trees", *Adv. in Appl. Math.*, vol. 32, pp. 669–680, 2004.

[22] Y. Song, and J. Hein, "On the minimum number of recombination events in the evolutionary history of DNA sequences", *J. Math. Biol.*, vol. 48, pp. 160–186, 2003.

[23] L. Wang, K. Zhang, and L. Zhang, "Perfect phylogenetic networks with recombination", *Journal of Computational Biology*, vol. 8, pp. 69–78, 2001.

Charles Semple is a Senior Lecturer in the Department of Mathematics and Statistics at the University of Canterbury. After receiving his BSc(Hons) from Massey University and teaching for five years, he returned to university and received his PhD in mathematics from Victoria University of Wellington in 1998. Initially a Postdoctoral Fellow, he has been a permanent staff member at the University of Canterbury since 2001. Other academic positions include Visiting Research Fellow at Merton College, University of Oxford (2003) and Visiting Professor at the Université of Montpellier II (2005). His main research interests are matroid theory and phylogenetics.

Mike Steel studied mathematics at Canterbury and Massey Universities (New Zealand) and received the PhD degree in 1989. From 1990-1993, he held various postdoctoral positions in Germany, the United States, and

New Zealand and was appointed to a tenured position at University of Canterbury in 1994. He is currently a professor and director of the Biomathematics Research Centre at the University of Canterbury and is a principal investigator in the Allan Wilson Centre for Molecular Ecology and Evolution.

Biomathematics Research Centre, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

E-mail address: c.semple@math.canterbury.ac.nz, m.steel@math.canterbury.ac.nz