Tree Reconstruction via a Closure Operation on Partial Splits ⋆

Charles Semple and Mike Steel

Biomathematics Research Centre
Department of Mathematics and Statistics
University of Canterbury
Private Bag 4800
Christchurch, New Zealand
c.semple@math.canterbury.ac.nz, m.steel@math.canterbury.ac.nz

Abstract. A fundamental problem in biological classification is the reconstruction of phylogenetic trees for a set X of species from a collection of either subtrees or qualitative characters. This task is equivalent to tree reconstruction from a set of partial X–splits (bipartitions of subsets of X). In this paper, we define and analyse a "closure" operation for partial X–splits that was informally proposed by Meacham [5]. In particular, we establish a sufficient condition for such an operation to reconstruct a tree when there is essentially only one tree that displays the partial X–splits. This result exploits a recent combinatorial result from [2].

1 Introduction

Trees that have some vertices labelled by elements from a finite set X are often used to represent evolutionary relationships, particularly in biology. Two closely related problems are

(i) determining how to combine such trees that classify overlapping subsets of X into a parent tree that displays each of the input trees, and
(ii) determining how to reconstruct a parent tree from (partial) qualitative characters (equivalently, partitions of X or subsets of X) so that each character could have evolved on the parent tree without any reverse or convergent transitions (this is equivalent to each tree displaying the partition associated with each character).

For either problem, a parent tree may not exist, and even deciding this turns out to be an NP-complete problem [3, 6]. However, in certain cases, various efficient "rules" for extending sets of trees or sets of characters can either determine

⋆ This work was supported by the New Zealand Marsden Fund (UOC-MIS-003).

that such a tree does not exist, or reconstruct the tree when there is essentially only one possible parent tree. Conditions for such an approach to succeed using extension rules on sets of trees were recently established in [1], using a combinatorial result from [2]. In this paper we take a related, but different, approach by considering a rule for extending sets of partial binary characters (called partial X–splits below) that was proposed informally by Meacham [5]. We formalise an iterative construction using this rule, and show that it always leads to the same set of partial X–splits, regardless of the possible choices by which the rule can be applied. Then, using the main combinatorial result from [2], we provide sufficient conditions for this construction to successfully recover a parent tree or determine that no such tree exists. Note that although the input to our tree reconstruction problem consists of partial X–splits, it could easily be modified to input partitions of subsets of X (in the case of problem (ii)) or trees classifying overlapping subsets of X (in the case of problem (i)) since all these problems are essentially equivalent [6].

2 Preliminaries

Throughout this paper, X denotes a finite set. We begin with some definitions.

Partial splits. A *partial split* of X, or more briefly a *partial X–split*, is a partition of a subset of X into two disjoint non-empty subsets. If these two subsets are A and B, we denote the partial split by $A|B$. Note that no distinction is made between $A|B$ and $B|A$. If $A \cup B = X$ we say that $A|B$ is a *(full) X–split.* We write $aa'|bb'$ to denote the partial split $A|B$ if $A = \{a, a'\}$ and $B = \{b, b'\}$, and we call this a *quartet X–split.* We say that the partial split $A'|B'$ *extends* the partial split $A|B$ precisely if either $A \subseteq A'$ and $B \subseteq B'$ or $A \subseteq B'$ and $B \subseteq A'$. A partial X–split $A|B$ is *trivial* if $\min\{|A|, |B|\} = 1$.

X–trees. Let T be a tree with vertex set V and edge set E, and suppose we have a map $\phi : X \to V$ with the property that, for all $v \in V$ with degree at most two, $v \in \phi(X)$. Then the ordered pair $(T; \phi)$, which we frequently denote by $\mathcal{T}$, is called an *X–tree.* Two X–trees $(T_1; \phi_1)$ and $(T_2; \phi_2)$, where $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$, are regarded as equivalent if there exists a bijection $\psi : V_1 \to V_2$ which induces a bijection between E_1 and E_2 and satisfies $\phi_2 = \psi \circ \phi_1$, in which case, ψ is unique.

Let $\mathcal{T} = (T; \phi)$ be an X–tree and let e be an edge of T. Then *corresponding* to e is the X–split $\phi^{-1}(V_1)|\phi^{-1}(V_2)$, where V_1 and V_2 denote the vertex sets of the two components obtained from T by deleting e. For an X–tree $\mathcal{T}$, let $\Sigma(\mathcal{T})$ (resp. $\Sigma^*(\mathcal{T})$) denote the collection of non-trivial X–splits (resp. all X–splits) corresponding to the edges of T.

Compatibility. Let $A|B$ be a partial X–split. An X–tree $\mathcal{T} = (T; \phi)$ *displays* $A|B$ if there is an edge e of $T = (V, E)$ such that, in $(V, E - \{e\})$, the sets $\phi(A)$ and $\phi(B)$ are subsets of the vertex sets of different components. For example, the X-tree shown in Figure 1, where $X = \{1, 2, \ldots, 7\}$, displays each of the partial X–splits in $\{\{1, 2\}|\{3, 4\}, \{2, 3\}|\{4, 7\}, \{1, 7\}|\{4, 5\}, \{2, 5\}|\{6, 7\}\}$. A collection Σ of partial X–splits is said to be *compatible* if there exists an X–tree that displays every X–split in Σ. This is equivalent to requiring that every non-trivial split in Σ is extended by a split in $\Sigma(\mathcal{T})$.

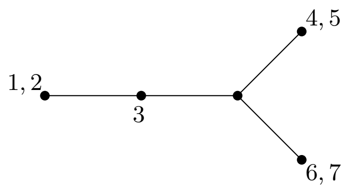

Fig. 1. An X–tree displaying $\{1, 2\}|\{3, 4\}$, $\{2, 3\}|\{4, 7\}$, $\{1, 7\}|\{4, 5\}$, and $\{2, 5\}|\{6, 7\}\}$.

The following result is well known, and follows immediately from results in [4].

Lemma 1. *Let $A_1|B_1$ and $A_2|B_2$ be partial X–splits. The following statements are equivalent:*

(i) *$A_1|B_1$ and $A_2|B_2$ are compatible.*
(ii) *At least one of the sets $A_1 \cap A_2$, $A_1 \cap B_2$, $B_1 \cap A_2$, and $B_1 \cap B_2$ is empty.*

A set Σ of partial X–splits is said to be *pairwise compatible* if each pair of splits in Σ is compatible. This condition is not sufficient for Σ to be compatible. For example, for $X = \{a, b, c, d, e\}$, the set $\{ab|cd, ab|ce, ad|be\}$ of partial X–splits of X is pairwise compatible, but not compatible. However, if Σ consists of full X-splits, then Σ is compatible precisely if Σ is pairwise compatible, in which case there is a unique X–tree $\mathcal{T}$ such that $\Sigma^*(\mathcal{T}) = \Sigma$ (see [4]).

Irreducible sets of partial X–splits. Let Σ be a set of partial splits of X. A partial split $A|B \in \Sigma$ is *redundant* if there exists a different partial split in Σ that extends $A|B$. If Σ has no redundant splits, then Σ is said to be *irreducible*. Let Σ_1 and Σ_2 be two irreducible sets of partial splits of X. We write $\Sigma_1 \preceq \Sigma_2$ if, for each $A_1|B_1 \in \Sigma_1$, there is an element $A_2|B_2$ in Σ_2 that extends $A_1|B_1$. It is not difficult to show that $\preceq$ is a partial order on the collection of irreducible sets of partial X–splits. Note that if we drop the irreducibility condition, then $\preceq$ may fail to satisfy the antisymmetric property ($a \preceq b$ and $b \preceq a$ implies $a = b$) required of a partial order. Observe that an X–tree $\mathcal{T}$ displays a set Σ of partial X–splits precisely if $\Sigma \preceq \Sigma^*(\mathcal{T})$.

We will let $\mathcal{P}(X)$ denote the collection of all sets Σ of partial X–splits that are both pairwise compatible and irreducible. Let us adjoin to $\mathcal{P}(X)$ a new element ω, and let $\mathcal{P}_\omega(X) = \mathcal{P}(X) \cup \{\omega\}$. If we extend the definition of $\preceq$ by setting $\Sigma \preceq \omega$ for all $\Sigma \in \mathcal{P}(X)$ (so that ω acts as a maximal element), then $\preceq$ is a partial order on $\mathcal{P}_\omega(X)$.

3 Split Closure

In this section, we define a "split closure" of a set Σ of partial X–splits. Informally, we construct an irreducible set of partial X–splits from Σ by repeatedly applying a pairwise replacement rule along the lines suggested by Meacham [5] (the replacement rule (SC) below corresponds to "Rule 2" in [5]). We then show that any two split closures of Σ are equal.

The replacement rule we consider for an irreducible set Σ of partial X–splits is the following:

(SC) If $A_1|B_1$ and $A_2|B_2$ are elements of Σ that satisfy

$$\emptyset \notin \{A_1 \cap A_2, A_1 \cap B_2, B_1 \cap B_2\} \text{ and } B_1 \cap A_2 = \emptyset, \tag{1}$$

replace $A_1|B_1$ and $A_2|B_2$ in Σ by $(A_1 \cup A_2)|B_1$ and $A_2|(B_1 \cup B_2)$, and then remove any redundant partial splits from the newly created set.

If $A_1|B_1$ and $A_2|B_2$ in the statement of (SC) have the property that $A_2 \subseteq A_1$, $B_1 \subseteq B_2$, and (1) applies, then $B_1 \cap A_2$ is empty, and the two newly created partial splits are $(A_1 \cup A_2)|B_1$ and $A_2|(B_1 \cup B_2)$, which are identical to $A_1|B_1$ and $A_2|B_2$, respectively. We call such an application of (SC) *trivial*; in all other (*non-trivial*) applications of (SC) at least one of the newly created partial splits differs from $A_1|B_1$ or $A_2|B_2$.

We say that $\Sigma \in \mathcal{P}(X)$ is *closed* under (SC) if (SC) applies only trivially to Σ.

The motivation for (SC) is the following result due to Meacham [5].

Lemma 2. *Let Σ be a set of partial X–splits, and let Σ' be a set of partial X–splits obtained from Σ by a single application of* (SC). *Then an X–tree $\mathcal{T}$ displays Σ if and only if $\mathcal{T}$ displays Σ'.*

Let Σ be a set of irreducible partial X–splits, and suppose that we construct a sequence

$$\Sigma_0, \Sigma_1, \ldots, \Sigma_i, \Sigma_{i+1}, \ldots$$

of irreducible partial X–splits such that $\Sigma_0 = \Sigma$ and, for all $i \geq 0$, Σ_{i+1} is obtained from Σ_i by one non-trivial application of (SC) provided Σ_i is pairwise compatible. Since $\Sigma_i \preceq \Sigma_{i+1}$, for all $i \leq 0$, it follows that this sequence is strictly increasing under $\preceq$. Consequently, since the set of all X–splits is finite, this sequence must terminate with a set, Σ_n say, of irreducible partial X–splits such that either Σ_n is pairwise compatible and closed under (SC), or Σ_n is not pairwise compatible. If the latter holds, we reset Σ_n to be the element ω.

Definition. We refer to the sequence $\Sigma_0, \Sigma_1, \ldots, \Sigma_n$ as a *split closure sequence* for Σ, and the terminal value Σ_n as a *split closure* of Σ. Note that (SC) applies only trivially to Σ_n (when $\Sigma_n \neq \omega$) and Σ_n is always an upper bound, under $\preceq$, to Σ.

We next provide an explicit bound on the length of any split closure sequence.

Lemma 3. *Let Σ be a set of irreducible partial X–splits, and let $\Sigma_0, \Sigma_1, \ldots, \Sigma_n$ be a split closure sequence for Σ. Then $n \leq |\Sigma| \times |X| - \sum_{A|B \in \Sigma} |A \cup B|$.*

Proof. It is straightforward to see that we can prove the lemma by making the additional assumption that $\Sigma_n \neq \omega$. For all $i \in \{0, 1, \ldots, n-1\}$, let $\lambda_i : \Sigma_i \to \Sigma_{i+1}$ be a function that maps an element, $A'|B'$ say, of Σ_i to an element of Σ_{i+1} that extends $A'|B'$. Furthermore, for each element, $A|B$ say, of Σ and for all $i \in \{0, 1, \ldots, n-1\}$, let $A_{i+1}|B_{i+1} = \lambda_i \lambda_{i-1} \cdots \lambda_0(A|B)$.

Since, for all i, Σ_{i+1} is obtained from Σ_i by a non-trivial application of (SC) and $\Sigma_i \preceq \Sigma_{i+1}$, it follows that

$$\sum_{A|B \in \Sigma} (|A_{i+1} \cup B_{i+1}| - |A_i \cup B_i|) \geq 1,$$

for all i. Consequently,

$$\sum_{A|B \in \Sigma} (|A_n \cup B_n| - |A \cup B|) \geq n.$$

Therefore, as $|A_n \cup B_n| - |A \cup B| \leq |X| - |A \cup B|$ for each element $A|B$ in Σ,

$$n \leq |\Sigma| \times |X| - \sum_{A|B \in \Sigma} |A \cup B|$$

as required. □

It will immediately follow from Lemma 4 that the split closure of a set Σ of irreducible partial X–splits is well-defined.

Lemma 4. *Let Σ be an irreducible set of partial X–splits. Then any two split closures of Σ are equal.*

Proof. If every split closure of Σ is ω, the lemma is (trivially) true, so we may assume that there exists a split closure, $\overline{\Sigma}$ say, of Σ which is not ω. We prove the lemma by showing that every other split closure of Σ equals $\overline{\Sigma}$. To this end, let $\Sigma_0, \Sigma_1, \ldots, \Sigma_n$ be a split closure sequence for Σ, where $\Sigma_0 = \Sigma$. We first claim that, for all $i \in \{0, 1, \ldots, n\}$,

$$\Sigma_i \neq \omega \text{ and } \Sigma_i \preceq \overline{\Sigma}. \tag{2}$$

We establish (2) by induction on i. If $i = 0$, then (2) holds as there exists a split closure of Σ not equal to ω. Now suppose that (2) holds for $i = r$, where $r \in \{0, 1, \ldots, n-1\}$, and Σ_{r+1} is obtained from Σ_r by applying (SC) to the pair $A_1|B_1$ and $A_2|B_2$. Without loss of generality, we may assume that $B_1 \cap A_2 = \emptyset$. By the induction hypothesis, $\Sigma_r \preceq \overline{\Sigma}$, and so there is a pair of partial X–splits $A'_1|B'_1$ and $A'_2|B'_2$ in $\overline{\Sigma}$ such that $A_i \subseteq A'_i$ and $B_i \subseteq B'_i$, for all $i \in \{1, 2\}$. Since $\overline{\Sigma}$ is pairwise compatible, it follows that $A'_1|B'_1$ and $A'_2|B'_2$ satisfy (1). Therefore, as (SC) applies only trivially to $\overline{\Sigma}$, it follows that $A'_2 \subseteq A'_1$ and $B'_1 \subseteq B'_2$. Consequently, $A'_1|B'_1$ and $A'_2|B'_2$ extend $(A_1 \cup A_2)|B_1$ and $A_2|(B_1 \cup B_2)$, respectively, and so

$$\Sigma_r \cup \{(A_1 \cup A_2)|B_1, A_2|(B_1 \cup B_2)\} - \{A_1|B_1, A_2|B_2\} \preceq \overline{\Sigma}.$$

Therefore, as $\overline{\Sigma}$ is pairwise compatible, $\Sigma_{r+1} \neq \omega$ and $\Sigma_{r+1} \preceq \overline{\Sigma}$. This completes the induction step and thereby establishes (2).

Applying (2) to $i = n$, we get $\Sigma_n \neq \omega$ and $\Sigma_n \preceq \overline{\Sigma}$. By interchanging the roles of Σ_n and $\overline{\Sigma}$ in the argument of the last paragraph, we deduce that $\overline{\Sigma} \preceq \Sigma_n$, and hence $\Sigma_n = \overline{\Sigma}$. □

Definition. In view of Lemma 4, we denote the split closure of a set Σ of irreducible partial X–splits by $\mathrm{spcl}(\Sigma)$.

Note that $|\mathrm{spcl}(\Sigma)| \leq |\Sigma|$. Also, provided we set $\mathrm{spcl}(\omega) = \omega$, then spcl satisfies the three properties one expects of a closure operation on the poset $\mathcal{P}_\omega(X)$. Namely, if $a, b \in \mathcal{P}_\omega(X)$ with $a \preceq b$, then

(i) $a \preceq \mathrm{spcl}(a)$,
(ii) $\mathrm{spcl}(a) \preceq \mathrm{spcl}(b)$, and
(iii) $\mathrm{spcl}(\mathrm{spcl}(a)) = \mathrm{spcl}(a)$.

The next result follows immediately from Lemma 2, however, the converse of this corollary is not true [6].

Corollary 1. *Let Σ be a set of irreducible partial X–splits. If* $\mathrm{spcl}(\Sigma) = \omega$, *then Σ is incompatible.*

4 Tree Reconstruction Using Split Closure

In this section, we establish a sufficient condition for the split closure of an irreducible set of partial X–splits to recover all the non-trivial splits of an X–tree. This result, Corollary 2, depends on a combinatorial theorem from [2]. In order to apply this theorem, we need to relate partial splits and split closure to quartet splits and a dyadic closure rule that operates on quartet splits. To this end, we introduce some further definitions.

Definition. For a partial X–split $A|B$, let

$$\mathcal{Q}(A|B) = \{aa'|bb' : \ a, a' \in A;\ b, b' \in B;\ a \neq a';\ \text{and } b \neq b'\}$$

and, for a set Σ of partial X–splits, let

$$\mathcal{Q}(\Sigma) = \bigcup_{A|B \in \Sigma} \mathcal{Q}(A|B).$$

For an X-tree $\mathcal{T}$, we denote $\mathcal{Q}(\Sigma(\mathcal{T}))$ by $\mathcal{Q}(\mathcal{T})$.

Proposition 1 relates partial splits to quartet splits.

Proposition 1. *Let Σ be an irreducible set of non-trivial partial X–splits and let $\mathcal{T}$ be an X–tree. Then $\Sigma = \Sigma(\mathcal{T})$ if and only if the following two conditions hold:*

(i) $|\Sigma| \leq |\Sigma(\mathcal{T})|$*; and*
(ii) $\mathcal{Q}(\Sigma) = \mathcal{Q}(\mathcal{T})$.

Proof. Evidently, if $\Sigma = \Sigma(\mathcal{T})$, then (i) and (ii) hold. For the converse, we first show that $\Sigma \preceq \Sigma(\mathcal{T})$. Let $\mathcal{T} = (T; \phi)$, and let $A|B$ be an element of Σ. By (ii), $\mathcal{Q}(A|B) \subseteq \mathcal{Q}(\mathcal{T})$. Therefore, for each quartet of elements a, a', b, and b' with $a, a' \in A$ and $b, b' \in B$, the cardinality, denoted $n(a, a', b, b')$, of $\{A'|B' \in \Sigma(\mathcal{T}) : aa'|bb' \in \mathcal{Q}(A'|B')\}$ satisfies $n(a, a', b, b') \geq 1$. Now suppose that a, a', b, and b' are chosen so that $n(a, a', b, b')$ is minimised, and $A'|B'$ is an element of $\Sigma(\mathcal{T})$ with $aa'|bb' \in \mathcal{Q}(A'|B')$. By considering the placement of the vertices $\phi(a)$, $\phi(a')$, $\phi(b)$, and $\phi(b')$ in T, we see that $A \subseteq A'$ and $B \subseteq B'$, thus showing that $\Sigma \preceq \Sigma(\mathcal{T})$.

Now let $n(A|B) = \min\{n(a, a', b, b') : a, a' \in A; b, b' \in B\}$, and let

$$\Sigma_1 = \{A|B \in \Sigma : n(A|B) = 1\}.$$

Using the fact that $\Sigma \preceq \Sigma(\mathcal{T})$, it is easily seen that, for each element, $A|B$ say, of Σ_1, there is a unique element of $\Sigma(\mathcal{T})$ that extends $A|B$. Let $\mu : \Sigma_1 \rightarrow \Sigma(\mathcal{T})$

denote the map that associates with each element $A|B$ of Σ_1 the unique element of $\Sigma(\mathcal{T})$ that extends $A|B$. We next show that μ is a bijection.

Let $C'|D'$ be an element of $\Sigma(\mathcal{T})$, and choose elements $c, c' \in C'$ and $d, d' \in D'$ so that $n(c, c', d, d') = 1$. Then, by (ii), there is an element $C|D$ of Σ_1 such that $cc'|dd' \in \mathcal{Q}(C|D)$ and, moreover, $\mu(C|D) = C'|D'$. Thus the map μ is surjective and so $|\Sigma_1| \geq |\Sigma(\mathcal{T})|$. It now follows from (i) that $\Sigma_1 = \Sigma$, and so μ is indeed a bijection. Hence $|\Sigma| = |\Sigma(\mathcal{T})|$. Since $\Sigma_1 = \Sigma$, we can complete the proof by showing that, for each $A|B \in \Sigma_1$, $\mu(A|B) = A|B$.

Suppose, to the contrary, that $\mu(A|B) = A'|B'$, where $A'|B'$ extends $A|B$ but is not equal to $A|B$, for some $A|B \in \Sigma_1$. Then there is an element x in $(A' \cup B') - (A \cup B)$. Without loss of generality, we may assume $x \in A'$. Then we can choose elements $a_1 \in A'$ and $b_1, b_2 \in B'$ so that $n(x, a_1, b_1, b_2) = 1$. By (ii), there is an element $C|D$ of Σ_1 such that $xa_1|b_1b_2 \in \mathcal{Q}(C|D)$. Since $x \notin A \cup B$, $C|D$ is not equal to $A|B$. Therefore, as μ is a bijection, $\mu(C|D) \neq A'|B'$, and so $xa_1|b_1b_2 \notin \mathcal{Q}(\mu(C|D))$. This contradiction completes the proof of Proposition 1. $\square$

Following [1], the *semi-dyadic closure* of a collection $\mathcal{Q}$ of quartet X–splits, denoted $\mathrm{scl}_2(\mathcal{Q})$, is the minimal set of quartet X–splits that contains $\mathcal{Q}$ and is closed under the following rule:

(SDC) If $ab|cd$ and $ac|de$ are elements of $\mathrm{scl}_2(\mathcal{Q})$, then $ab|ce$, $ab|de$, and $bc|de$ are elements of $\mathrm{scl}_2(\mathcal{Q})$.

The next proposition relates split closure to semi-dyadic closure.

Proposition 2. *If Σ is a set of compatible irreducible partial X–splits, then* $\mathrm{scl}_2(\mathcal{Q}(\Sigma)) \subseteq \mathcal{Q}(\mathrm{spcl}(\Sigma))$.

Proof. We can obtain $\mathrm{scl}_2(\mathcal{Q}(\Sigma))$ by constructing a sequence $\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_m$ of collections of quartet X–splits such that $\mathcal{Q}_0 = \mathcal{Q}(\Sigma)$, $\mathcal{Q}_m = \mathrm{scl}_2(\mathcal{Q}(\Sigma))$, and, for all $i \in \{0, 1, \ldots, m-1\}$, $\mathcal{Q}_{i+1} = \mathcal{Q}_i \cup \mathrm{scl}_2(\{q_i, q'_i\})$, where $q_i, q'_i \in \mathcal{Q}_i$ but $\mathrm{scl}_2(\{q_i, q'_i\}) \not\subseteq \mathcal{Q}_i$. We prove the proposition by showing that one can construct a sequence $\Sigma_0, \Sigma_1, \ldots, \Sigma_m$ of sets of irreducible partial X–splits such that $\Sigma_0 = \Sigma$ and, for all $j \in \{0, 1, \ldots, m\}$,

$$\Sigma_j \preceq \mathrm{spcl}(\Sigma) \text{ and } \mathcal{Q}_j \subseteq \mathcal{Q}(\Sigma_j). \tag{3}$$

For then, taking $j = m$, establishes the proposition.

The proof of the latter construction is by induction on j. Clearly, the result holds if $j = 0$ as $\Sigma_0 = \Sigma$. Now let r be an element of $\{0, 1, \ldots, m-1\}$, and suppose that Σ_j has been defined for all $j \leq r$ and (3) holds for $j = r$. Then $q_r, q'_r \in \mathcal{Q}(\Sigma_r)$. If $\mathrm{scl}_2(\{q_r, q'_r\}) \subseteq \mathcal{Q}(\Sigma_r)$, then set $\Sigma_{r+1} = \Sigma_r$. On the other hand, suppose that $\mathrm{scl}_2(\{q_r, q'_r\}) \not\subseteq \mathcal{Q}(\Sigma_r)$. Since $\mathcal{Q}_r \subseteq \mathcal{Q}(\Sigma_r)$, there are two distinct elements $A|B$ and $A'|B'$ in Σ_r that extend q_r and q'_r, respectively. As Σ is compatible, $\mathrm{spcl}(\Sigma)$ is compatible, so Σ_r is pairwise compatible. It now follows that we can apply (SC) to $A|B$ and $A'|B'$. Set Σ_{r+1} to be the resulting set of irreducible partial X–splits. In both cases, $\Sigma_{r+1} \preceq \mathrm{spcl}(\Sigma)$ and, moreover, one can easily check that $\mathrm{scl}_2(\{q_r, q'_r\}) \subseteq \mathcal{Q}(\Sigma_{r+1})$. Hence (3) holds for $j = r + 1$, and so we can indeed construct such a sequence. □

Definition. A set Σ of non-trivial partial X–splits *weakly defines* an X–tree $\mathcal{T}$ if there is a unique X–tree $\mathcal{T}'$ that displays $\Sigma \cup \{\{x\}|X - \{x\} : x \in X\}$, in which case $\Sigma^*(\mathcal{T}) = \Sigma(\mathcal{T}')$.

Let $\mathcal{T} = (T; \phi)$ be an X–tree, and let v be a vertex of T. Suppose that there is a set of non-trivial partial X–splits that weakly defines $\mathcal{T}$. Then it is easily seen that each of the following hold in T:

(i) If v is a pendant vertex, then $|\phi^{-1}(v)| = 2$.
(ii) If v is a degree-two vertex, then $|\phi^{-1}(v)| = 1$.
(iii) If v is neither a pendant vertex nor a degree-two vertex, then v is a degree-three vertex and $\phi^{-1}(v) = \emptyset$.

Conversely, if $\mathcal{T}$ satisfies all of (i)–(iii), then $\Sigma(\mathcal{T})$ weakly defines $\mathcal{T}$. As an example, the set $\{\{1, 2\}|\{3, 4\}, \{2, 3\}|\{4, 7\}, \{1, 7\}|\{4, 5\}, \{2, 5\}|\{6, 7\}\}$ of partial X–splits weakly defines the X–tree in Figure 1.

Two characterisations for when a minimum-sized set of quartet X–splits weakly defines an X–tree are given in [2]. Theorem 1 gives a third such characterisation. Before stating this theorem, we note that it immediately follows from [6, Proposition 6] that $|X| - 3$ is the minimum number of quartet X–splits that can weakly define an X–tree $\mathcal{T}$. Observe that $|X| - 3 = |\Sigma(\mathcal{T})|$.

Theorem 1. *Let Σ_Q be a set of $|X|-3$ quartet X–splits, and let $\mathcal{T}$ be an X–tree. Then the following statements are equivalent:*

(i) *Σ_Q weakly defines $\mathcal{T}$.*
(ii) $\mathrm{spcl}(\Sigma_Q) = \Sigma(\mathcal{T})$.

Proof. If $\mathrm{spcl}(\Sigma_Q) = \Sigma(\mathcal{T})$, then one can easily check using Lemma 2 that Σ_Q weakly defines $\mathcal{T}$. For the converse, suppose that Σ_Q weakly defines $\mathcal{T}$. Then, by [2, Theorem 3.11] (also see [1]), $\mathrm{scl}_2(\Sigma_Q) = \mathcal{Q}(\mathcal{T})$. Since Σ_Q is compatible and irreducible, we can apply Proposition 2 to Σ_Q and get $\mathrm{scl}_2(\mathcal{Q}(\Sigma_Q)) \subseteq$

$\mathcal{Q}(\mathrm{spcl}(\Sigma_Q))$. As $\mathcal{Q}(\Sigma_Q) = \Sigma_Q$, it follows that $\mathcal{Q}(\mathcal{T}) \subseteq \mathcal{Q}(\mathrm{spcl}(\Sigma_Q))$. Now $\mathrm{spcl}(\Sigma_Q)$ is compatible, so $\mathcal{Q}(\mathcal{T}) = \mathcal{Q}(\mathrm{spcl}(\Sigma_Q))$. Moreover, $|\mathrm{spcl}(\Sigma_Q)| \leq |\Sigma_Q| = |\Sigma(\mathcal{T})|$, and so, by Proposition 1, $\mathrm{spcl}(\Sigma_Q) = \Sigma(\mathcal{T})$ as required. □

An immediate consequence of Theorem 1 is Corollary 2.

Corollary 2. *Let Σ_Q be a set of quartet X–splits, and suppose that there exists a subset of Σ_Q of size $|X|-3$ that weakly defines an X–tree $\mathcal{T}$. If Σ_Q is compatible, then* $\mathrm{spcl}(\Sigma_Q) = \Sigma(\mathcal{T})$*; otherwise* $\mathrm{spcl}(\Sigma_Q) = \omega$.

Suppose that Σ_Q and $\mathcal{T}$ satisfy the assumptions of their namesake in the statement of Corollary 2. The potential utility of Corollary 2 lies in the fact that $\mathcal{T}$ can be reconstructed from $\Sigma(\mathcal{T})$ and, in turn, $\Sigma(\mathcal{T}) = \mathrm{spcl}(\Sigma_Q)$ can be reconstructed from Σ_Q; moreover, both tasks can be carried out in polynomial time. Thus we obtain an alternative polynomial-time algorithm for the special case of this tree reconstruction problem to that described in [1]. Furthermore, if $|\Sigma_Q| = O(n)$, then, by Lemma 3, every split closure sequence for Σ_Q has length at most $O(n^2)$, and so the algorithm described here should be reasonably fast.

Acknowledgements

We thank the anonymous referee for their helpful comments.

References

[1] Böcker, S., Bryant, D., Dress, A. W. M., and Steel, M. A.: Algorithmic aspects of tree amalgamation. To appear in *Journal of Algorithms*.

[2] Böcker, S., Dress, A. W. M., and Steel, M.: Patching up X–trees. *Annals of Combinatorics* **3** (1999) 1-12.

[3] Bodlaender, H. L., Fellows, M. R., and Warnow, T. J.: Two strikes against perfect phylogeny. In *Proceedings of the International Colloquium on Automata, Languages and Programming*, Vol. 623 of Lecture Notes in Computer Science, Springer-Verlag, Berlin (1993) 273–283.

[4] Buneman, P.: The recovery of trees from measures of dissimilarity. In F. R. Hodson, D. G. Kendall, and P. Tautu (eds.): *Mathematics in the Archaeological and Historical Sciences*, Edinburgh University Press (1971) 387–395.

[5] Meacham, C. A.: Theoretical and computational considerations of the compatibility of qualitative taxonomic characters. In J. Felsenstein (ed.): *Numerical Taxonomy*, NATO ASI Series Vol. G1, Springer-Verlag (1983) 304–314.

[6] Steel, M.: The complexity of reconstructing trees from qualitative characters and subtrees. *Journal of Classification* **9(1)** (1992) 91–116.