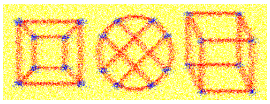
Journal of Graph Algorithms and Applications
http://jgaa.info/ vol. 30, no. 1, pp. 363–383 (2026)
DOI: 10.7155/jgaa.v30i1.3194

A Kernel for the Maximum Agreement Forest Problem on Multiple Binary Phylogenetic Trees

Steven Kelk[1] *Ruben Meuwese*[2] *Leo van Iersel*[2]

[1]Department of Advanced Computing Sciences,
Maastricht University, The Netherlands
[2]Delft Institute of Applied Mathematics,
Delft University of Technology, The Netherlands

Submitted: March 2026 Accepted: July 2026 Published: August 2026

Article type: Regular paper Communicated by: Alexandru I. Tomescu

Abstract. The maximum agreement forest (MAF) problem in phylogenetics takes as input a set of $t \geq 2$ binary phylogenetic trees $\mathcal{T}$ on the same set of taxa X. It asks for a partition of X into the smallest number of blocks such that the subtrees induced by these blocks are disjoint and have common topology across all the trees in $\mathcal{T}$. We produce a modified version of the well-known chain reduction rule in order to prove that after exhaustive application of reduction rules each tree has $O(t \cdot r \cdot k)$ leaves, where k is the natural parameter (the number of blocks) and $r = \min\{\max\{k, 3\}, t+1\}$. We prove this bound for both the unrooted and rooted version of the problem, and demonstrate that the bound r, the length to which common chains are truncated, is tight. Our results constitute the first kernels for MAF in the $t > 2$ regime.

1 Introduction

A phylogenetic tree T on a set of species (or more generally, *taxa*) X is a tree whose leaves are bijectively labelled by X. Such trees are commonplace in the study of the evolution of species; interior nodes represent 'branching' events in history, such as speciation events. Taken as a whole, T therefore represents a hypothesis about how the species X evolved from common ancestors [14]. In this article we are only concerned with *binary* phylogenetic trees, i.e. in which each interior node represents a bifurcation of lineages. These trees come in two variants: rooted, in which the direction of evolution is designated, and unrooted in which the direction of evolution has not yet been determined.

Given a set of $t \geq 2$ phylogenetic trees $\mathcal{T}$, all on the same set X, the *maximum agreement forest* problem (MAF) asks for a partition of X into the smallest number of blocks such that the subtrees

Ruben Meuwese was supported by the Dutch Research Council (NWO), project OCENW.GROOT.2019.015.

E-mail addresses: steven.kelk@maastrichtuniversity.nl (Steven Kelk) R.H.Meuwese@tudelft.nl (Ruben Meuwese) L.J.J.vanIersel@tudelft.nl (Leo van Iersel)

induced by these blocks are disjoint and have common topology across all the trees in $\mathcal{T}$. MAF can be used as a measure of dissimilarity; this can be particularly useful if the tree inference process has constructed a set of competing tree hypotheses, not just one, and we wish to summarize their shared structure.

Unfortunately, the variants of MAF on rooted trees (rMAF) and unrooted trees (uMAF) are both already NP-hard when $t = 2$ [1, 3, 8]. This has led to more than two decades of research aiming to tame this worst-case intractability, particularly in the area of parameterized complexity and approximation algorithms. Rather than list the full history of results in this area we refer to papers with detailed surveys of the literature such as [4, 11, 15, 16].

In this article our focus is on parameterized complexity. For $t = 2$, both kernelization and branching algorithms are known. For uMAF, a branching algorithm with running time $O^*(2.846^k)$ was recently proposed [12], where here k refers to the number of blocks in an optimal solution and '*' suppresses polynomial terms. The same article presents a branching algorithm for rMAF with running time $O^*(2.3391^k)$, very slightly improving upon the previous best of $O^*(2.3431^k)$ [5]. Linear kernels for uMAF and rMAF were established in 2001 and 2004 respectively [1, 3], using two reduction rules known as the *common subtree* and *common chain* reduction rules. The state of the art kernel for rMAF augments this with a third reduction rule and (up to constant terms) achieves a kernel size of $9k$ [10]. For uMAF, augmenting the two original reduction rules with eight new rules also yields a kernel of size $9k$ [11].

For $t > 2$ progress has been comparatively slow. For rMAF, a branching algorithm exists that runs in time $O^*(2.42^k)$ [16], and for uMAF the best known branching algorithm runs in time $O^*(4^k)$[15]. However, to date no explicit kernelization results have been proven[1].

In this article, we fill this gap by showing that the common subtree reduction rule, when combined with a modified version of the common chain reduction rule, reduces the number of leaves in each tree to $O(t \cdot r \cdot k)$ where $r = \min\{\max\{k, 3\}, t + 1\}$ and k is the natural parameter (the number of blocks). This bound holds for both uMAF and rMAF. In proving this we have to forego the 'generator' machinery that has been used in the recent $t = 2$ kernelization literature to produce small kernels: this machinery breaks down for $t > 2$. For this reason the analysis of the size of our kernel bears more resemblance to earlier work in this area, specifically [1, 3]. The dependency on t in the size of the kernel stems from the fact that the modified common chain rule truncates chains to length r, which is a function of t and k; in contrast, in the $t = 2$ kernelization literature truncation to constant length is sufficient. We show that our choice of r is tight, i.e. there exist inputs in which shortening common chains to length less than r alters the size of an optimum solution. Since, for $t = 2$, our bounds perfectly match the known bounds under subtree and chain reduction, another contribution of this article is to provide a simpler proof of these kernelization results.

Our kernel might be of interest to participants of the Parameterized Algorithms and Computational Experiments competition 2026 (PACE 2026), since the exact track of this competition asks participants to solve rMAF instances on potentially multiple trees [13].

We conclude with a discussion of open problems and future research directions. Most notably, do rMAF and uMAF admit kernels whose size is at most *polynomial* in *only* k, as opposed to polynomial in k and t, or exponential in k?

[1]The aforementioned branching algorithms in [15, 16] automatically imply the existence of a kernel whose size is *exponential* in k. This leverages the well-known equivalence between the existence of fixed parameter tractable (FPT) algorithms and kernels, see e.g. Lemma 2.2 of [6].

2 Preliminaries

An *unrooted binary phylogenetic tree* on X is an undirected tree $T = (V(T), E(T))$ where every internal node has degree 3 and whose leaves (degree-1 nodes) are bijectively labelled by a set of leaf labels (or more formally *taxa*) X, where $|X| = n$. When it is clear from the context we will often refer to an unrooted binary phylogenetic tree as simply a tree. For every taxon $x \in X$ its *parent* in a tree T is the node adjacent to x. A *cherry* in T is a pair of distinct taxa $x, y \in X$ which have a common parent in T.

We refer to $B \subseteq X$, with $|B| \geq 1$, as a *block*. For a block B we write $T[B]$ to denote the minimal subtree of T spanning exactly all the taxa in B. Furthermore we write $T|B$ to denote the phylogenetic tree obtained from $T[B]$ by suppressing nodes of degree 2. Subtree $T[B]$ is called *pendant* in T if there exists an edge such that deleting it splits T into exactly two subtrees, where one tree has taxa B and the other has taxa $X \setminus B$. The *degree* of a block B in a tree T, denoted as $deg^T(B)$, is the number of edges in T that have exactly one endpoint in $T[B]$ (informally, the number of edges which need to be cut to split B from $X \setminus B$). Hence, $deg^T(B) = 1$ precisely if $T[B]$ is pendant.

For $m \geq 1$, let $C = (x_1, x_2 \ldots, x_m)$ be a sequence of distinct taxa in X. We call C an *m-chain* in T, or simply a *chain* in T, if for each $i \in \{1, \ldots m-1\}$ the parent of x_i is either equal to, or adjacent to, the parent of x_{i+1}[2]. Note that if $(x_1, x_2 \ldots, x_m)$ is an m-chain in T then so is $(x_m, x_{m-1} \ldots, x_1)$. Chain $(x_1, \ldots, x_m)$ is *maximal* if there is no chain $(y_1, \ldots, y_{m'})$ with $\{x_1, \ldots, x_m\} \subsetneq \{y_1, \ldots, y_{m'}\}$. Examples of subtrees and chains are given in Figure 1.

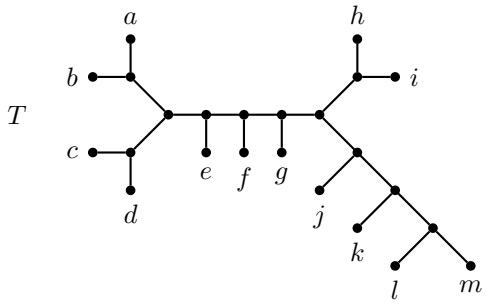

Figure 1: An unrooted binary phylogenetic tree T on $X = \{a, \ldots, m\}$. For example, $T[\{a, b, c, d\}]$ is a pendant subtree, (e, f, g) is a 3-chain with $deg^T(\{e, f, g\}) = 2$ and (j, k, l, m) is a 4-chain in T with $deg^T(\{j, k, l, m\}) = 1$.

For leaf-labelled (not necessarily binary) trees T, T' both having leaf set X, we write $T = T'$ if there is an isomorphism between T and T' that preserves the leaf labels X. We are now ready to define the main structure studied in this paper.

Definition 1. Let $\mathcal{T}$ be a set of unrooted phylogenetic trees on X. An *(unrooted) agreement forest* $F = \{B_1, B_2, \ldots, B_k\}$ for $\mathcal{T}$ is a partition of X such that the following conditions hold.

1. For all $i \in \{1, 2, \ldots, k\}$, $T|B_i = T'|B_i$ for every pair of trees $T, T' \in \mathcal{T}$.

[2]Due to the fact we are working with binary trees x_i and x_{i+1} can only share a parent if $i = 1$ or $i = m - 1$, and $\{x_i, x_{i+1}\}$ form a cherry.

2. For each pair $i, j \in \{1, 2, \ldots, k\}$ with $i \neq j$, the subtrees $T[B_i]$ and $T[B_j]$ are vertex-disjoint in T for every tree $T \in \mathcal{T}$.

Figure 2: Example of an unrooted agreement forest $F = \{\{a, b, c, d\}, \{e\}, \{f\}\}$ for phylogenetic trees T and T'. This is a maximum agreeement forest: a smaller agreement forest is not possible.

The *size* of F is simply $|F|$, i.e., the number of blocks in the partition. An agreement forest with minimum size is called a *maximum* agreement forest (MAF), see Figure 2 for an example. In this paper, we consider the following problem:

The Unrooted Maximum Agreement Forest Problem

Input: A set $\mathcal{T}$ of unrooted binary phylogenetic trees on X and an integer k.
Question: *Does there exist an unrooted agreement forest F for $\mathcal{T}$ of size at most k?*

There is also a *rooted* version of the agreement forest problem. A *rooted* binary phylogenetic tree on X is a tree with a single indegree-0 and outdegree-2 *root*, where the leaves are bijectively labelled by X and all other nodes have indegree 1 and outdegree 2. Edges are directed away from the root. If $|X| = 1$ then the single node labelled with the unique element of X is regarded as a rooted binary phylogenetic tree. In this rooted context $T = T'$ if there is an isomorphism between T and T' that respects the direction of the edges (and, as usual, preserves the leaf labels X). Core definitions, such as $T[B]$ and $T|B$, go through unchanged, with one crucial exception: when constructing $T|B$ from $T[B]$, the unique node of $T[B]$ with indegree-0 and outdegree-2 (i.e. its root) is not suppressed. The substructures for rooted trees are similar to their unrooted counterparts, as shown in Figure 3. However, chains are slightly different. In a rooted phylogenetic tree chains have an orientation, starting with the taxon closest to the root. For example in Figure 3 (g, f, e) is a chain of T, but (e, f, g) is not.

Let $\mathcal{T}$ be a set of rooted binary phylogenetic trees on X. A *(rooted) agreement forest* for $\mathcal{T}$ is defined identically to the unrooted case; the only difference is the different semantics of isomorphism in the $T|B_i = T'|B_i$ condition. See Figure 4 for an example.

Now the problem becomes:

The Rooted Maximum Agreement Forest Problem

Input: A set $\mathcal{T}$ of rooted binary phylogenetic trees on X and an integer k_r.
Question: *Does there exist a rooted agreement forest F_r for $\mathcal{T}$ of size at most k_r?*

Where necessary we will write uMAF and rMAF to distinguish the unrooted and rooted versions of the problem.

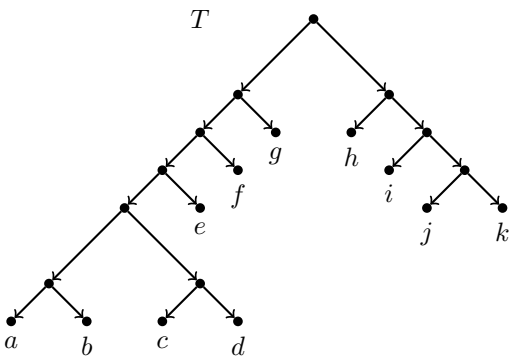

Figure 3: A rooted binary phylogenetic tree T. For example, $T[\{a,b,c,d\}]$ is a pendant subtree, (g,f,e) is a 3-chain and (h,i,j,k) a 4-chain in T.

Figure 4: Example of a rooted agreement forest $F_r = \{\{a,b,c\},\{d,e\}\}$ for the rooted phylogenetic trees T and T'. This is a maximum agreement forest because an agreement forest with fewer blocks is not possible.

We will use the following observation freely in the remainder of the article, which holds for both uMAF and rMAF, and which follows directly from the definitions. It shows that the size of a maximum agreement forest is non-increasing under the action of deleting taxa from the trees.

Observation 1: *Let $\mathcal{T}$ be a set of (rooted or unrooted) binary phylogenetic trees on X and F an agreement forest for $\mathcal{T}$. If $X' \subseteq X$ then*

$$F' = \{B \cap X' \mid B \in F, B \cap X' \neq \emptyset\}$$

is an agreement forest for $\mathcal{T}' = \{T|X' \mid T \in \mathcal{T}\}$ and $|F'| \leq |F|$.

In the main part of the article we will focus solely on uMAF. However, as we will point out later in Section 4.1 the results go through almost entirely unchanged for rMAF.

3 Reduction Rules

Let $\mathcal{T}$ be a set of unrooted binary phylogenetic trees on X and $S \subseteq X$. We say that S *induces a common pendant subtree* of $\mathcal{T}$ if $T[S]$ is pendant in every tree $T \in \mathcal{T}$ and $T[S] = T'[S]$ for each pair of trees $T, T' \in \mathcal{T}$[3]. If S induces a common pendant subtree of $\mathcal{T}$, then the *subtree reduction rule* picks an arbitrary $s \in S$ and replaces each tree $T \in \mathcal{T}$ by $T|((X \setminus S) \cup \{s\})$, see Figure 5.

[3]Note the use of $T[.]$ here: the pendant subtree should be 'rooted' at the same location in every tree in $\mathcal{T}$.

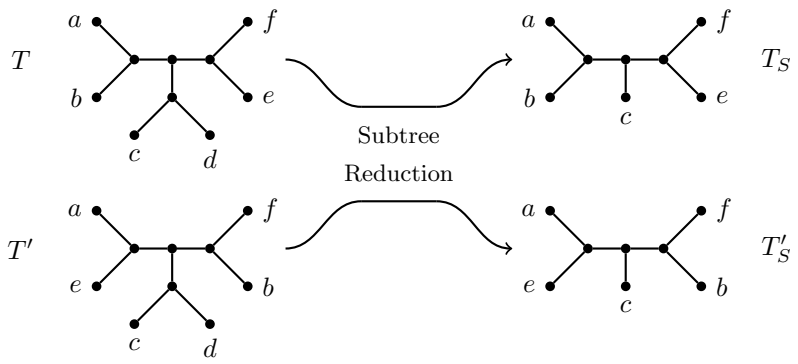

Figure 5: Example of applying the subtree reduction rule on $\mathcal{T} = \{T, T'\}$ with $S = \{c, d\}$, reducing the common pendant subtree to a single taxon c.

The following lemma is folklore in the phylogenetics literature; correctness for two trees is straightforward (see [1, 3] for related discussions) and extends without effort to three or more trees.

Lemma 1: *Let $\mathcal{T}$ be a set of unrooted binary phylogenetic trees on X and let $\mathcal{T}_S$ be the result of applying the subtree reduction rule. Then there exists an unrooted agreement forest of $\mathcal{T}_S$ of size at most k if and only if $\mathcal{T}$ has an unrooted agreement forest of size at most k.*

There is another well-known reduction rule called the *chain reduction* rule. It is known that when the input consists of precisely two trees, we can reduce a common chain of length 4 or more to a common 3-chain. When combined with the subtree reduction rule this is sufficient to obtain a kernel of size $15k$ [9]. However, there is no chain reduction rule known for more than two trees. In the next section we extend the chain reduction to multiple trees and then use it to produce a kernel.

3.1 Chain Reduction for Multiple Trees

Let $\mathcal{T}$ be a set of unrooted binary phylogenetic trees on X. We say that a sequence of taxa C is a *common chain* of $\mathcal{T}$ if C is a chain in each $T \in \mathcal{T}$. The intuition behind the following reduction rule is to shorten long common chains without altering the size of an optimal solution. Let $C' = (x_1, \ldots, x_m)$ be a common m-chain where $m > r$ and $r = \min\{\max\{k, 3\}, t+1\}$. Here, k is the target parameter, i.e. the value k in the question "Is there an agreement forest with at most k blocks?" Note that $r \geq 3$, because $t \geq 2$. The *chain reduction rule* reduces every tree $T \in \mathcal{T}$ by removing all but r taxa from the chain. More formally: let $X_C = (X \setminus \{x_{r+1}, x_{r+2} \ldots, x_m\})$, then replace each $T \in \mathcal{T}$ by $T_C = T|X_C$. (Note that deleting any $m - r$ taxa from C' has the same effect; there is nothing special about deleting the last $m - r$ taxa).

When $t = 2$ we get the well-known reduction rule which reduces a common chain to a common 3-chain. This is proven in [1] to be safe. In the next lemma we prove safeness for $t \geq 2$.

Lemma 2: *Let $\mathcal{T}$ be a set of t unrooted binary phylogenetic trees on X and C' a common m-chain of $\mathcal{T}$ with $m > r = \min\{\max\{k, 3\}, t+1\}$. Let $\mathcal{T}_C$ be the result of applying the chain reduction rule with respect to C'. Then there exists an unrooted agreement forest of $\mathcal{T}_C$ of size at most k if and only if $\mathcal{T}$ has an unrooted agreement forest of size at most k.*

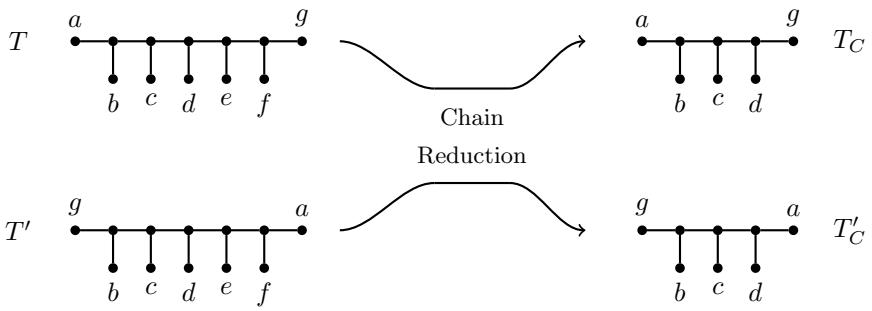

Figure 6: Example of applying the chain reduction rule on $\mathcal{T} = \{T, T'\}$ where $C' = (b, c, d, e, f)$ is a long common chain. We reduce it to the smaller chain $C = (b, c, d)$.

Proof: Let F be an unrooted agreement forest for $\mathcal{T}$ of size k. It then follows from Observation 1 (take X' equal to X_C) that there exists an unrooted agreement forest of $\mathcal{T}_C$ of size at most k. This completes one direction of the proof.

To prove the other direction, suppose there exists an unrooted agreement forest F_C of $\mathcal{T}_C$ of size at most k. We say that a chain is *preserved* by an unrooted agreement forest if all its elements are in the same block of the unrooted agreement forest. Let C be the truncated common chain $(x_1, \ldots, x_r)$. For convenience we overload C to also denote the set of taxa $\{x_1, \ldots, x_r\}$. Suppose that C is not preserved in F_C.

We will prove that there exists an unrooted agreement forest F'_C of $\mathcal{T}_C$ of size at most k in which C is preserved and contained in block B_C. From this (and the fact that $r \geq 3$) it will follow that the taxa $x_{r+1}, \ldots, x_m$ can simply be added to the block B_C to obtain an agreement forest of size at most k for the original set of trees $\mathcal{T}$[4].

We call a block $B \in F_C$ an *inside-out* block if B contains at least one taxon of C and at least one taxon not in C. We claim that F_C has at most two inside-out blocks. To see this, consider any $T \in \mathcal{T}_C$ and observe that there are at most two edges that are incident to a vertex of $T[X_C]$ and to a vertex not in $T[X_C]$. Let e_1, e_r be these edges (if they exist), where e_1 is incident to the parent of x_1 and e_r is incident to the parent of x_r. For each inside-out block B, $T[B]$ contains at least one of e_1, e_r. By the definition of an unrooted agreement forest, for any two blocks $B_1, B_2 \in F_C$, the subtrees $T[B_1]$ and $T[B_2]$ are vertex-disjoint and hence edge-disjoint. It follows that there can be at most two inside-out blocks. We will split into three cases: two, one or zero inside-out blocks.

The first case we consider is that F_C **has exactly two** inside-out blocks B, B'. In this case, we merge the two inside-out blocks and any blocks that are completely contained within C into a single block B_C, giving partition F'_C of X_C. This is depicted schematically in Figure 7, where for simplicity we assume $C = (x_1, x_2, x_3, x_4, x_5)$.

Clearly the size of F'_C is at most the size of F_C. We now argue that F'_C is an unrooted agreement forest of $\mathcal{T}_C$. F'_C satisfies the non-overlapping condition because any block that would overlap with B_C would also overlap with B or B' contradicting that F_C is an unrooted agreement forest. To see that the topological equality condition is also satisfied, consider the two edges e_1, e_r defined above. Assume without loss of generality that $T[B]$ uses e_1 for some $T \in \mathcal{T}_C$. Then $T[B]$ uses e_1 and

[4]It is well-known that for both rooted and unrooted binary phylogenetic trees common chains of length at least 3 can be extended in this way; informally, three taxa are sufficient to ensure that the truncated chain is 'oriented' the same way in every tree in $\mathcal{T}_C$.

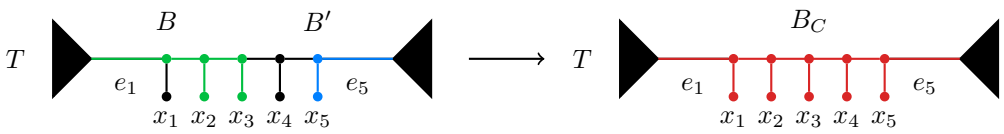

Figure 7: If F_C has two inside-out blocks B (green) and B' (blue) then in every tree $T \in \mathcal{T}_C$, $T[B]$ and $T[B']$ enter the chain in the same way. In this example $T[B]$ will always use e_1 but not e_5, and $T[B']$ will always use e_5 but not e_1, in every $T \in \mathcal{T}_C$. This allows B, B' and the remainder of the chain C to be merged together into a single block B_C (red).

$T[B']$ uses e_r in each tree $T \in \mathcal{T}_C$. From this together with the notion that C has the exact same topology in every tree $T \in \mathcal{T}_C$ it follows that for the merged block B_C the equality $T|B_C = T'|B_C$ holds for all $T, T' \in \mathcal{T}_C$.

The second case is that F_C has **exactly one inside-out block** B. If $|B \cap C| \geq 3$ then observe that either $T[B]$ uses e_1 (but not e_r) in every tree $T \in \mathcal{T}_C$, e_r (but not e_1) in every tree $T \in \mathcal{T}_C$, or both e_1 and e_r in every tree $T \in \mathcal{T}_C$. Whichever situation applies, merging B with all other blocks that intersect C yields the desired agreement forest. This is depicted in Figure 8.

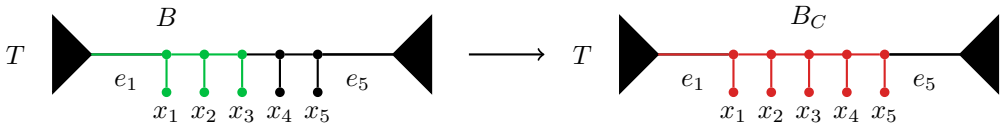

Figure 8: If F_C has exactly one inside-out block B (green) and $|B \cap C| \geq 3$ then $T[B]$ must use the edges e_1 and e_r in the same way for every $T \in \mathcal{T}_C$. In this example $T[B]$ uses e_1 but not e_5, so this will hold for all $T \in \mathcal{T}_C$. This allows us to merge B with the remainder of C to obtain B_C (red).

If $|B \cap C| = 2$ then we distinguish two subcases. In the first subcase, $T[B]$ always uses e_1 (but not e_r), always uses e_r (but not e_1), or always uses both e_1 and e_r, ranging across all trees $T \in \mathcal{T}_C$. Here, as in the case $|B \cap C| = 3$, merging B with all other blocks that intersect C yields the desired agreement forest. This is summarized in Figure 9.

In the second subcase there are distinct trees $T, T' \in \mathcal{T}_C$ where $T[B]$ uses e_1 but not e_r, while $T'[B]$ uses e_r but not e_1[5]. Combined with the fact that $r \geq 3$, this means that $|C \setminus B| \geq 1$ and all taxa in $C \setminus B$ are singleton blocks in the agreement forest. We split B into $B \setminus C$ and $B \cap C$ and then merge $B \cap C$ with the singleton blocks in $C \setminus B$. This does not increase the size of the agreement forest, so we are done with the $|B \cap C| = 2$ case. This is depicted in Figure 10.

If $|B \cap C| = 1$ we proceed as follows. Consider an arbitrary tree $T \in \mathcal{T}_C$. Then $T[B]$ uses one or both of the edges e_1, e_r, and this can vary for different $T \in \mathcal{T}_C$. Observe that due to $r \geq 3$ and $|B \cap C| = 1$ there is at least one block in the agreement forest distinct from B that intersects C.

[5]If $|B \cap C| = 2$ and some tree T has the property that $T[B]$ uses both e_1 and e_r, then all trees $T \in \mathcal{T}_C$ have this property.

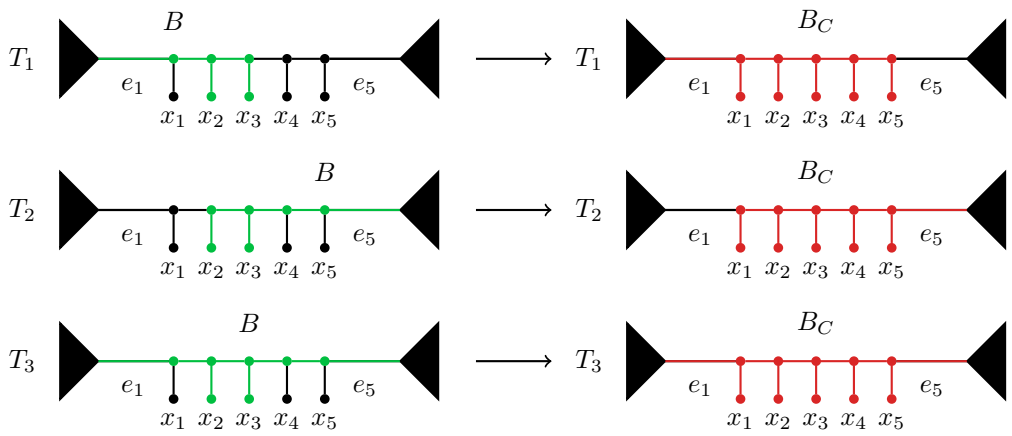

Figure 9: If F_C has exactly one inside-out block B (green) and $|B \cap C| = 2$, and $T[B]$ uses the edges e_1 and e_r in the same way for every $T \in \mathcal{T}_C$ - so in every tree it behaves like in T_1, or in every tree it behaves like in T_2, or in every tree it behaves like in T_3 - then we can merge B with the remainder of C to obtain B_C (red).

If $T[B]$ uses e_1 (but not e_r) for all $T \in \mathcal{T}_C$, or $T[B]$ uses e_r (but not e_1) for all $T \in \mathcal{T}_C$, we can simply merge B with all other blocks that intersect C and we are done; see Figure 11.

If there exist distinct trees $T, T' \in \mathcal{T}_C$ such that $T[B]$ uses e_1 (but not e_r) and $T'[B]$ uses e_r (but not e_1), then the at least two taxa in $C \setminus B$ are both singleton taxa. Now, select an arbitrary $T \in \mathcal{T}_C$ and let x be the single taxon in $|B \cap C|$. Let L and R be the taxa in the two sibling subtrees of x in $T|B$. (If $|B| = 2$ then only one of L and R will be non-empty, say L, otherwise they will both be non-empty). Figure 12 clarifies this concept.

Now, we split B into $\{x\}$, L and (if it exists) R but compensate for this by merging $\{x\}$ with the at least two singleton blocks corresponding to taxa in $C \setminus B$. This does not increase the size of the agreement forest overall, so we are done. See Figure 13.

Suppose then that there exists a tree $T \in \mathcal{T}_C$ such that $T[B]$ uses both e_1 and e_r. The same split-and-merge tactic with L and R as used in the previous paragraph also works here. This concludes the case that $|B \cap C| = 1$. This is also depicted in Figure 13.

The third case is that F_C has **no inside-out blocks**. We say that B is a *bypass block* with respect to $T \in \mathcal{T}_C$ if $B \cap C = \emptyset$ but $T[B]$ uses both e_1 and e_r. If F_C has no bypass blocks with respect to any tree $T \in \mathcal{T}_C$, then we can construct F'_C from F_C simply by merging all blocks contained in C into a single block C and we are done.

Now assume that F_C has at least one bypass block B with respect to some $T \in \mathcal{T}_C$. This implies that all elements of C are singleton blocks in F_C, so F_C has size at least $r + 1$. Since F_C has size at most k by assumption, it follows that $3 \leq r < k$ and hence that $r = t + 1$. In this case, we construct F'_C from F_C as follows. While there exists a tree $T \in \mathcal{T}_C$ for which B is a bypass block with respect to T, split B into B_L, B_R where $T[B_L]$ is incident to edge e_1 and $T[B_R]$ is incident to edge e_r.

After repeating this for all trees for which there is a bypass block (possibly splitting the same block multiple times), merge all blocks contained in C (which are all singletons) into a single block

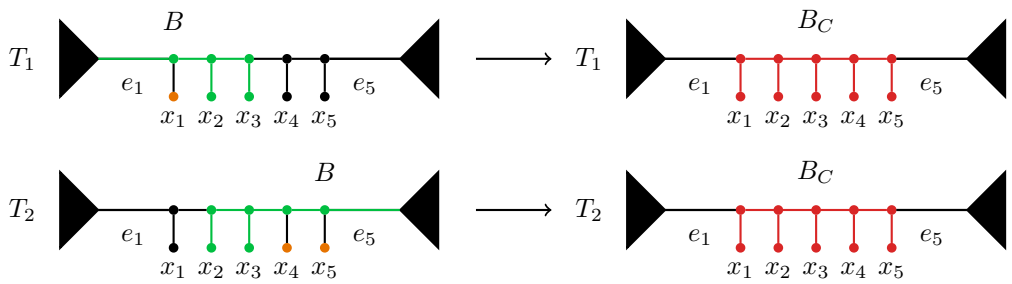

Figure 10: If F_C has exactly one inside-out block B (green) and $|B \cap C| = 2$, and $T[B]$ uses the edges e_1 and e_r in different ways - so in at least one tree it behaves like in T_1, and in at least one tree it behaves like in T_2 - then all the taxa in $C \setminus (B \cap C)$ are necessarily singletons in F_C (shown in orange). In this case we can split $B \cap C$ off from B and merge it with these singletons to obtain B_C (red).

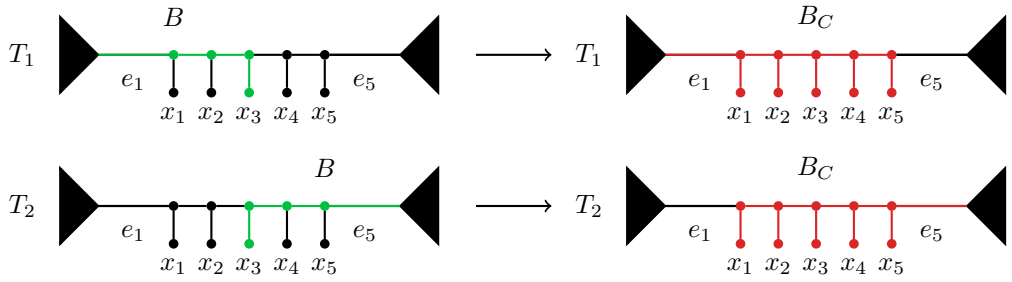

Figure 11: If F_C has exactly one inside-out block B (green) and $|B \cap C| = 1$, and $T[B]$ always behaves like in T_1, or always behaves like in T_2, then we can merge B with the remainder of C to obtain B_C (red).

C. See Figure 14. It is again easy to see that F'_C is an unrooted agreement forest. Clearly there is not overlap because there are no bypass blocks any more. Isomorphism again holds because C has the exact same topology in every tree and we only split bypasses.

The total number of times a block is split is at most t and after that we merged $t+1$ singleton blocks into a single block. Hence, the size of F'_C is at most k.

□

3.2 Tightness of the Truncation Length

It is natural to ask whether the truncation length $r = \min\{\max\{k, 3\}, t+1\}$ used in the chain reduction can be improved i.e. reduced in size. We show that for various combinations of t and k the bound is tight. (For $t = 2$ reducing chains to length 3 is already well-known to be tight).

Our first family of tight examples is as follows; we show for every $t \geq 3$ that there is an instance where reducing the $t+1$ term to t in $r = \min\{\max\{k, 3\}, t+1\}$ is not safe.

Let $t \geq 3$. Construct t different trees $T_1, \ldots T_t$ by starting with a $(t+1)$-chain and denote the taxa in this chain $1, 2, \ldots, t, t+1$. Then for each T_i in $\mathcal{T}$ insert a common chain $C' =$

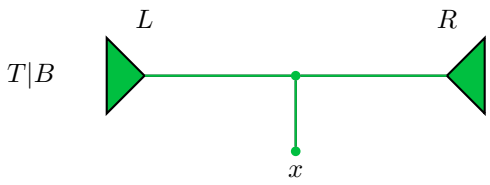

Figure 12: The definition of the sets L and R as used in Figure 13 (where $x = x_3$). Here T can be any tree in $\mathcal{T}_C$.

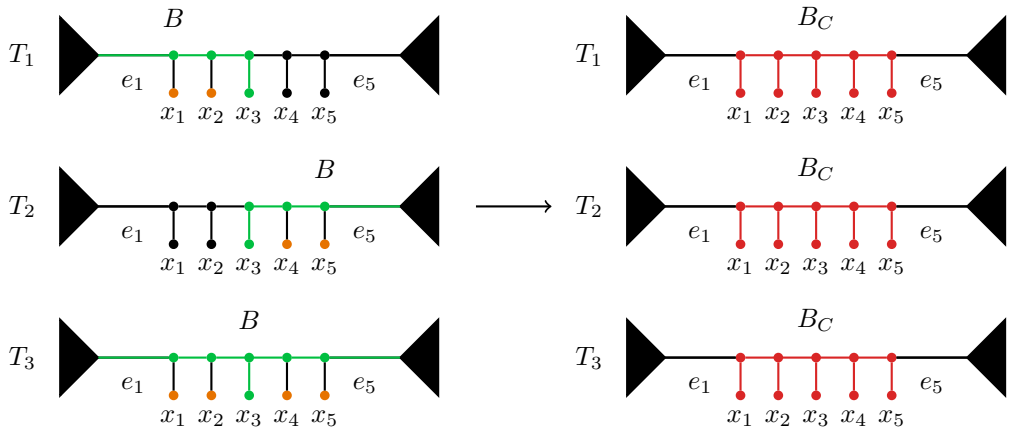

Figure 13: If F_C has exactly one inside-out block B (green) and $|B \cap C| = 1$, and at least one of the two following situations holds – (i) in at least one tree $T[B]$ behaves like in T_1 and in at least one tree $T[B]$ behaves like in T_2, (ii) in at least one tree $T[B]$ behaves like in T_3 – then all the taxa in $(C \setminus B)$ are necessarily singletons (shown in orange). After splitting B into L, R and $(B \cap C)$ (see Figure 12 for the definition of L and R) all the blocks intersecting C can be merged into a single block B_C (red).

$(x_1, x_2, \cdots, x_m)$ after the ith taxon in the t-chain; we let $m = 2(t+2)$. Finally, in T_2 (but not in any of the other trees) we reverse the order of the chain C', resulting in x_m being sibling to taxon 1 in that tree. See Figure 15.

Note that if we truncate C' to a chain C of length t, then the resulting modified set of trees has an agreement forest of size at most $t+1$: we take the taxa from the $(t+1)$-chain as a single block, together with $|C| = t$ singleton blocks. We will show that the original trees did not have an agreement forest of size $t+1$ or less. From this we will conclude that taking a truncation length of $\min\{\max\{k, 3\}, t\}$ and parameter $k = (t+1)$ is not safe, because the first set of trees do not have an agreement forest of size at most $t+1$, but the reduced trees do.

Consider then the original set of trees. Note firstly that if any block B of an agreement forest of the original trees contains two or more taxa from C', then no two taxa from $\{1, 2, \ldots, t, t+1\}$ can be in the same block of the forest. Moreover, if $|B \cap C'| \geq 3$, then B cannot contain any taxa from $\{1, 2, \ldots, t, t+1\}$, meaning that the forest has at least $t+2$ blocks. So suppose there is no block B with the property that $|B \cap C'| \geq 3$. Then due to the choice of $m = 2(t+2)$ it follows that the forest contains at least $t+2$ blocks; we are done.

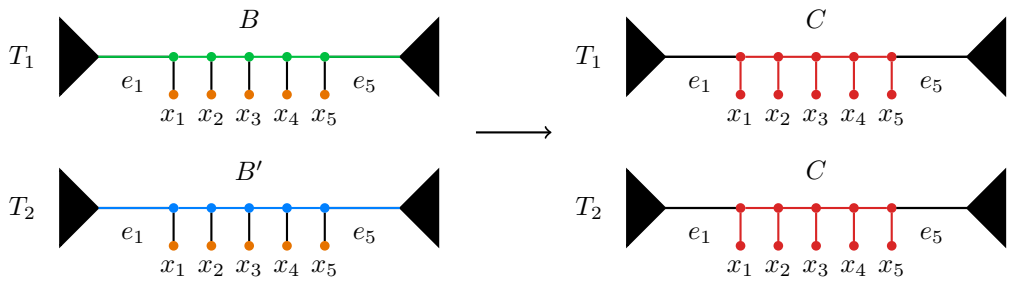

Figure 14: B (green) and B' (blue) are examples of bypass blocks. As soon as there is at least one bypass block in F_C, all the taxa in C are necessarily singletons (shown in orange). After splitting all bypass blocks these singletons can be merged into a single block C, show in in red.

Our second family of tight examples shows that for every $k \geq 4$ a truncation length of $\min\{\max\{k-1,3\}, t+1\}$ is also not safe. The construction is similar to the previous case, except that we take $m = 2k+1$, $t = k+2$ and the chain C' in T_2 is not reversed. That is, there are $t = k+2$ trees and each of these trees is obtained by inserting the chain C' into a starting chain comprising $t+1 = k+3$ taxa. After truncation of the common chain to length $\min\{\max\{k-1,3\}, t+1\} = k-1$, observe that the reduced trees have an agreement forest of size k (i.e. the original taxa $\{1,2,\ldots,t,t+1\}$ in one block, and $k-1$ singleton blocks corresponding to taxa in C). Before truncation, note that if a block B has the property that $|B \cap C'| \geq 3$, then the taxa $\{2,\ldots,t\}$ must be in separate blocks, meaning that there will be at least $t-1 = k+1 > k$ blocks in the forest. But if no such block B exists, then there will be at least $(2k+1)/2 > k$ blocks. Hence, prior to truncation, an agreement forest contains at least $k+1$ blocks.

4 Bounding the Number of Leaves in Each Tree

Lemma 3: *Let $\mathcal{T}$ be a set of t unrooted binary phylogenetic trees on X and F an unrooted agreement forest for $\mathcal{T}$ of size $k \geq 2$. If neither the subtree nor the chain reduction are applicable in $\mathcal{T}$, then for each block $B \in F$, it holds that*

$$|B| \leq 2r \sum_{T \in \mathcal{T}} deg^T(B) - 3r$$

with $r = \min\{\max\{k,3\}, t+1\}$ and $t \geq 2$.

Proof: Let $B \in F$ and $d = \sum_{T \in \mathcal{T}} deg^T(B)$. Note that if $|B| = 1$ then $d = t \geq 2$ and the right hand side of the claimed inequality evaluates to at least $4r - 3r = r \geq 3$, so the claim holds immediately. Hence we can focus on $|B| \geq 2$.

For an arbitrary tree $T^* \in \mathcal{T}$, consider the tree T_B obtained from $T^*|B$ as follows. For an edge e of $T^*|B$ and tree $T \in \mathcal{T}$, let $n(e,T)$ be the number of internal vertices on the path in $T[B]$ corresponding to[6] edge e. Subdivide each edge e of $T^*|B$ by $\sum_{T \in \mathcal{T}} n(e,T)$ degree-2 vertices. Call the obtained tree T_B^{full}. Then, in T_B^{full}, delete all leaves and suppress their parents if they become

[6] If $e = \{u,v\}$ then the path corresponding to e in $T[B]$ is the path that starts at u and ends at v. The interior vertices of this path are the degree-2 vertices that are suppressed in order to create e in $T|B$.

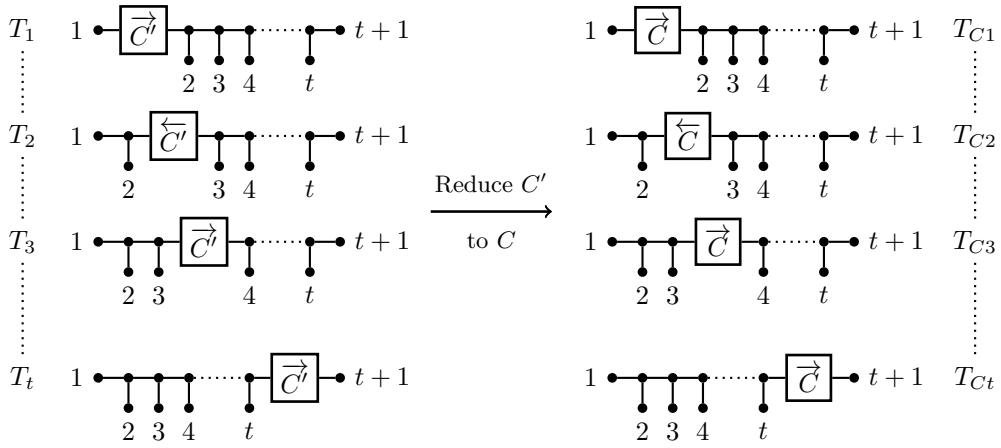

Figure 15: If we take the common chain C' to have length 2(t+2) then truncating it to a length t common chain C alters the size of a maximum agreement forest from at least $t+2$ to at most $t+1$.

degree-2. (Note however that we do not suppress vertices that were degree 2 *before* the deletion of the leaves). This gives tree T_B, see Figure 16 for an example.

Observe that the total number of degree-2 vertices and leaves of T_B is d and that $d \geq t \geq 2$.

We prove by induction on d that a tree S with d vertices of degree at most 2 has at most $2d-3$ edges. If $d = 2$ then S has exactly one edge and we are done. If $d \geq 3$, let S' be the result of deleting a leaf and suppressing its neighbour if it becomes degree-2. This reduces the number of vertices of degree at most 2 by 1. By induction, S' has at most $2(d-1)-3 = 2d-5$ edges and therefore S has at most $2d-3$ edges. Hence, T_B has at most $2d-3$ edges.

We now show that the number of leaves of T_B^{full} is at most rm with m the number of edges of T_B. Consider any maximal chain C of T_B^{full}. Then the leaves of C are deleted from T_B^{full} in the construction of T_B and their neighbours are suppressed if they get degree-2. Observe that it is not possible that a neighbour of a leaf had degree-3 in T_B^{full} and becomes degree-1 in T_B because then the subtree reduction would have been applicable in $\mathcal{T}$. Hence, we can distinguish the following three cases. In each case we will map or "charge" each leaf to an edge of T_B. We will do this in such a way that leaves from different maximal chains of T_B^{full} are charged to different edges of T_B, which will make it possible to bound the total number of leaves.

The **first case** is that all neighbours of leaves in C are suppressed. Let e be the edge of T_B created by suppressing these neighbours. Charge all leaves of C to edge e.

The **second case** is that $|C| \geq 2$ and there is at least one leaf x in C such that the neighbour v of x in T_B^{full} is a degree-2 vertex, and thus is not suppressed when x is deleted. Note that there can be at most two such leaves since C is a chain. If there are exactly *two* such leaves, then B is a common chain, so $B = C$, and $d = 2$. In this case, T_B has exactly one edge and we charge all leaves of B to it. If there is exactly *one* such leaf x, then all other neighbours of leaves in C are suppressed, creating an edge e. Charge all the leaves of C, including x, to e.

It remains to consider the **third case**: that $C = \{x\}$ and the neighbour v of x in T_B^{full} has degree 2. Then v is not suppressed. Consider the edge e of T_B^{full} incident to v but not to x.

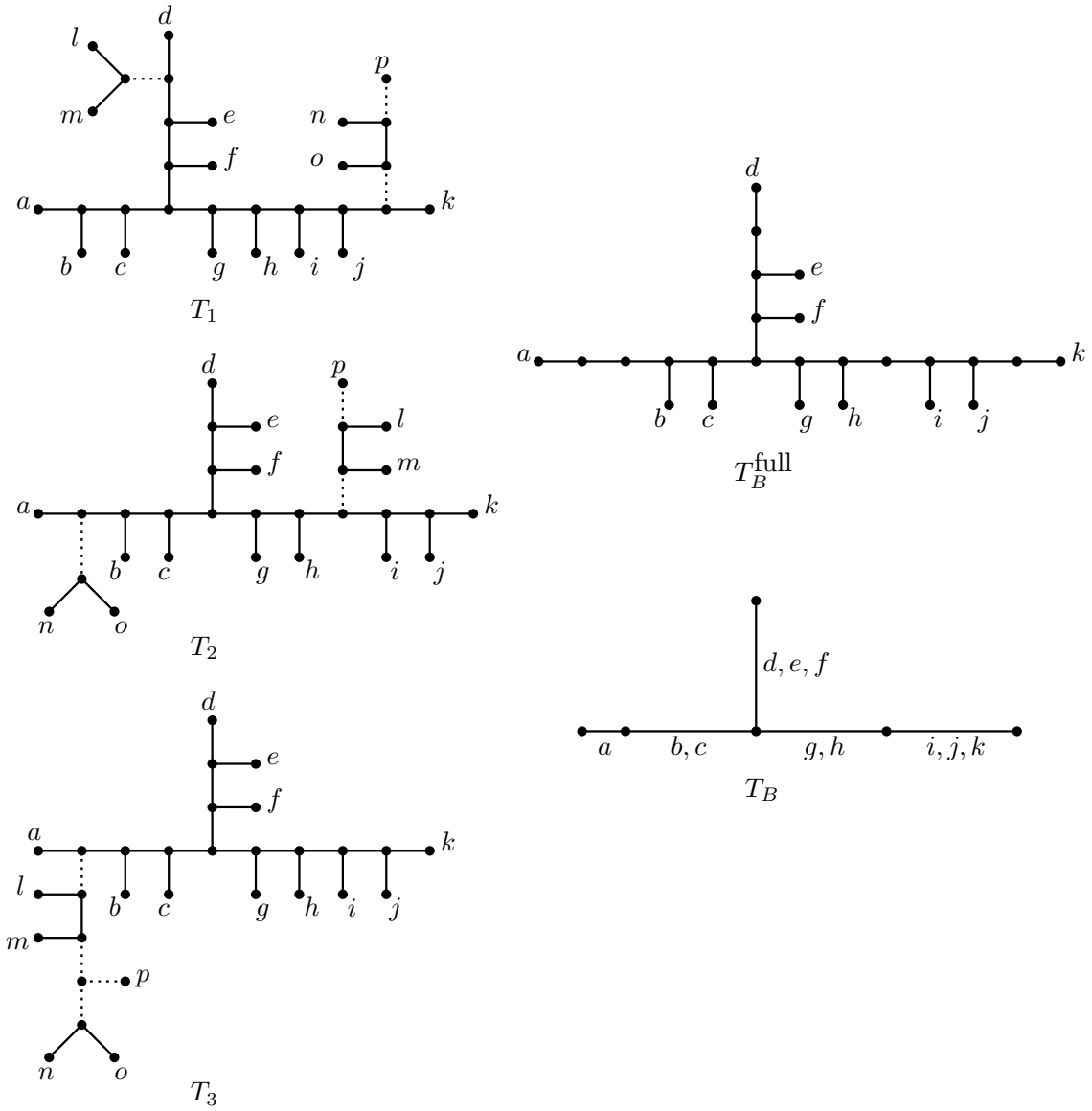

Figure 16: Example of the construction of trees T_B^{full} and T_B in the proof of Lemma 3 with $B = \{a, b, \ldots, k\}$. Each edge of T_B is labelled by the leaves that are charged to it. The edges labelled b, c and g, h are considered in the first case of the proof, the edges labelled d, e, f and i, j, k are considered in the second case, while the edge labelled a is considered in the third case.

Charge x to e. Observe that no other leaves are charged to e because otherwise C would not have been maximal.

In general, leaves from different maximal chains of T_B^{full} are never charged to the same edge of T_B. Moreover, each maximal chain contains at most r leaves because otherwise the chain reduction would have been applicable in $\mathcal{T}$. Hence, it follows that the number of leaves of T_B^{full} is at most the number of edges of T_B times r.

This concludes the proof since $|B|$ is equal to the number of leaves of T_B^{full}, which we have shown to be at most r times the number of edges of T_B, which we have shown to be at most $2d-3$. □

Lemma 4: *For every $T \in \mathcal{T}$ the sum of the degrees of each block B in F is bounded as follows, where k is the size of F:*

$$\sum_B deg^T(B) \leq 2k-2.$$

Proof: Let T be any tree in $\mathcal{T}$ and let B be a block of F. Consider $T[B]$. If T has an edge e such that one endpoint u of e is a node of $T[B]$ and the other endpoint v is not, and v does not lie on $T[B']$ for any $B' \in F$, then v is necessarily not a leaf (because every taxon is in some block of the forest), and thus has degree larger than 1. If we expand $T[B]$ with the edge e, then the degree of this expanded block (i.e. the number of edges that have exactly one endpoint on the expanded block) is at least as large as the degree of B in T: this is because there is at least one edge incident to v that is not equal to e. We iterate this procedure to exhaustion i.e. until the point that no block $B \in F$ can be further expanded in T. At this point, there will be exactly $k-1$ edges in T that are not part of any expanded block, and each such edge will have its endpoints on two distinct expanded blocks. Hence, the total degree of the expanded blocks in T will be exactly $2(k-1)$. By construction, this is an upper bound on the sum of the degrees of the original blocks in T, so $\sum_B deg^T(B) \leq 2k-2$. □

Theorem 1: *Let $\mathcal{T}$ be a set of t unrooted binary phylogenetic trees on X. If there exists an unrooted agreement forest of $\mathcal{T}$ of size at most k and neither the subtree nor chain reduction is applicable, then every tree in $\mathcal{T}$ has at most $4trk-4tr-3rk$ taxa, with $r = \min\{\max\{k,3\}, t+1\}$.*

Proof: Let F be an unrooted agreement forest of $\mathcal{T}$ of size at most k. We can bound the number of leaves in each tree as follows.

$$\begin{aligned}
|X| &\leq \sum_{B \in F} |B| \\
&\leq \sum_B \left(2r \sum_T deg^T(B) - 3r \right) && \text{(Lemma 3)} \\
&= 2r \sum_T \sum_B deg^T(B) - \sum_B 3r \\
&= 2r \sum_T \sum_B deg^T(B) - 3rk \\
&\leq 2r \sum_T (2k-2) - 3rk && \text{(Lemma 4)} \\
&= t2r(2k-2) - 3rk \\
&= 4trk - 4tr - 3rk.
\end{aligned}$$

□

We note that for $t = 2$, at which point $r = 3$, the above bound becomes $24k-24-9k = 15k-24$. This matches exactly with the analysis in [9] which shows that for these two reduction rules the kernel has at most $15d_{\text{TBR}} - 9$ taxa where d_{TBR}, the *Tree Bisection and Reconnect* distance, is defined to be equal to the size of a maximum agreeement forest *minus one*. Their analysis is tight: they show two trees for which the kernel has exactly $15d_{\text{TBR}} - 9$ taxa. Hence, for $t = 2$, our analysis is also tight.

4.1 Adapting to Rooted Trees

It is well-known that the subtree reduction rule holds in the case of multiple rooted trees. The chain reduction rule is also safe for the case of multiple rooted trees, as the next lemma shows.

Lemma 5: *Let $\mathcal{T}$ be a set of rooted binary phylogenetic trees on X and C' a common chain of $\mathcal{T}$. Let $\mathcal{T}_C$ be the result of applying the chain reduction rule. Then there exists a rooted agreement forest of $\mathcal{T}_C$ of size at most k_r if and only if $\mathcal{T}$ has a rooted agreement forest of size at most k_r.*

Proof: In the proof for the unrooted variation of the chain reduction rule, Lemma 2, there were three main cases to discuss: F_C had 2, 1 or 0 inside-out blocks. A careful re-reading of that proof shows that all these cases and their various subcases go through in the rooted case. □

Let F_r be a rooted agreement forest. We define the *outdegree* of a block B in T, $outdeg^T(B)$, to be the number of edges of T which have their tail on $T[B]$ but not their head. We state the following modification of Lemma 4 without proof.

Lemma 6: *For every $T \in \mathcal{T}$ the sum of the outdegrees of each block B in F_r is bounded as follows, where k_r is the size of F_r:*

$$\sum_B outdeg^T(B) \leq k_r - 1.$$

Lemma 3 also requires some adjustment. Consider a block B of F_r where $T|B$ has $c \geq 1$ cherries (here T is an arbitrary tree from $\mathcal{T}$). To avoid triggering the subtree reduction it must hold that, for each such cherry $\{x, y\}$, at least one tree $T \in \mathcal{T}$ has the property that some edge of T not within $T[B]$ has its tail on the (undirected) path between x and y in T. Hence, $\sum_T outdeg^T(B) \geq c$. If $\sum_T outdeg^T(B) = c$ then it is possible that B has up to $r(2c-1)$ taxa. This is because a rooted binary tree on c leaves has $2(c-1) = 2c-2$ edges, each of which can correspond to an r-chain, but B can also have an extra r-chain that feeds into the top of the tree. This is summarized in the next lemma; note that the lemma only holds for non-singleton blocks, as $\sum_T outdeg^T(B) = 0$ for singleton blocks. See also Figure 17.

Lemma 7: *The number of taxa in a non-singleton block B of F_r is at most*

$$|B| \leq 2r \sum_T outdeg^T(B) - r$$

with $r = \min\{\max\{k_r, 3\}, t+1\}$ and t the number of trees.

This means that Theorem 1 can also be adjusted to incorporate rooted agreement forests and provide a very similar result.

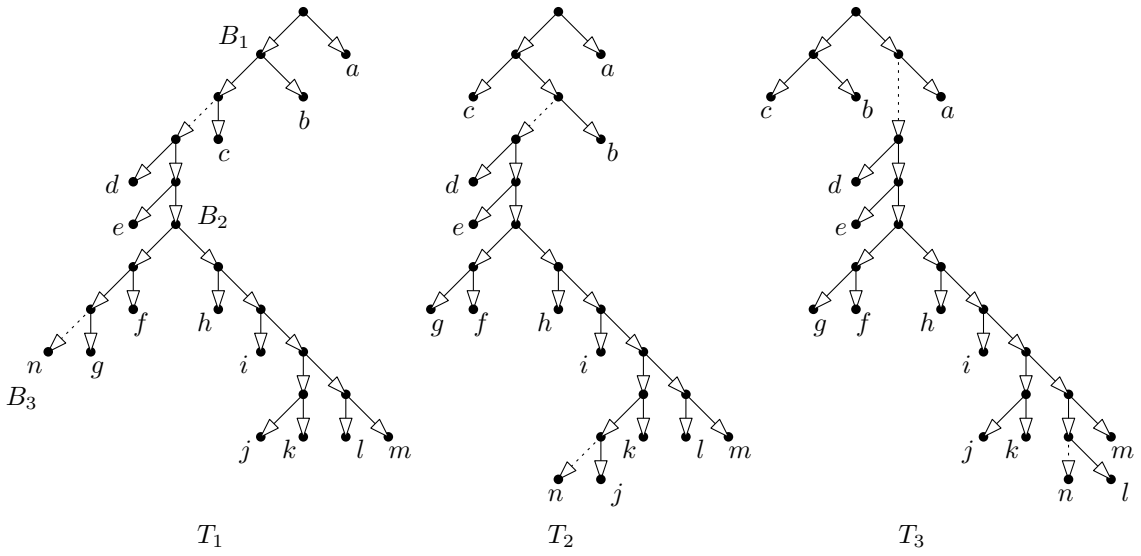

Figure 17: Illustration of Lemma 7. Block $B_2 = \{d, e, \ldots, m\}$ has total outdegree, over all trees, of 3, so it may contain up to $5r$ taxa. For ease of illustration, we used chains of length 2, leading to $|B_2| = 10$.

Theorem 2: *Let $\mathcal{T}$ be a set of t rooted binary phylogenetic trees on X. If there exists a rooted agreement forest for $\mathcal{T}$ of size at most k_r and neither the subtree nor chain reduction is applicable, then every tree in $\mathcal{T}$ has at most $(2tr + 1)(k_r - 1) - r$ taxa, with $r = \min\{\max\{k_r, 3\}, t + 1\}$.*

Proof: Consider a rooted agreement forest F_r of size at most k_r. We may assume that at least one block of F_r has size at least 2 since otherwise we can simply merge two arbitrary blocks. We can then bound the number of leaves in each tree as follows.

$$
\begin{aligned}
|X| &\leq \sum_{B \in F_r} |B| \\
&= \sum_{|B|=1} 1 + \sum_{|B|>1} |B| \\
&\leq k_r - 1 + \sum_{|B|>1} \left(2r \sum_{T} outdeg^T(B) - r \right) && \text{(Lemma 7)} \\
&\leq k_r - 1 + \sum_{B} \left(2r \sum_{T} outdeg^T(B) \right) - r \\
&\leq k_r - 1 + 2r \sum_{T} (k_r - 1) - r && \text{(Lemma 6)} \\
&= k_r - 1 + 2tr(k_r - 1) - r \\
&= (2tr + 1)(k_r - 1) - r.
\end{aligned}
$$

□

For $t = 2$ and thus $r = 3$ this gives $13k_r - 16$. We note that when restricted to the subtree and chain reduction rule the analysis in [10] also yields a bound of $13k_r - 16$ for $t = 2$; more precisely, $13d_{\text{rSPR}} - 3$ where d_{rSPR} (the *rooted Subtree Prune and Regraft* distance) is equal to $k_r - 1$. Their bound is tight[7], including additive terms, and thus so is ours.

Remark. The second family of tight examples given for unrooted trees in Section 3.2, which shows that the k cannot be reduced to $k - 1$ in $r = \min\{\max\{k, 3\}, t + 1\}$, works unchanged for rooted trees; the trees can simply be rooted on the edge incident to taxon 1.

However, a slightly more complex construction is required to show that for $t \geq 2$ rooted trees, $t+1$ cannot be reduced to t in $r = \min\{\max\{k, 3\}, t+1\}$. We construct the following set of t trees. Every tree contains a chain C' with $m = 2t + 1$ taxa, and $2t$ other distinct taxa labeled x_i and y_i, $1 \leq i \leq t$. To construct T_i we start by taking a cherry on $\{x_1, y_1\}$, placing this on the left side of the root, and on the right side of the root we place the remaining taxa $x_2, y_2, x_3, y_3, \ldots, x_t, y_t$ in a chain (so there are two cherries in total). We then cyclically increment the indices of the labels $(i-1)$ times. Note in particular that applying a cyclical increment to x_t and y_t sends them back to x_1 and y_1 respectively. We complete the construction of T_i by inserting the chain C' between taxa x_{i+1} and y_{i+1} (for $i < t$) or between taxa x_1 and y_1 (for $i = t$). The construction is summarized in Figure 18. Note that for clarity the figure assumes $t \geq 3$, but the construction also works for $t = 2$.

For $t = 2$, truncating C' to length 2 reduces the size of a maximum agreement forest from 4 to 3 (we omit details), so truncation to length t is not safe for $t = 2$ (at $k = 3$). So, we assume $t \geq 3$. It is easy to verify that, if the chain C' is truncated to length t, the resulting trees have an agreement forest with at most $2t$ blocks: the t taxa from the truncated chain become singletons, together with blocks $\{x_i, y_i\}$ for $1 \leq i \leq t$. We will show that prior to truncation an agreement forest has at least $2t + 1$ blocks. From this we will conclude that a truncation length of $\min\{\max\{k, 3\}, t\}$ is not safe, because when taking $k = 2t$ (noting that $\min\{\max\{2t, 3\}, t\} = t$) the original trees do not have an agreement forest of size at most $2t$, but the reduced trees do. So, consider a maximum agreement forest of the original trees. If there is no block B such that $|B \cap C'| \geq 2$, then there are at least $m = 2t + 1$ blocks in the forest, so we are done. Hence, assume that there is at least one block B where $|B \cap C'| \geq 2$ and let c, d be distinct taxa in $B \cap C'$. Observe that B cannot contain any taxon from outside C'. To see why, suppose it contained x_j for some j (the same reasoning will hold for y_j). In some tree x_j appears in the cherry left of the root, i.e. 'above' C', while in some tree x_j appears in the cherry 'underneath' C': these two trees induce different topology subtrees on $\{c, d, x_j\}$, violating the definition of an agreement forest. Moving on, it follows that no two taxa from outside C' can be in the same block $B' \neq B$ of the agreement forest. This is because, by construction, for every two distinct taxa in $\{x_1, y_1, ..., x_t, y_t\}$ there exists some tree T_j in which one of the two taxa is above C', and the other taxon is below C', so $T_j[B]$ and $T_j[B']$ would intersect, again contradicting the definition of an agreement forest. This means that the taxa $\{x_1, y_1, ..., x_t, y_t\}$ induce $2t$ singleton blocks in the forest. Counting B too, that yields at least $2t + 1$ blocks.

5 Future Work

Although the truncation length r used in our chain reduction is in some sense tight, it is unclear whether the overall bound on the size of the kernel is tight for $t > 2$. For $t = 2$ our bounds perfectly

[7] The construction shown in Figure 3 of [10] to demonstrate tightness for the three reduction rules considered there, can be adapted to show tightness when only the subtree and chain reduction rule are considered. This can be achieved by placing 3 taxa on each "edge side" of the underlying generator.

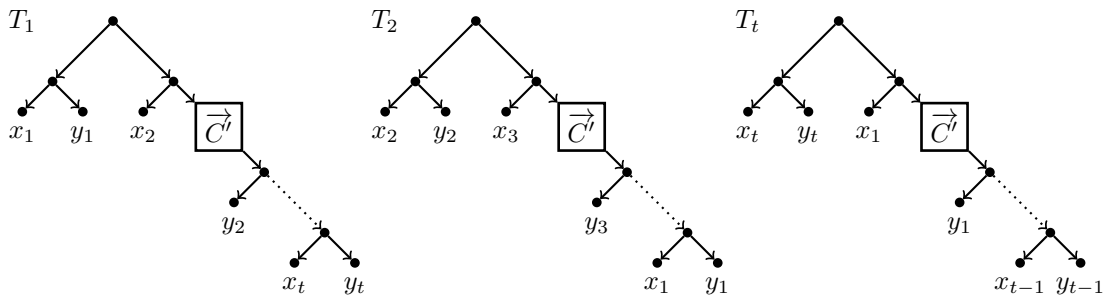

Figure 18: If we take the common chain C' to have length $2t+1$ then truncating it to a length t common chain C alters the size of a maximum agreement forest from at least $2t+1$ to at most $2t$.

match existing bounds which are known to be tight under subtree and chain reduction. The main question is whether for $t > 2$ the counting in Lemmas 3 and 4 (and their rooted analogues) can be undertaken more carefully. For $t = 2$ the introduction of generators made the counting much easier [9], and subsequently became the basis for new, more powerful reduction rules [10, 11]. Generators are in essence a static, graph-based representation of maximum agreement forests, but they do not work for $t > 2$. Finding an alternative to generators for $t > 2$ seems an important direction for future research.

Relatedly, it is natural to ask whether the design and deployment of new, additional reduction rules can produce a smaller kernel. A natural starting point would be to analyze the extra rules that have already been developed for $t = 2$. For the rooted problem this is the '3-2 chain reduction' described in [10]. In contrast, for the unrooted problem there are eight extra reduction rules known [11]. Although the generator machinery is not available to us in the $t > 2$ regime some of these reduction rules might still be correct. In such a case the main challenge will be measuring the impact of the reduction rules on the size of the kernel.

Currently, depending on whether $t \ll k$, $k \ll t$ or $t \approx k$ we obtain a kernel with $O(t^2k)$, $O(tk^2)$ or $O(t^3)$ (equivalently $O(k^3)$) leaves respectively.[8] In an applied context it is plausible that the first scenario (comparing a small number of highly discordant gene trees) or the second scenario (comparing a large number of broadly similar gene trees) will often occur. Possibly specialized reduction rules can be designed for each of these scenarios separately.

Perhaps the most tantalizing open question is whether rMAF and uMAF admit a kernel whose size is polynomial in *only* k. The results in the present article show that they admit a kernel whose size is at most polynomial in k *and* t. On the other hand, the branching algorithms for rMAF and uMAF with running times $O^*(c^k)$ in [15, 16] automatically imply the existence of a 'trivial' kernel whose size is (single) exponential in only k (see e.g. Lemma 2.2 of [6] for a proof of this implication; it is a foundational fact in parameterized complexity). The number of distinct binary phylogenetic trees on n leaves is bounded by a function of n, but the bound is super-exponential (see Proposition 2.1.4 and Corollary 2.2.4 of [14]), so to obtain a polynomial kernel in k it will be necessary to produce a reduced instance on at most poly(k) trees and poly(k) leaves. A potentially relevant fact here is that, for rMAF and uMAF, there is no obvious function f of k whereby we can say "If $t \geq f(k)$ then a MAF with at most k blocks does not exist." To see

[8] Measured in bits the size of the kernel is slightly larger. This is because a reasonable encoding of t binary phylogenetic trees each on n leaves requires $N = \Theta(tn \log n)$ bits.

this, observe that two distinct trees, each on $n/2$ leaves, can be connected together with a single edge to obtain $O(n^2)$ trees on n leaves, which have a MAF of size only 2. Indeed, it might be that a kernel whose size is bounded by a polynomial of k simply does not exist, under reasonable complexity-theoretic assumptions. The literature on kernel-size lower bounds (using techniques such as *cross-composition* [2]; see also the chapter on lower bounds in [7]) could be very useful in this regard.

Acknowledgements

We thank the anonymous reviewers for their helpful feedback. We also thank Enrico Iurlano and Johannes Varg for pointing out an error in the tightness construction in an earlier version of the article.

References

[1] B. Allen and M. Steel. Subtree transfer operations and their induced metrics on evolutionary trees. *Annals of Combinatorics*, 5(1):1–15, 2001. doi:10.1007/s00026-001-8006-8.

[2] H. Bodlaender, B. Jansen, and S. Kratsch. Kernelization lower bounds by cross-composition. *SIAM Journal on Discrete Mathematics*, 28(1):277–305, 2014. doi:10.1137/120880240.

[3] M. Bordewich and C. Semple. On the computational complexity of the rooted subtree prune and regraft distance. *Annals of Combinatorics*, 8(4):409–423, January 2004. doi:10.1007/s00026-004-0229-z.

[4] L. Bulteau and M. Weller. Parameterized algorithms in bioinformatics: An overview. *Algorithms*, 12(12):256, 2019. doi:10.3390/a12120256.

[5] Z.-Z. Chen, Y. Fan, and L. Wang. Faster exact computation of rSPR distance. *Journal of Combinatorial Optimization*, 29(3):605–635, 2015. doi:10.1007/s10878-013-9695-8.

[6] M. Cygan, F. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms*. Springer, 2015. doi:10.1007/978-3-319-21275-3.

[7] F. Fomin, D. Lokshtanov, S. Saurabh, and M. Zehavi. *Kernelization: theory of parameterized preprocessing*. Cambridge University Press, 2019. doi:10.1017/9781107415157.

[8] J. Hein, T. Jiang, L. Wang, and K. Zhang. On the complexity of comparing evolutionary trees. *Discrete Applied Mathematics*, 71(1-3):153–169, 1996. doi:10.1016/S0166-218X(96)00062-5.

[9] S. Kelk and S. Linz. A tight kernel for computing the tree bisection and reconnection distance between two phylogenetic trees. *SIAM Journal on Discrete Mathematics*, 33(3):1556–1574, 2019. doi:10.1137/18M122724X.

[10] S. Kelk, S. Linz, and R. Meuwese. Cyclic generators and an improved linear kernel for the rooted subtree prune and regraft distance. *Information Processing Letters*, 180:106336, 2023. doi:10.1016/j.ipl.2022.106336.

[11] S. Kelk, S. Linz, and R. Meuwese. Deep kernelization for the tree bisection and reconnection (TBR) distance in phylogenetics. *Journal of Computer and System Sciences*, 142:103519, 2024. doi:10.1016/j.jcss.2024.103519.

[12] D. Mestel, S. Chaplick, S. Kelk, and R. Meuwese. Split-or-decompose: Improved FPT branching algorithms for maximum agreement forests. *Journal of Computer and System Sciences*, 160:103800, 2026. doi:10.1016/j.jcss.2026.103800.

[13] PACE Challenge. PACE 2026 — (rooted) maximum-agreement forest. https://pacechallenge.org/2026/maf/, 2026. Accessed: 2026-01-23.

[14] C. Semple and M. Steel. *Phylogenetics*. Oxford University Press, 2003. doi:10.1093/oso/9780198509424.001.0001.

[15] F. Shi, J. Wang, J. Chen, Q. Feng, and J. Guo. Algorithms for parameterized maximum agreement forest problem on multiple trees. *Theoretical Computer Science*, 554:207–216, 2014. doi:10.1016/j.tcs.2013.12.025.

[16] F. Shi, J. Chen, Q. Feng, and J. Wang. A parameterized algorithm for the maximum agreement forest problem on multiple rooted multifurcating trees. *Journal of Computer and System Sciences*, 97:28–44, 2018. doi:10.1016/j.jcss.2018.03.002.