# Maximum parsimony distance on phylogenetic trees: A linear kernel and constant factor approximation algorithm 

Mark Jones, Steven Kelk, Leen Stougie

## - To cite this version:

Mark Jones, Steven Kelk, Leen Stougie. Maximum parsimony distance on phylogenetic trees: A linear kernel and constant factor approximation algorithm. Journal of Computer and System Sciences, 2021, 117, pp.165-181. 10.1016/j.jcss.2020.10.003 . hal-03498430

HAL Id: hal-03498430
https://inria.hal.science/hal-03498430
Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Maximum parsimony distance on phylogenetic trees: A linear kernel and constant factor approximation algorithm 

Mark Jones ${ }^{\text {a,b,* }}$, Steven Kelk ${ }^{\text {c }}$, Leen Stougie ${ }^{\text {b,d,e }}$<br>${ }^{\text {a }}$ Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Delft, the Netherlands<br>${ }^{\text {b }}$ Centrum Wiskunde E Informatica (CWI), 1098 XG Amsterdam, the Netherlands<br>${ }^{\text {c }}$ Department of Data Science and Knowledge Engineering (DKE), Maastricht University, 6200 MD Maastricht, the Netherlands<br>${ }^{\text {d }}$ Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands<br>e INRIA-Erable, France

## A R T I C L E I N F O

## Article history:

Received 7 April 2020
Received in revised form 23 October 2020
Accepted 26 October 2020
Available online 7 December 2020

## Keywords:

Phylogenetics
Maximum parsimony
Fixed parameter tractability
Maximum agreement forest


#### Abstract

Maximum parsimony distance is a measure used to quantify the dissimilarity of two unrooted phylogenetic trees. It is NP-hard to compute, and very few positive algorithmic results are known due to its complex combinatorial structure. Here we address this shortcoming by showing that the problem is fixed parameter tractable. We do this by establishing a linear kernel i.e., that after applying certain reduction rules the resulting instance has size that is bounded by a linear function of the distance. As powerful corollaries to this result we prove that the problem permits a polynomial-time constantfactor approximation algorithm; that the treewidth of a natural auxiliary graph structure encountered in phylogenetics is bounded by a function of the distance; and that the distance is within a constant factor of the size of a maximum agreement forest of the two trees, a well studied object in phylogenetics.


© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Phylogenetics is the science of inferring and comparing trees (or more generally, graphs) that represent the evolutionary history of a set of species [34]. In this article we focus on trees. The inference problem has been comprehensively studied: given only data about the species in $X$ (such as DNA data) construct a phylogenetic tree which optimizes a particular objective function $[17,40]$. Informally, a phylogenetic tree is simply a tree whose leaves are bijectively labelled by $X$. Due to different objective functions, multiple optima and the phenomenon that certain genomes are the result of several evolutionary paths (rather than just one) we are often confronted with multiple "good" phylogenetic trees [32]. In such cases we wish to formally quantify how dissimilar these trees really are. This leads naturally to the problem of defining and computing the distance between phylogenetic trees [36]. Many such distances have been proposed, some of which can be computed in polynomial-time, such as Robinson-Foulds (RF) distance [33], and some of which are NP-hard, such as Subtree Prune and Regraft (SPR) distance [9] or Tree Bisection and Reconnection (TBR) distance [1].

Interestingly, distances are not only relevant as a numerical quantification of difference: they also appear in constructive methods for the inference of phylogenetic networks [20], which generalise trees to graphs, and phylogenetic supertrees,

[^0]which seek to merge multiple trees into a single summary tree [42]. In recent decades NP-hard phylogenetic distances have attracted quite some attention from the discrete optimization and parameterized complexity communities, see e.g. [12,16].

In this article we focus on a relatively new distance measure, maximum parsimony distance, henceforth denoted $d_{M P}$. Let $T_{1}$ and $T_{2}$ be two unrooted (i.e. undirected) binary phylogenetic trees, with the same set of leaf labels $X$. Consider an arbitrary assignment of colours ("states") to $X$; we call such an assignment a character. The parsimony score of $T_{1}$ with respect to the character is the minimum number of bichromatic edges in $T_{1}$, ranging over all possible colourings of the internal vertices of $T_{1}$. The parsimony distance of $T_{1}$ and $T_{2}$ is the maximum absolute difference between parsimony scores of $T_{1}$ and $T_{2}$, ranging over all characters $[18,31]$.

The distance has several attractive properties; it is a metric, and (unlike e.g. RF distance) it is not confounded by the influence of horizontal evolutionary events [18]. Furthermore, the concept of parsimony, which lies at the heart of $d_{M P}$, is fundamental in phylogenetics since it articulates the idea that explanations of evolutionary history should be no more complex than necessary. Alongside its historical significance for applied phylogenetics [17], the study of character-based parsimony has given rise to many beautiful combinatorial and algorithmic results; we refer to e.g. [37,29,38,2,30] for overviews.

Unfortunately, it is NP-hard to compute $d_{M P}$ [22]. A simple exponential-time algorithm is known [26], which runs in time $O\left(\phi^{n} \cdot \operatorname{poly}(n)\right)$, where $|X|=n$ and $\phi \approx 1.618$ is the golden ratio, but beyond this few positive results are known. This is frustrating and surprising, since a number of results link $d_{M P}$ to the well-studied TBR distance, henceforth denoted $d_{T B R}$. Namely, it has been proven that $d_{M P}$ is a lower bound on $d_{T B R}$ [18], which, informally, asks for the minimum number of topological rearrangement operations to transform one tree into the other; an empirical study has suggested that in practice the distances are often very close [23]. Also, $d_{M P}$ has been used to prove the tightness of the best-known kernelization results for $d_{T B R}$ [24,25]. What, exactly, is the relationship between $d_{M P}$ and $d_{T B R}$ ? This is a pertinent question, which transcends the specifics of TBR distance because, crucially, $d_{T B R}$ can be characterized using the powerful maximum agreement forest abstraction.

Distances based on agreement forests have been intensively and successfully studied in recent years, as the use of the agreement forest abstraction almost always yields fixed parameter tractability and constant-factor approximation algorithms [10], many of which are effective in practice. We refer to [41,39,14,35] for recent overviews of the agreement forest literature, and books such as [15] for an introduction to fixed parameter tractability. In particular, $d_{T B R}$ can be computed in $O\left(3^{d_{T B R}} \cdot \operatorname{poly}(n)\right)$ time [13], permits a polynomial-time 3-approximation algorithm, and a kernel of size $11 d_{T B R}-9$ [25].

In contrast, prior to this paper very little was known about $d_{M P}$ : nothing was known about the approximability of $d_{M P}$; it was not known whether it is fixed parameter tractable (where $d_{M P}$ is the parameter); and, while, as mentioned above, it is known that $d_{M P} \leq d_{T B R}$, it remained unclear how much smaller $d_{M P}$ can be than $d_{T B R}$ in the worst case. Despite promising partial results it even remained unclear whether questions such as "Is $d_{M P} \geq k$ ?" can be solved in polynomial time when $k$ is a constant [8,23]. This is another important difference with distances such as $d_{T B R}$, where corresponding questions are trivially polynomial time solvable for fixed $k$. The apparent extra complexity of $d_{M P}$ seems to stem from the unusual max-min definition of the problem, and the fact that unlike $d_{T B R}$, which is based on topological rearrangements of subtrees, $d_{M P}$ is based only on characters.

In this article we take a significant step forward in understanding the deeper complexity of $d_{M P}$ and resolve all of the above questions. Our central result is that we prove that two common polynomial-time reduction rules encountered in phylogenetics, the subtree and chain reductions [1], are sufficient to produce a linear kernel for $d_{M P}$. This means that, after exhaustive application of these rules, which preserve $d_{M P}$, the reduced trees will have at most $\alpha \cdot\left(d_{M P}+1\right)$ leaves, with $\alpha=560$. The fixed parameter tractability of computing $d_{M P}$ (parameterized by itself) then follows, by solving the kernel using the exact algorithm from [26]. The fact that the reduction rules preserve $d_{M P}$ was already known [23]. However, proving the bound on the size of the reduced trees requires rather involved combinatorial arguments, which have a very different flavour to the arguments typically encountered in the maximum agreement forest literature. The main goal of this article is to present these arguments as clearly as possible, rather than to optimize the resulting constants.

The kernel confirms that questions such as "Is $d_{M P} \geq k$ ?" can, indeed, be solved in polynomial time: it is striking that here the proof of fixed parameter tractability has preceded the weaker result of polynomial-time solveability for fixed $k$.

Next, by producing a modified, constructive version of the bounding argument underpinning the kernelization, we are able to demonstrate a polynomial-time $\alpha(1+1 / r)$-factor approximation algorithm for computation of $d_{M P}$ for any constant $r$, placing the problem in APX.

A number of other powerful corollaries result from the kernelization. We leverage the fact that the reduction rules also preserve $d_{T B R}$, to show that $1 \leq \frac{d_{T B R}}{d_{M P}} \leq 2 \alpha$, which limits how much smaller $d_{M P}$ can be than $d_{T B R}$. Subsequently, we show that the treewidth of an auxiliary graph structure known as the display graph [11] is bounded by a linear function of $d_{M P}$, resolving an open question posed several times [28,23]. The treewidth bound, and the existence of a non-trivial approximation algorithm for $d_{M P}$, were specified as sufficient conditions for proving the fixed parameter tractability of $d_{M P}$ via Courcelle's Theorem [23]; our linear kernel implies them. Summarising, our central result shows how kernelization can open the gateway to a host of strong auxiliary results and bypass intermediate steps in the algorithm design process.

The structure of the paper is as follows. In Section 2 we give formal definitions and insightful preliminary results. In Section 3 we prove our main result: the linear kernel. The section starts with Subsection 3.1 that gives a high-level overview of how a sequence of lemmas and theorems lead to the kernel, whereas in the rest of the section these lemmas and theorems are proved. Interesting corollaries of the existence of a linear kernel are derived in Section 4: A constant approximation algorithm in Section 4.1; A bound on the ratio between $d_{M P}$ and $d_{T B R}$ in Section 4.2; A bound on the

(a) $T_{1}$

(b) $T_{2}$

Fig. 1. Two unrooted binary phylogenetic trees $T_{1}, T_{2}$ on $X=\{a, \ldots, g\}$. Solid edges are monochromatic and dashed edges are bichromatic under an optimal extension for the character $\chi: X \rightarrow\{$ RED, BLUE $\}$, where $\chi(a)=\chi(b)=\chi(c)=$ RED, $\chi(d)=\chi(e)=\chi(f)=\chi(g)=$ bLUE. As there is one bichromatic edge in $T_{1}$ and two in $T_{2}$, we have that $l_{\chi}\left(T_{1}\right)=1, l_{\chi}\left(T_{2}\right)=2$, proving that $d_{M P}\left(T_{1}, T_{2}\right) \geq|1-2|=1$. In fact, it can be verified that no character can cause the parsimony scores of these two trees to differ by more, so $d_{M P}\left(T_{1}, T_{2}\right)=1$. We will show in Section 4.2 that $d_{T B R}\left(T_{1}, T_{2}\right)=2$, because a maximum agreement forest of these two trees contains three blocks [23]. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)
treewidth of the so-called display graph in terms of $d_{M P}$ in Section 4.3. Section 5 concludes with some directions for future research.

## 2. Definitions and preliminaries

An unrooted binary phylogenetic tree on a set of species (or taxa) $X$ is an undirected tree in which all internal vertices have degree 3, and the degree-1 vertices (the leaves) are bijectively labelled with elements from $X$. For brevity we will refer to unrooted binary phylogenetic trees as phylogenetic trees, or even shorter trees. See Fig. 1 for an example.

Given a set $S \subseteq X$ and a tree $T$ on $X$, we denote by $T[S]$ the spanning subtree on $S$ in $T$, that is, the minimal connected subgraph $T^{\prime}$ of $T$ such that $T^{\prime}$ contains every element of $S$. The induced subtree $\left.T\right|_{S}$ by $S$ in $T$ is the tree derived from $T[S]$ by suppressing any vertices of degree 2.

Given a subset $S \subseteq X$ and a tree $T$ on $X$, we say that $S$ has degree $d$ in $T$ if there are exactly $d$ edges $u v$ in $T$ for which $u$ is in $T[S]$ and $v$ is not; in other words, $d$ is the number of edges separating $T[S]$ from the rest of $T$. We call these edges pending edges of $S$ in $T$.

For two disjoint subsets $S_{1}, S_{2} \subseteq X$, we say $S_{1}$ and $S_{2}$ are spanning-disjoint in $T$ if the spanning subtrees $T\left[S_{1}\right]$ and $T\left[S_{2}\right.$ ] are edge-disjoint. (Observe that as $T$ is binary, this also implies that $T\left[S_{1}\right]$ and $T\left[S_{2}\right]$ are vertex-disjoint.) Similarly, we say a collection $S_{1}, \ldots S_{m}$ of subsets of $X$ are spanning-disjoint in $T$ if $S_{i}, S_{j}$ are spanning-disjoint in $T$ for any $i \neq j$.

### 2.1. Characters and parsimony

A character on $X$ is a function $\chi: X \rightarrow \mathbf{C}$, where $\mathbf{C}$ is a set of states. In this paper there is no limit on the size of $\mathbf{C}$, in contrast to some contexts where $|\mathbf{C}|$ is assumed to be quite small (for example, in genetic data the nucleobases $\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}$ ). Think of the states as colours, say $1,2, \ldots, t=:[t]$.

For a given character $\chi$ and tree $T$ on $X$, the parsimony score measures how well $T$ fits $\chi$. It is defined in the following way. Call a colouring $\phi: V(T) \rightarrow[t]$ an extension of $\chi$ to $T$ if $\phi(x)=\chi(x)$ for all $x \in X$. Denote by $\Delta_{T}(\phi)$ the number of bichromatic edges $u v$ in $T$, i.e. for which $\phi(u) \neq \phi(v)$. We usually omit subscript $T$ when the tree is clear from context. The parsimony score for $T$ with respect to $\chi$ is defined as

$$
l_{\chi}(T)=\min _{\phi} \Delta_{T}(\phi)
$$

where the minimum is taken over all possible extensions $\phi$ of $\chi$ to $T$. An extension $\phi$ that achieves this bound is called an optimal extension of $\chi$ to $T$. An optimal extension, and thus the parsimony score, can be easily computed in polynomial time using dynamic programming or e.g. Fitch's algorithm [19].

Observe that for any $T$ and $\chi$, the parsimony score for $T$ with respect to $\chi$ is at least $|\chi(X)|-1$, i.e. the number of colours assigned by $\chi$ minus 1 . If $l_{\chi}(T)$ is exactly $|\chi(X)|-1$, we say that $T$ is a perfect phylogeny for $\chi$. For trees $T_{1}, T_{2}$ and a character $\chi$ on $X$, the parsimony distance with respect to $\chi$ is defined as

$$
d_{M P \chi}\left(T_{1}, T_{2}\right)=\left|l_{\chi}\left(T_{1}\right)-l_{\chi}\left(T_{2}\right)\right|
$$

Now we are ready to define the maximum parsimony distance between two trees (see also Fig. 1). For two trees $T_{1}, T_{2}$ on $X$, the maximum parsimony distance is defined as

$$
d_{M P}\left(T_{1}, T_{2}\right)=\max _{\chi} d_{M P \chi}\left(T_{1}, T_{2}\right)
$$

where the maximum is taken over all possible characters $\chi$ on $X[18,31]$. Equivalently, we may write it as

$$
d_{M P}\left(T_{1}, T_{2}\right)=\max _{\chi}\left|\Delta\left(\phi_{1}\right)-\Delta\left(\phi_{2}\right)\right|
$$

where $\phi_{1}$ is an optimal extension of $\chi$ to $T_{1}$, and $\phi_{2}$ an optimal extension of $\chi$ to $T_{2}$. This measure satisfies the properties of a distance metric on the space of unrooted binary phylogenetic trees [18,31]. For two trees on $n$ taxa it is known that $d_{M P}$ is at most $n-2 \sqrt{n}+1$ [18]. A weaker bound of $n-1$ is easily obtained by observing that the parsimony score of a character on a tree is at least 0 and at most $n-1$.

Given a tree $T$ on $X$ and a colouring $\phi: V(T) \rightarrow[t]$, the forest induced by $\phi$ is derived from $T$ by deleting every bichromatic edge under $\phi$. Observe that the number of connected components in the forest induced by $\phi$ is exactly $\Delta(\phi)+1$.

Lemma 1. If $\chi: X \rightarrow[t]$ is a character with $S_{i}=\chi^{-1}(i) \neq \emptyset$ (i.e. at least one taxa is coloured $i$ ) for each $i \in[t]$, and $T$ is a tree on $X$, then

$$
l_{T}(\chi) \geq t-1
$$

with equality if and only if $S_{1}, \ldots S_{t}$ are spanning-disjoint in $T$.
Proof. To see that $l_{T}(\chi) \geq t-1$, consider an optimal extension $\phi$ of $\chi$ to $T$, and let $F$ be the forest induced by $\phi$. As each connected component in $F$ is monochromatically coloured by $\phi$, there must be at least $t$ connected components, and thus $\Delta(\phi) \geq t-1$, which implies $l_{\chi}(T) \geq t-1$.

Now suppose that $S_{1}, \ldots, S_{t}$ are spanning-disjoint in $T$. Then construct an extension $\phi$ of $\chi$ to $T$ by first setting $\phi(u)=i$ for every vertex $u$ in $T\left[S_{i}\right]$, for each $i \in[t]$. (As the spanning trees are edge-disjoint and thus vertex-disjoint in $T$, this is welldefined). For any remaining unassigned vertices $v$, if $v$ has a neighbour $u$ for which $\phi(u)$ is defined, then set $\phi(v)=\phi(u)$. Repeat this process until every vertex is assigned a colour by $\phi$. Now observe that by construction, the vertices assigned colour $i$ by $\phi$ form a connected subtree for each $i \in[t]$. Thus the forest induced by $\phi$ has exactly $t$ connected components, and so $\Delta(\phi)=t-1$.

Finally, suppose $l_{\chi}(T)=t-1$, and let $\phi$ be an optimal extension of $\chi$. Then the forest $F$ induced by $\phi$ has exactly $t$ connected components, which implies by the pigeonhole principle that each $S_{i}$ is a subset of one connected component in $F$. Then as each $S_{i}$ is contained within a different connected component of $F$, the spanning trees $T\left[S_{i}\right]$ are also contained within these components, and so $S_{1}, \ldots S_{t}$ are spanning-disjoint.

### 2.2. Parameterized complexity and kernelization

A parameterized problem is a problem for which the inputs are of the form $(x, k)$, where $k$ is an non-negative integer, called the parameter. A parameterized problem is fixed-parameter tractable (FPT) if there exists an algorithm that solves any instance $(x, k)$ in $f(k) \cdot|x|^{O(1)}$ time, where $f()$ is a computable function depending only on $k$. A parameterized problem has a kernel of size $g(k)$, where $g()$ is a computable function depending only on $k$, if there exists a polynomial time algorithm transforming any instance ( $x, k$ ) into an equivalent problem ( $x^{\prime}, k^{\prime}$ ), with $\left|x^{\prime}\right|, k^{\prime} \leq g(k)$. If $g(k)$ is a polynomial in $k$ then we call this a polynomial kernel; if $g(k)=O(k)$ then it is a linear kernel. It is well-known that a parameterized problem is fixed-parameter tractable if and only if it has a (not necessarily polynomial) kernel. For more information, we refer the reader to [15].

For a maximization problem $\Pi$ and $\rho \geq 1$, we say $\Pi$ has a constant factor approximation with approximation ratio $\rho$ if there exists a polynomial-time algorithm such that for any instance $\pi$ of $\Pi$, the following inequalities hold, where opt $(\pi)$ denotes the maximum value of a solution to $\pi$, and $\operatorname{alg}(\pi)$ denotes the value of the solution to $\pi$ returned by the algorithm:

$$
1 \leq \frac{o p t(\pi)}{\operatorname{alg}(\pi)} \leq \rho
$$

In this paper we study the following maximization problem:

## Maximum Parsimony Distance (dmp)

Input: Two trees $T_{1}, T_{2}$ on a set of taxa $X$.
Output: A character $\chi$ on $X$ that maximizes $\left|l_{\chi}\left(T_{1}\right)-l_{\chi}\left(T_{2}\right)\right|$.

## 3. Kernel bound

### 3.1. Overview

In this section we give an overview of the constituent parts of our kernelization result, and how they fit together.
The first step is to apply two reduction rules, the Cherry rule and the Chain rule, described in the next section. These rules correspond roughly to reduction rules that often appear in papers on computational phylogenetics. The correctness of these rules was proved in [23]; our contribution is to show that the exhaustive application of these rules grants a linear kernel, as stated in the following theorem.

Theorem 1. There exists a constant $\alpha(\alpha=560)$ for which the following holds. Let $\left(T_{1}, T_{2}\right)$ be a pair of binary unrooted phylogenetic trees on $X$ that are irreducible under Reduction Rules 1 and 2.

Then if $|X| \geq \alpha k$, it holds that $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, and we can find a witnessing character, i.e. a character $\chi$ yielding $d_{M P} \chi\left(T_{1}, T_{2}\right) \geq$ $k$, in polynomial time.

This theorem, together with the correctness of the reduction rules as proved in [23], immediately implies a linear kernel for DMP.

To show how we prove the theorem, we will need to introduce some terminology as we go.
A quartet $Q$ is any set of 4 elements in $X$. If $\left.T_{1}\right|_{Q} \neq\left. T_{2}\right|_{Q}$, we say that $Q$ is a conflicting quartet for $\left(T_{1}, T_{2}\right)$.
As a crucial step we prove that for any $S$ large enough with respect to the degree of $S$ in both $T_{1}$ and $T_{2}$, either there exists a conflicting quartet or one of the reduction rules applies.

Lemma 2. Let $S$ be a subset of $X$ with $d_{1}$ the degree of $S$ in $T_{1}$, and $d_{2}$ the degree of $S$ in $T_{2}$. If $|S|>9\left(d_{1}+d_{2}\right)-12$, then either $T_{1}\left|s \neq T_{2}\right| s$ or one of Reduction Rules 1 or 2 applies to ( $T_{1}, T_{2}$ ). In particular if ( $T_{1}, T_{2}$ ) is irreducible under Rules 1 or 2 and $|S| \geq$ $9\left(d_{1}+d_{2}\right)-11$, then there exists a conflicting quartet $Q \subseteq S$, and such a quartet can be found in polynomial time.

The next result implies that if we have a large enough number of conflicting quartets that are also spanning-disjoint in both $T_{1}$ and $T_{2}$, then we are done. While it is intuitively clear that such quartets can be leveraged to create a high parsimony score in one tree, some care has to be taken to keep the parsimony score low in the other tree.

Lemma 3. Let $\mathcal{Q}=\left\{Q_{1}, \ldots, Q_{k}\right\}$ be a set of conflicting quartets for $T_{1}, T_{2}$, such that $Q_{1}, \ldots Q_{k}$ are spanning-disjoint in $T_{1}$ and in $T_{2}$.

Then $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, and we can find a witnessing character in polynomial time.
In combination, Lemmas 2 and 3 allow us to show that $d_{M P}\left(T_{1}, T_{2}\right) \geq k$ provided that we can find at least $k$ sets $S_{1}, \ldots S_{k}$ that are spanning-disjoint in both trees and satisfy the conditions of Lemma 2.

We will find $k$ such sets as part of the construction of a character that witnesses $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, for any reduced instance with $|X| \geq \alpha k$. In order to construct this character, we first create a partition of $X$ into large subsets, as described by the following lemma.

Lemma 4. Suppose that $|X| \geq 2 c t$ for some integers $c$ and $t$, and let $T_{1}$ be a phylogenetic tree on $X$.
Then in polynomial time we can construct a partition $S_{1}, \ldots, S_{t}$ of $X$ with $S_{1}, \ldots, S_{t}$ spanning-disjoint in $T_{1}$, such that $\left|S_{i}\right| \geq c$ for each $i$.

We note that there is a one-to-one correspondence between partitions and characters on $X$, in the following sense. Given a partition $S_{1}, \ldots S_{t}$ of $X$, we may define a character $\chi: X \rightarrow[t]$ such that $\chi(x)=i$ if $x \in S_{i}$, for each $i \in[t]$. Call such a character the character defined by $S_{1}, \ldots S_{t}$.

Thus let us consider the character $\chi$ on $X$ defined by the partition described by Lemma 4 . Since $S_{1}, \ldots S_{t}$ are spanningdisjoint in $T_{1}$, Lemma 1 tells that the parsimony score of $T_{1}$ with respect to $\chi$ is exactly $t-1$.

Lemma 5. Let $\chi$ be the character defined by the partition $S_{1}, \ldots, S_{t}$ where $S_{1}, \ldots, S_{t}$ are spanning-disjoint in $T_{1}$, let $d_{1}$, $d_{2}$ be positive integers such that $d_{1} d_{2}-d_{1}-d_{2}>0$, and assume

$$
t \geq\left\lceil\frac{\left(2 d_{1} d_{2}+d_{1}\right)}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil k
$$

Then either $d_{M P \chi}\left(T_{1}, T_{2}\right) \geq k$, or in polynomial time we can find a set of indices $i_{1}, \ldots i_{k^{\prime}}$ with $k^{\prime} \geq k$ such that:

- $S_{i_{1}}, \ldots S_{i_{k^{\prime}}}$ are spanning-disjoint in $T_{2}$ (as well as in $T_{1}$ );
- $S_{i_{j}}$ has degree at most $d_{1}$ in $T_{1}$ for each $j \in\left[k^{\prime}\right]$; and
- $S_{i_{j}}$ has degree at most $d_{2}$ in $T_{2}$ for each $j \in\left[k^{\prime}\right]$.

We will prove Theorem 1 by combining these results in the following way. Fix integers $d_{1}, d_{2}$ to be determined later. Assume ( $T_{1}, T_{2}$ ) is irreducible under Reduction Rules 1 and 2, and assume that

$$
|X| \geq 2 c t, \text { where } c=9\left(d_{1}+d_{2}\right)-11 \text { and } t \geq\left\lceil\frac{\left(2 d_{1} d_{2}+d_{1}\right)}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil k
$$

(this holds if $|X| \geq \alpha k$ ).
By Lemma 4, there exists a partition $S_{1}, \ldots S_{t}$ of $X$ with $S_{1}, \ldots S_{t}$ spanning-disjoint in $T_{1}$ and $\left|S_{i}\right| \geq c$ for each $i \in[t]$. Let $\chi$ be the character defined by this partition. If $d_{M P \chi}\left(T_{1}, T_{2}\right) \geq k$, we may return $\chi$. Otherwise, we may apply Lemma 5 to
get a set of indices $i_{1}, \ldots i_{k}$ such that $S_{i_{1}}, \ldots S_{i_{k}}$ are spanning-disjoint in $T_{2}$ (as well as in $T_{1}$ ), each $S_{i_{j}}$ has degree at most $d_{1}$ in $T_{1}$, and each $S_{i_{j}}$ has degree at most $d_{2}$ in $T_{2}$. But then each $S_{i_{j}}$ satisfies the conditions of Lemma 2, and therefore for each $j \in[k]$ there exists a conflicting quartet $Q_{j} \subseteq S_{i_{j}}$. Moreover, as $S_{i_{1}}, \ldots S_{i_{k}}$ are spanning-disjoint in $T_{1}$ and $T_{2}$, the quartets $Q_{1}, \ldots Q_{k}$ are also spanning-disjoint in $T_{1}$ and $T_{2}$. Then Lemma 3 implies that $d_{M P}\left(T_{1}, T_{2}\right) \geq k$.

By setting $d_{1}=4$ and $d_{2}=5$, we get that $\alpha=560$, giving the desired bound.
In the next subsections we prove each of these lemmas, and then the main theorem, in turn.

### 3.2. Reduction rules

We begin by stating the reduction rules for our kernelization result. In what follows, a pair ( $x, y$ ) with $x, y \in X$ is a cherry in a tree $T$ if there exists an internal vertex $u$ in $T$ adjacent to both $x$ and $y$. A cherry is also sometimes known in the literature as a sibling-pair. A sequence of leaves $x_{1}, \ldots x_{r} \in X$ is a chain in $T$ if there exists a path of internal vertices $p_{1}, \ldots, p_{r}$ (possibly with $p_{1}=p_{2}$ and possibly with $p_{r-1}=p_{r}$ ), such that for each $i \in[r] p_{i}$ is the internal vertex adjacent to $x_{i}$. We call $r$ the length of this chain.

Reduction Rule 1. [Cherry reduction rule] If there exist $x, y \in X$ such that $(x, y)$ is a cherry in each of $T_{1}, T_{2}$, then replace $\left(T_{1}, T_{2}\right)$ with $\left(\left.T_{1}\right|_{X \backslash\{x\}},\left.T_{2}\right|_{X \backslash\{x\}}\right)$.

Reduction Rule 2. [Chain reduction rule] If there exists a sequence of leaves $x_{1}, \ldots x_{r} \in X$ such that $x_{1}, \ldots x_{r}$ is chain in both $T_{1}$ and $T_{2}$, and $r \geq 5$, then replace $\left(T_{1}, T_{2}\right)$ with $\left(T_{1}\left|X \backslash\left\{x_{5}, \ldots, x_{r}\right\}, T_{2}\right|_{X \backslash\left\{x_{5}, \ldots x_{r}\right\}}\right)$ (thus, the common chain is reduced to length 4).

The correctness of these rules (in the sense that they preserve $d_{M P}$ ) was previously proved in [23].
Theorem 2. Let $\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ be an instance of DMP derived from $\left(T_{1}, T_{2}\right)$ by an application of Reduction Rules 1 or 2. Then

$$
d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{M P}\left(T_{1}, T_{2}\right)
$$

Correctness of the chain reduction rule follows from Theorem 3.1 in [23]. Correctness of the cherry reduction rule follows as a subcase of Theorem 4.1 in [23].

Our main contribution is to show that if an instance is reduced by these rules then its size is bounded by a linear function of $d_{M P}$.

### 3.3. Small degree sets

In this section we prove Lemma 2.

Lemma 2. Let $S$ be a subset of $X$ with $d_{1}$ the degree of $S$ in $T_{1}$, and $d_{2}$ the degree of $S$ in $T_{2}$. If $|S|>9\left(d_{1}+d_{2}\right)-12$, then either $T_{1}\left|s \neq T_{2}\right|_{s}$ or one of Reduction Rules 1 or 2 applies to ( $T_{1}, T_{2}$ ). In particular if ( $T_{1}, T_{2}$ ) is irreducible under Rules 1 or 2 and $|S| \geq 9\left(d_{1}+d_{2}\right)-11$, then there exists a conflicting quartet $Q \subseteq S$, and such a quartet can be found in polynomial time.

Proof. Since unrooted binary trees are characterized by their quartets [34, Theorem 6.3.5(iii)], the last statement of the lemma follows directly.

We will show that if $\left.T_{1}\right|_{S}=\left.T_{2}\right|_{S}$ and neither of the reduction rules applies to ( $T_{1}, T_{2}$ ), then $|S| \leq 9\left(d_{1}+d_{2}\right)-12$. This implies the main claim of the lemma. Let us denote $\left.T\right|_{S}=\left.T_{1}\right|_{S}=\left.T_{2}\right|_{s}$.

Consider the backbone graph of $\left.T\right|_{S}$ obtained by deleting all leaves (see Fig. 2 for an example). Let $P_{C}$ be the set of nodes having degree 1 on the backbone, which we refer to as parents of a cherry in $\left.T\right|_{s}$. Let $P_{L}$ be the set of nodes having degree 2 on the backbone, which we refer to as parents of a leaf of $\left.T\right|_{S}$. All remaining vertices on the backbone have degree 3 . Thus $|S|$, the total number of leaves of $\left.T\right|_{S}$ is $2\left|P_{C}\right|+\left|P_{L}\right|$. We call the path between any two odd degree vertices on the backbone, having internal nodes only in $P_{L}$, a side of the backbone.

First notice that for each cherry in $\left.T\right|_{S}$, there must exist in $T_{1}[S]$, the spanning tree on $S$ in $T_{1}$, or in $T_{2}[S]$ a node, incident to a pending edge of $S$, between at least one of its two leaves and its corresponding node in $P_{C}$. Otherwise Reduction Rule 1 can be applied. In particular this implies that $\left|P_{C}\right| \leq d_{1}+d_{2}$.

Thus at least $P_{C}$ of the $d_{1}+d_{2}$ pending edges must be used for "cutting" the cherries, each of them cutting 1 leaf of a cherry. Let us choose one such leaf from each cherry, and call these the cut-leaves.

After removing cut-leaves, every node in $P_{C}$ and $P_{L}$ is now the parent of 1 leaf in $T \mid s$. Every side of the backbone contains at most 4 vertices in $P_{C}$ and $P_{L}$, unless $T_{1}[S]$ or $T_{2}[S]$ has a node of a pending edge of $S$ or a node adjacent to a node of a pending edge on that side. We show that every such pending edge on a side may increase the number of $P_{L}$-nodes on that side by at most 5 (see Fig. 2). Indeed, suppose a side of the backbone has in total $d$ pending edges of $S$ in both $T_{1}$ and $T_{2}$, but more than $4+5 d$ nodes in $P_{L}$, i.e. at least $5(d+1)$. Then $\left.T\right|_{S}$ contains a chain of length $5(d+1)$,


Fig. 2. Example illustration of the backbone of $T\left|s=T_{1}\right| s=T_{2} \mid s$ within $T_{1}$ and $T_{2}$, where $S=\left\{s_{1}, \ldots, s_{29}\right\}$. Edges and vertices of the backbone are in bold. Observe that $\left.T\right|_{S}$ has the chain $s_{1}, \ldots, s_{9}$, but $\left(T_{1}, T_{2}\right)$ do not have a common chain of length greater than 4 , as the leaf $s_{5}$ has a sibling $a$ in $T_{2}$.
which we can split up into $d+1$ chains of length 5 . Clearly at least one of these chains has no pending edge in either $T_{1}$ or $T_{2}$, and so $T_{1}, T_{2}$ have a common chain of length 5 , a contradiction.

Thus the total number of nodes from $P_{C}$ and $P_{L}$ on a side is at most five times the number of pending edges of $S$ (in $T_{1}[S]$ or $\left.T_{2}[S]\right)$ on that side, plus 4 . Otherwise Reduction Rule 2 can be applied. Given that we already used $\left|P_{C}\right|$ pending edges for cutting the cherries, we have $d_{1}+d_{2}-\left|P_{C}\right|$ pending edges left to be distributed over the sides.

The number of sides on the backbone is the number of edges in an unrooted binary tree with $\left|P_{C}\right|$ leaves, which is $2\left|P_{C}\right|-3$. Therefore the total number of leaves of $\left.T\right|_{S}$ is

$$
\begin{aligned}
|S|=2\left|P_{C}\right|+\left|P_{L}\right| & \leq\left|P_{C}\right|+4\left(2\left|P_{C}\right|-3\right)+5\left(d_{1}+d_{2}-\left|P_{C}\right|\right) \\
& \leq 4\left|P_{C}\right|+5\left(d_{1}+d_{2}\right)-12 .
\end{aligned}
$$

Clearly, this attains its largest value if $\left|P_{C}\right|=d_{1}+d_{2}$, in which case $|S| \leq 9\left(d_{1}+d_{2}\right)-12$, as was to be proven.

### 3.4. Combining conflicting quartets

In this section we prove Lemma 3.

Lemma 3. Let $\mathcal{Q}=\left\{Q_{1}, \ldots, Q_{k}\right\}$ be a set of conflicting quartets for $T_{1}, T_{2}$, such that $Q_{1}, \ldots Q_{k}$ are spanning-disjoint in $T_{1}$ and in $T_{2}$.

Then $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, and we can find a witnessing character in polynomial time.
Proof. For a quartet $Q$ and tree $T$, we say that $\left.T\right|_{Q}=a b \mid c d$ if $Q=\{a, b, c, d\}$ and in $T$ the path between $a$ and $b$ is edgedisjoint from the path between $c$ and $d$. Without loss of generality, we may assume $Q_{i}=\left\{a_{i}, b_{i}, c_{i}, d_{i}\right\}, T_{1}\left|Q_{i}=a_{i} b_{i}\right| c_{i} d_{i}$ and $T_{2}\left|Q_{i}=a_{i} c_{i}\right| b_{i} d_{i}$ for each $i \in[k]$.

We will show how to build a character $\chi$ with two states, such that $l_{\chi}\left(T_{1}\right) \leq k$, and $l_{\chi}\left(T_{2}\right) \geq 2 k$. This shows that $d_{M P \chi}\left(T_{1}, T_{2}\right) \geq k$, as required.

The idea is to construct $\chi$ in such a way that, for each quartet $Q_{i}, \chi\left(a_{i}\right)=\chi\left(b_{i}\right) \neq \chi\left(c_{i}\right)=\chi\left(d_{i}\right)$. This will ensure that $l_{\chi}\left(T_{2}\right)$ is at least $2 k$, as $T_{2}$ will have at least $2 k$ edge-disjoint paths (from $a_{i}$ to $c_{i}$ and from $b_{i} d_{i}$, for each $i \in[k]$ ) that each require at least one change in state along some edge.

For each $Q_{i}$, let $e_{Q_{i}}$ denote an edge in $T_{1}$ such that in $T_{1}\left[Q_{i}\right], e_{i}$ is on the path that separates $\left\{a_{i}, b_{i}\right\}$ from $\left\{c_{i}, d_{i}\right\}$.
Now we construct a function $\phi: V\left(T_{1}\right) \rightarrow\{$ RED, BLUE $\}$ as follows. Start by choosing an arbitrary leaf in $T_{1}$, say without loss of generality $a_{1}$, and set $\phi\left(a_{1}\right)=$ RED. Now proceed as follows. For any edge $u v$ in $T_{1}$ such that $\phi(u)$ is defined but $\phi(v)$ is not, we set $\phi(v)=\phi(u)$, unless $u v=e_{Q_{i}}$ for some $i$. In that case, we set $\phi(v)=$ BLUE if $\phi(u)=$ RED, and set $\phi(v)=$ RED otherwise.

Now we can let $\chi$ be the restriction of $\phi$ to $X$. By construction, $\phi$ is an extension of $\chi$ to $T_{1}$ and $\Delta(\phi)=\left|e_{Q_{i}}: i \in[k]\right|=k$. This is enough to show that $l_{\chi}\left(T_{1}\right) \leq k$.

We now show that $\chi\left(a_{i}\right)=\chi\left(b_{i}\right) \neq \chi\left(c_{i}\right)=\chi\left(d_{i}\right)$, for each $i \in[k]$. To see this, consider the spanning tree $T_{1}\left[Q_{i}\right]$. By construction, $T_{1}\left[Q_{i}\right]$ contains the edge $e_{Q_{i}}$ and $e_{Q_{i}}$ separates $\left\{a_{i}, b_{i}\right\}$ from $\left\{c_{i}, d_{i}\right\}$. Let $u_{i}, v_{i}$ be the vertices of $e_{Q_{i}}$, with $u_{i}$ the vertex closer to $a_{i}$ and $b_{i}$. Note that $T_{1}\left[Q_{i}\right]$ cannot contain $e_{Q_{j}}$ for any $j \neq i$, as $T_{1}\left[Q_{i}\right]$ and $T_{1}\left[Q_{j}\right]$ are edge-disjoint. It follows that $u_{i}, a_{i} b_{i}$ are all assigned the same value by $\phi$ and $v_{i}, c_{i}, d_{i}$ are assigned the opposite value. Thus by definition of $\chi$, we have $\chi\left(a_{i}\right)=\chi\left(b_{i}\right)=\phi\left(u_{i}\right) \neq \phi\left(v_{i}\right)=\chi\left(c_{i}\right)=\chi\left(d_{i}\right)$.

It remains to observe that as $Q_{1}, \ldots Q_{k}$ are spanning-disjoint in $T_{2}$, the $a_{i}-c_{i}$ and $b_{i}-d_{i}$ paths in $T_{2}$ are pairwise edgedisjoint for all $i \in[k]$. Then as $\chi\left(a_{i}\right) \neq \chi\left(c_{i}\right)$ and $\chi\left(b_{i}\right) \neq \chi\left(d_{i}\right)$, there exist at least $2 k$ edges $u v$ in $T_{2}$ with $\phi_{2}(u) \neq \phi_{2}(v)$, for any extension $\phi_{2}$ of $\chi$ to $T_{2}$. It follows that $l_{\chi}\left(T_{2}\right) \geq 2 k$, and so $d_{M P}\left(T_{1}, T_{2}\right) \geq d_{M P}\left(T_{1}, T_{2}\right)=\left|l_{\chi}\left(T_{1}\right)-l_{\chi}\left(T_{2}\right)\right| \geq 2 k-k=k$.

Since each edge is processed at most once in the construction of $\chi$, it is clear that this construction takes polynomial time.

### 3.5. Constructing an initial partition

In this section we prove Lemma 4.
Lemma 4. Suppose that $|X| \geq 2 c t$ for some integers $c$ and $t$, and let $T_{1}$ be a phylogenetic tree on $X$.
Then in polynomial time we can construct a partition $S_{1}, \ldots, S_{t}$ of $X$ with $S_{1}, \ldots, S_{t}$ spanning-disjoint in $T_{1}$, such that $\left|S_{i}\right| \geq c$ for each $i$.

Proof. We prove the claim by induction on $t$. For the base case, if $t=1$ then we may let $S_{1}=X$, and we have the desired partition.

For the inductive step, assume $|X| \geq 2 c t$ and that the claim is true for smaller values of $t$. We first fix an arbitrary rooting on $T_{1}$. That is, choose an arbitrary edge $e$ in $T_{1}$ and subdivide it with a new (temporary) vertex $r$, then orient all edges in $T_{1}$ away from $r$. Under this rooting, let $u$ be a lowest vertex in $T_{1}$ for which $u$ has at least $c$ descendants in $X$. Let $S_{t} \subseteq X$ be the set of these descendants. Note that since $T_{1}$ is binary, $\left|S_{t}\right|<2 c$, as otherwise one of the two children of $u$ would be a lower vertex with at least $c$ descendants.

Now consider the induced subtree $T_{1} \mid X^{\prime}$, where $X^{\prime}=X \backslash S_{t}$. As $\left|S_{t}\right|<2 c$, we have $X^{\prime} \geq 2 c(t-1)$. Then by the inductive hypothesis, we can construct a partition $S_{1}, \ldots, S_{t-1}$ of $X^{\prime}$ with $S_{1}, \ldots, S_{t-1}$ spanning-disjoint in $\left.T_{1}\right|_{X^{\prime}}$, such that $\left|S_{i}\right| \geq c$ for each $i$. By construction it is clear that $S_{t}$ is spanning-disjoint in $T_{1}$ from $S_{1}, \ldots, S_{t-1}$. Thus $S_{1}, \ldots, S_{t}$ is the desired partition.

As the construction of $S_{t}$ can be done in polynomial time and this process is repeated $t \leq|X|$ times, the entire process takes polynomial time.

### 3.6. Well-behaved sets

In this section we prove Lemma 5 . We start with an observation:
Observation 1. For any (not necessarily binary) unrooted tree $T$ with $n$ vertices, and any integer $d \geq 1$, the number of vertices in $T$ with degree strictly greater than $d$ is at most $n / d .{ }^{1}$

[^1]Proof. For each vertex $v$ in $T$ let $d(v)$ denote the degree of $v$. Recall that an unrooted tree with $n$ vertices has exactly $n-1$ edges. It follows that

$$
\sum_{v \in V(T)} d(v)=2|E(T)|=2 n-2
$$

Now suppose that $T$ has $m>n / d$ vertices with degree strictly greater than $d$, i.e. at least $d+1$. The remaining $n-m$ vertices all have degree at least 1 , from which it follows that

$$
\sum_{v \in V(T)} d(v) \geq m(d+1)+n-m=m d+n \geq(n / d) d+n=2 n,
$$

a contradiction.

Lemma 5. Let $\chi$ be the character defined by the partition $S_{1}, \ldots, S_{t}$ where $S_{1}, \ldots, S_{t}$ are spanning-disjoint in $T_{1}$, let $d_{1}$, $d_{2}$ be positive integers such that $d_{1} d_{2}-d_{1}-d_{2}>0$, and assume

$$
t \geq\left\lceil\frac{\left(2 d_{1} d_{2}+d_{1}\right)}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil k
$$

Then either $d_{M P \chi}\left(T_{1}, T_{2}\right) \geq k$, or in polynomial time we can find a set of indices $i_{1}, \ldots i_{k^{\prime}}$ with $k^{\prime} \geq k$ such that:

- $S_{i_{1}}, \ldots S_{i_{k^{\prime}}}$ are spanning-disjoint in $T_{2}$ (as well as in $T_{1}$ );
- $S_{i_{j}}$ has degree at most $d_{1}$ in $T_{1}$ for each $j \in\left[k^{\prime}\right]$; and
- $S_{i_{j}}$ has degree at most $d_{2}$ in $T_{2}$ for each $j \in\left[k^{\prime}\right]$.

Proof. By Lemma $1, l_{\chi}\left(T_{1}\right)=t-1$. If $l_{\chi}\left(T_{2}\right) \geq t+k-1$, then $d_{M P \chi}\left(T_{1}, T_{2}\right) \geq k$ as required. So we may assume that $l_{\chi}\left(T_{2}\right) \leq t+k-2$. Let $\delta=l_{\chi}\left(T_{2}\right)-l_{\chi}\left(T_{1}\right)$, and observe that $0 \leq \delta \leq k-1$.

We now construct a partition $P_{1}, \ldots P_{s}$ of $X$ which is spanning-disjoint in $T_{2}$ (see Fig. 3 for an illustration). Let $\phi_{2}$ be an optimal extension of $\chi$ to $T_{2}$. As $l_{\chi}\left(T_{2}\right)=l_{\chi}\left(T_{1}\right)+\delta=t+\delta-1$, the forest induced by $\phi_{2}$ has exactly $s$ monochromatic connected components, where $s=t+\delta$. Let $P_{1}, \ldots, P_{s}$ be the partition of $X$ formed by taking the intersection of $X$ with the vertex set of each tree in this forest. Observe that by construction $P_{1}, \ldots P_{s}$ are spanning-disjoint in $T_{2}$, and that furthermore each $P_{j}$ is a subset of $S_{i}$ for some $i \in[t]$ (as each element of $P_{j}$ is assigned the same value by $\phi_{2}$, and thus by $\chi)$.

Now let $\mathcal{I} \subseteq[t]$ denote the set of indices $i$ in $[t]$ such that

- $S_{i}=P_{j}$ for some $j \in[s]$;
- $S_{i}$ has degree at most $d_{1}$ in $T_{1}$; and
- $S_{i}$ has degree at most $d_{2}$ in $T_{2}$.

Note that since $P_{1}, \ldots P_{j}$ are spanning-disjoint in $T_{2}$, the sets $\left\{S_{i}: i \in \mathcal{I}\right\}$ are also spanning-disjoint in $T_{2}$. Notice that it is sufficient to prove that $|\mathcal{I}| \geq k$, whence any subset of $k$ indices from $\mathcal{I}$ satisfies the lemma. We will prove this by providing upper bounds on the number of indices in $[t]$ that do not satisfy the conditions of $\mathcal{I}$.

Let $\mathcal{I}_{0}$ denote the set of indices $i \in[t]$ such that $P_{j} \neq S_{i}$ for any $j \in[s]$. We first claim that $\left|\mathcal{I}_{0}\right| \leq \delta$. Indeed, since every $P_{j}$ is a subset of some $S_{i}$ and $S_{1}, \ldots S_{t}$ and $P_{1}, \ldots, P_{s}$ are both partitions of $X$, we have that for every $i \in \mathcal{I}_{0}$, there exist at least two distinct indices $j, j^{\prime} \in[s]$ for which $P_{j}, P_{j^{\prime}} \subset S_{i}$. Hence,

$$
s \geq 2\left|\mathcal{I}_{0}\right|+\left|[t] \backslash \mathcal{I}_{0}\right|=t+\left|\mathcal{I}_{0}\right|
$$

Therefore if $\left|\mathcal{I}_{0}\right|>\delta$ then $s>t+\delta$, contradicting the definition of $s$. Thus, we have $\left|\mathcal{I}_{0}\right| \leq \delta$.
Next, let $\mathcal{I}_{>d_{1}}$ denote the set of indices $i \in[t]$ for which $S_{i}$ has degree greater than $d_{1}$ in $T_{1}$. We will show that $\left|\mathcal{I}_{>d_{1}}\right| \leq t / d_{1}$. For each $i \in[t]$, compress the spanning subtree $T_{1}\left[S_{i}\right]$ to a single vertex, and observe that the degree of this vertex is equal to the degree of $S_{i}$ in $T_{1}$. Any vertex $u$ which is not part of any $T_{1}$ [ $S_{i}$ ] is merged with one of its neighbours. Note that this merging process can only increase the degrees of the remaining vertices. Call the resulting tree $T_{1}^{\prime}$. See Fig. 4. $T_{1}^{\prime}$ has $t$ vertices, each of them corresponding to a subset $S_{i}$, and having degree at least the degree of the corresponding $S_{i}$ in $T_{1}$. Now by Observation 1, there are at most $t / d_{1}$ vertices in $T_{1}^{\prime}$ with degree greater than $d_{1}$. It follows that there are at most $t / d_{1}$ values of $i \in[t]$ for which $S_{i}$ has degree greater than $d_{1}$ in $T_{1}$, and thus $\left|\mathcal{I}_{>d_{1}}\right| \leq t / d_{1}$ as we wanted to show.

Similarly let $\mathcal{J}_{>d_{2}}$ denote the set of indices $j \in[s]$ for which $P_{j}$ has degree greater than $d_{2}$ in $T_{2}$. By similar arguments as used for $\mathcal{I}_{>d_{1}}$ above, we can show that $\left|\mathcal{J}_{>d_{2}}\right| \leq s / d_{2}$.

Notice that for any $i \in[t]$, if $i$ is not in $\mathcal{I}$, then either $i \in \mathcal{I}_{0}$, or $i \in \mathcal{I}_{>d_{1}}$, or there exists $j \in \mathcal{J}_{>d_{2}}$ such that $S_{i}=P_{j}$. We therefore have that

$$
|\mathcal{I}| \geq t-\left|\mathcal{I}_{0}\right|-\left|\mathcal{I}_{>d_{1}}\right|-\left|\mathcal{J}_{>d_{2}}\right| \geq t-\delta-t / d_{1}-s / d_{2}
$$


(a) The initial partition $S_{1}=\{a, b, c, d\}, S_{2}=\{e, f, g, h\}, S_{3}=\{i, j, k, l\}$. Solid edges are monochromatic and dashed edges are bichromatic under an optimal extension of the character $\chi$ to $T_{1}$, where $\chi$ is the character corresponding to $S_{1}, S_{2}, S_{3}$.

(b) An optimal extension of $\chi$ to $T_{2}$. Note that the original partition $S_{1}, S_{2}, S_{3}$ is not spanningdisjoint in $T_{2}$, and as such, there are distinct monochromatic components assigned the same color (such as the monochromatic components on $\{a, d\}$ and $\{b, c\}$ ). The leaf sets of these components give us the new partition $P_{1}=\{a, d\}, P_{2}=\{b, c\}, P_{3}=\{e, f, g, h\}, P_{4}=\{i, j, l\}, P_{5}=\{k\}$.

Fig. 3. Illustration of the construction of partition $P_{1}, P_{2}, P_{3}, P_{4}, P_{5}$ from $S_{1}, S_{2}, S_{3}$. Solid edges are monochromatic and dashed edges are bichromatic under an optimal extension for $\chi$, where $\chi$ is the character induced by $S_{1}, S_{2}, S_{3}$.


Fig. 4. Illustration of the construction of auxiliary tree $T_{1}^{\prime}$, given a partition of $X$ with $S_{1}=\{a, b, c\}, S_{2}=\{d, e, f\}, S_{3}=\{g, h, i\}, S_{4}=\{j, k\}, S_{5}=\{l, m\}$. Note that the internal vertex labelled $u$ is not part of $T_{1}\left[S_{i}\right]$ for any $i$, so we merge it with an arbitrary adjacent vertex. In this case we merge $u$ into $S_{1}=\{a, b, c\}$, which is why $S_{1}$ has degree 1 in $T_{1}$ but degree 2 in $T_{1}^{\prime}$.

Now, using that $t \geq \frac{\left(2 d_{1} d_{2}+d_{1}\right)}{d_{1} d_{2}-d_{1}-d_{2}} k, s=t+\delta$ and $\delta \leq k-1$, we have:

$$
\begin{aligned}
|\mathcal{I}| & \geq t-\left|\mathcal{I}_{0}\right|-\left|\mathcal{I}_{>d_{1}}\right|-\left|\mathcal{J}_{>d_{2}}\right| \\
& \geq t-\delta-t / d_{1}-s / d_{2} \\
& =t-\delta-t / d_{1}-(t+\delta) / d_{2} \\
& =\frac{d_{1} d_{2} t-d_{1} d_{2} \delta-d_{2} t-d_{1} t-d_{1} \delta}{d_{1} d_{2}} \\
& =\frac{\left(d_{1} d_{2}-d_{1}-d_{2}\right) t-\left(d_{1} d_{2}+d_{1}\right) \delta}{d_{1} d_{2}} \\
& \geq \frac{\left(d_{1} d_{2}-d_{1}-d_{2}\right) t-\left(d_{1} d_{2}+d_{1}\right)(k-1)}{d_{1} d_{2}} \\
& \geq \frac{\left(2 d_{1} d_{2}+d_{1}\right) k-\left(d_{1} d_{2}+d_{1}\right)(k-1)}{d_{1} d_{2}} \\
& =\frac{d_{1} d_{2} k+d_{1} d_{2}+d_{1}}{d_{1} d_{2}} \\
& >\frac{d_{1} d_{2} k}{d_{1} d_{2}} \\
& =k
\end{aligned}
$$

as we needed to prove. To see that $\mathcal{I}$ can be constructed in polynomial time, it suffices to observe that the partition $P_{1}, \ldots, P_{s}$ can be constructed in polynomial time (as the $\phi_{2}$ can be found in polynomial time), and after this each $S_{i}$ can be checked for membership in $\mathcal{I}$ in polynomial time.

### 3.7. Proof of Theorem 1

Lemma 6. Let $d_{1}, d_{2}$ be positive integers such that $d_{1} d_{2}-d_{1}-d_{2}>0$. Let $\left(T_{1}, T_{2}\right)$ be a pair of binary unrooted phylogenetic trees on $X$ that are irreducible under Reduction Rules 1 and 2.

Then if $|X| \geq 2 c t$, where $c=9\left(d_{1}+d_{2}\right)-11$ and $t=\left\lceil\frac{\left(2 d_{1} d_{2}+d_{1}\right)}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil k$, it holds that $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, and we can find a witnessing character in polynomial time.

Proof. By Lemma 4, there exists a partition $S_{1}, \ldots S_{t}$ of $X$, all spanning-disjoint in $T_{1}$, and with $\left|S_{i}\right| \geq c$ for all $i \in[t]$. Let $\chi$ be the character defined by $S_{1}, \ldots, S_{t}$. If $\chi$ is a witness to $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, then we may return $\chi$ and we are done. Otherwise, we may apply Lemma 5 to find indices $i_{1}, \ldots i_{k}$ such that:

- $S_{i_{1}}, \ldots S_{i_{k}}$ are all spanning-disjoint in $T_{2}$ (as well as in $T_{1}$ );
- each $S_{i_{j}}$ has degree at most $d_{1}$ in $T_{1}$; and
- each $S_{i_{j}}$ has degree at most $d_{2}$ in $T_{2}$.

Now for each $S_{i_{j}}$, we have that $S_{i_{j}}$ has degree $d_{1}^{j} \leq d_{1}$ in $T_{1}$ and $d_{2}^{j} \leq d_{2}$ in $T_{2}$, and that

$$
\left|S_{i_{j}}\right| \geq c>9\left(d_{1}+d_{2}\right)-11 \geq 9\left(d_{1}^{j}+d_{2}^{j}\right)-11
$$

and also that $\left(T_{1}, T_{2}\right)$ is irreducible under Rules 1 and 2. Thus we may apply Lemma 2, to find a conflicting quartet $Q_{j} \subseteq S_{i_{j}}$ for each $i_{j}$.

Finally, as $S_{i_{1}}, \ldots S_{i_{k}}$ are spanning-disjoint in both $T_{1}$ and $T_{2}$, and as each $Q_{j}$ is a subset of $S_{i_{j}}$, we have that $Q_{1}, \ldots, Q_{k}$ are also spanning-disjoint in both $T_{1}$ and $T_{2}$. Therefore we may apply Lemma 3 to find a witnessing character for $d_{M P}\left(T_{1}, T_{2}\right) \geq k$. As each step of this process takes polynomial time, the construction of a witnessing character takes polynomial time.

It remains to complete the proof of Theorem 1.

Theorem 1. There exists a constant $\alpha(\alpha=560)$ for which the following holds. Let $\left(T_{1}, T_{2}\right)$ be a pair of binary unrooted phylogenetic trees on $X$ that are irreducible under Reduction Rules 1 and 2.

Then if $|X| \geq \alpha k$, it holds that $d_{M P}\left(T_{1}, T_{2}\right) \geq k$, and we can find a witnessing character, i.e. a character $\chi$ yielding $d_{M P} \chi\left(T_{1}, T_{2}\right) \geq$ $k$, in polynomial time.

Proof. The proof boils down to choosing the appropriate values of $d_{1}$ and $d_{2}$ such that $2 c t=\left(9\left(d_{1}+d_{2}\right)-11\right)$. $\left\lceil\frac{\left(2 d_{1} d_{2}+d_{1}\right)}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil k=\alpha k$. For $d_{1}=4, d_{2}=5$ we get $c=70$ and $t=4 k$, yielding the value of $\alpha=560$ for $\alpha k=2 c t$.

In the appendix, we show that $d_{1}=4, d_{2}=5$ is in fact the optimal choice of values for $d_{1}$ and $d_{2}$.
As a corollary to Theorem 1 and Theorem 2, we have that DMP is fixed-parameter tractable with respect to $d_{M P}$. Specifically, the kernel can be solved using the exponential-time algorithm described in [26], which computes the maximum parsimony distance of two trees on $n$ leaves in time $O\left(1.619^{n} \cdot \operatorname{poly}(n)\right)$.

Corollary 1. DMP has a kernel with at most $\alpha(k+1)$ taxa, and can be solved in time $O\left(1.619^{\alpha k} \cdot \operatorname{poly}(\alpha k)+p o l y(n)\right)$, with $k=$ $d_{M P}\left(T_{1}, T_{2}\right)$.

Proof. Given an instance ( $T_{1}, T_{2}$ ) of DMP where $T_{1}, T_{2}$ are trees on a set of taxa $X$, let $\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ be the instance derived from ( $T_{1}, T_{2}$ ) by exhaustively applying Reduction Rules 1 and 2 . As each reduction rule can be applied in polynomial time, and each application of the rule reduces the number of taxa, ( $T_{1}^{\prime}, T_{2}^{\prime}$ ) can be derived in polynomial time. Moreover $\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ is irreducible under Rules 1 and 2, and by Theorem 2, $d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{M P}\left(T_{1}, T_{2}\right)=k$.

Let $X^{\prime}$ be the leaf set of ( $T_{1}^{\prime}, T_{2}^{\prime}$ ), and suppose for a contradiction that $\left|X^{\prime}\right| \geq \alpha(k+1)$. Then Theorem 1 implies that $d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \geq k+1$, a contradiction. Thus $\left|X^{\prime}\right|<\alpha(k+1)$, and ( $\left.T_{1}^{\prime}, T_{2}^{\prime}\right)$ is the desired kernel.

To see that DMP can be solved in time $O\left(1.619^{\alpha k} \cdot \operatorname{poly}(\alpha k)+\operatorname{poly}(n)\right)$, recall that DMP has a simple exponential-time algorithm with running time $O\left(\phi^{n} \cdot \operatorname{poly}(n)\right)$, where $|X|=n$ and $\phi<1.619$ is the golden ratio [26]. Applying this algorithm to our kernel, we get an algorithm with running time

$$
\operatorname{poly}(n)+O\left(1.619^{\alpha(k+1)} \cdot \operatorname{poly}(\alpha(k+1))\right)=O\left(1.619^{\alpha k} \cdot \operatorname{poly}(\alpha k)+\operatorname{poly}(n)\right)
$$

For completeness, we clarify that these results also prove that the decision problems " $d_{M P} \leq k$ ?", " $d_{M P} \geq k ?$ " and " $d_{M P}=$ $k$ ?" can all be answered in time $f(k) \cdot p o l y(n)$. To answer " $d_{M P} \leq k$ ?", note that if the kernel has size at least $\alpha(k+1)$ the answer is definitely NO, and otherwise the algorithm from [26] can be applied to compute $d_{M P}$ directly; this can then be compared to $k$ to resolve the question. The " $d_{M P} \geq k$ ?" question can be answered by asking " $d_{M P} \leq k-1$ ?" and negating the answer; and " $d_{M P}=k$ ?" can be answered by combining the answers to the $\leq$ and $\geq$ questions.

## 4. Corollaries: leveraging the kernel

### 4.1. A polynomial-time constant-factor approximation algorithm for DMP

We present how a constant factor approximation algorithm for DMP can be designed using Theorem 1 together with Reduction Rules 1 and 2.

In order to incorporate Reduction Rules 1 and 2 into our approximation algorithm, we require a way to construct a witnessing character for the original instance from a witnessing character for the reduced instance.

Lemma 7. Let $\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ be an instance of DMP derived from $\left(T_{1}, T_{2}\right)$ by an application of Reduction Rule 1 or 2, with $T_{1}^{\prime}$, $T_{2}^{\prime}$ trees on $X^{\prime} \subset X$. Then given a character $\chi^{\prime}$ on $X^{\prime}$, we can derive a character $\chi$ on $X$ in polynomial time such that $d_{M P}\left(T_{1}, T_{2}\right) \geq$ $d_{M P \chi^{\prime}}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$.

Proof. First observe that by definition of the reduction rules, we may assume that $T_{1}^{\prime}=\left.T_{1}\right|_{X^{\prime}}$ and $T_{2}^{\prime}=\left.T_{2}\right|_{X^{\prime}}$ for some $X^{\prime} \subseteq X$. Assume without loss of generality that $l_{\chi^{\prime}}\left(T_{2}^{\prime}\right) \geq l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)$, and let $\phi_{1}^{\prime}$ be an optimal extension of $\chi^{\prime}$ to $T_{1}^{\prime}$. We will now define a function $\phi: V\left(T_{1}\right) \rightarrow \mathbf{C}$ such that $\phi(u)=\phi^{\prime}(u)$ for all $u \in V\left(T_{1}^{\prime}\right)$, and such that $\Delta_{T_{1}}\left(\phi_{1}\right)=\Delta_{T_{1}^{\prime}}\left(\phi_{1}^{\prime}\right)=l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)$.

Recall that $T_{1} \mid X^{\prime}$ is derived from the spanning tree $T_{1}\left[X^{\prime}\right]$ by suppressing vertices of degree 2 , and therefore $T_{1}\left[X^{\prime}\right]$ can be derived from $T_{1}^{\prime}=\left.T_{1}\right|_{X^{\prime}}$ by repeatedly subdividing edges with degree- 2 vertices. Now construct $\phi_{1}$ as follows. For each vertex $v$ in $T_{1}^{\prime}$, set $\phi_{1}(v)=\phi_{1}^{\prime}(v)$. For every edge $e=u v$ that gets subdivided with one or more degree- 2 vertices, set $\phi_{1}\left(u^{\prime}\right)=\phi_{1}^{\prime}(u)$ for each such degree-2 vertex $u^{\prime}$. Thus, $\phi_{1}$ assigns a colour to every vertex in $T_{1}\left[X^{\prime}\right]$, and by construction $\Delta_{T_{1}\left[X^{\prime}\right]}\left(\phi_{1}\right)=\Delta_{T_{1}^{\prime}}\left(\phi_{1}^{\prime}\right)$.

In order to assign $\phi(v)$ to vertices $v$ of $T_{1}$ not in $T_{1}\left[X^{\prime}\right]$, take any edge $e=u v$ in $T_{1}$ such that $\phi_{1}(u)$ has been assigned but $\phi_{1}(v)$ has not, and set $\phi_{1}(v)=\phi_{1}(u)$. After completing this process, we have that $\phi_{1}$ assigns a colour to every vertex in $T_{1}$ (including its leaves) and $\Delta_{T_{1}}\left(\phi_{1}\right)=\Delta_{T_{1}^{\prime}}\left(\phi_{1}^{\prime}\right)$, as required.

Now let the character $\chi$ be the restriction of $\phi_{1}$ to $X$. Then by construction $\phi_{1}$ is an extension of $\chi$ on $X$, whence

$$
l_{\chi}\left(T_{1}\right) \leq \Delta_{T_{1}}\left(\phi_{1}\right)=l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)
$$

Moreover, we must have that $l_{\chi}\left(T_{1}\right) \geq l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)$ and thus $l_{\chi}\left(T_{1}\right)=l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)$. Indeed, if $\Delta_{T_{1}}(\phi)<\Delta_{T_{1}}\left(\phi_{1}\right)$ for some extension $\phi$ of $\chi$ to $T_{1}$, then by considering the restriction of $\phi$ to $T_{1}[S]$, we can see that $l_{\chi^{\prime}}\left(T_{1}^{\prime}\right) \leq \Delta_{T_{1}}(\phi)<\Delta_{T_{1}}\left(\phi_{1}\right)$, a contradiction as $\Delta_{T_{1}}\left(\phi_{1}\right)=\Delta_{T_{1}^{\prime}}\left(\phi_{1}^{\prime}\right)=l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)$.

Next we show that $l_{\chi}\left(T_{2}\right) \geq l_{\chi^{\prime}}\left(T_{2}^{\prime}\right)$. Consider any optimal extension $\phi_{2}$ of $\chi$ to $T_{2}$, and take the restriction $\phi_{2}^{\prime}$ of this function to $T_{2}^{\prime}=\left.T_{2}\right|_{X^{\prime}}$. Then clearly $\Delta_{T_{2}^{\prime}}\left(\phi_{2}^{\prime}\right) \leq \Delta_{T_{2}}\left(\phi_{2}\right)$ and therefore

$$
l_{\chi^{\prime}}\left(T_{2}^{\prime}\right) \leq \Delta_{T_{2}^{\prime}}\left(\phi_{2}^{\prime}\right) \leq \Delta_{T_{2}}\left(\phi_{2}\right)=l_{\chi}\left(T_{2}\right)
$$

Thus we have

$$
d_{M P \chi}\left(T_{1}, T_{2}\right) \geq l_{\chi}\left(T_{2}\right)-l_{\chi}\left(T_{1}\right) \geq l_{\chi^{\prime}}\left(T_{2}^{\prime}\right)-l_{\chi^{\prime}}\left(T_{1}^{\prime}\right)=d_{M P \chi^{\prime}}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)
$$

Theorem 3. For any positive integer $r$, given an instance $\left(T_{1}, T_{2}\right)$ of DMP, we can find in polynomial time a character $\chi$ such that

$$
1 \leq \frac{d_{M P}\left(T_{1}, T_{2}\right)}{d_{M P \chi}\left(T_{1}, T_{2}\right)} \leq(1+1 / r) \alpha
$$

where $\alpha=560$. That is, DMP has a constant factor approximation with approximation ratio $(1+1 / r) \alpha$.
Proof. First apply Reduction Rules 1 and 2 exhaustively, to derive an irreducible instance ( $T_{1}^{\prime}, T_{2}^{\prime}$ ). By Theorem 2 , $d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{M P}\left(T_{1}, T_{2}\right)$. Let $X^{\prime}$ be the leaf set of this reduced instance. Now let $k$ be the maximum integer such that $\left|X^{\prime}\right| \geq \alpha k$, where $\alpha=560$. If $k<r$, then we can determine a character $\chi^{\prime}$ for which $d_{M P} \chi^{\prime}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ exactly in time $O\left(1.619^{\alpha r} \cdot p o l y(n)\right)$ using the algorithm of [26]. Otherwise by Theorem 1, we can in polynomial time construct a character $\chi^{\prime}$ on $X^{\prime}$ such that $d_{M P} \chi^{\prime}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \geq k$. In either case, by Lemma 7 we can extend $\chi^{\prime}$ to a character $\chi$ on $X$ such that

$$
d_{M P \chi}\left(T_{1}, T_{2}\right) \geq d_{M P \chi^{\prime}}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \geq k
$$

We return $\chi$.
It remains to show that

$$
d_{M P}\left(T_{1}, T_{2}\right) /(1+1 / r) \alpha \leq d_{M P \chi}\left(T_{1}, T_{2}\right) \leq d_{M P}\left(T_{1}, T_{2}\right)
$$

from which the theorem follows. The second inequality is by definition of $d_{M P}\left(T_{1}, T_{2}\right)$. To see the first inequality: if $k<r$ then by construction

$$
d_{M P}\left(T_{1}, T_{2}\right)=d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{M P \chi^{\prime}}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \leq d_{M P \chi}\left(T_{1}, T_{2}\right)
$$

So now assume that $k \geq r$, and so by construction $d_{M P} \chi^{\prime}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \geq k \geq r$. As stated in the preliminaries, the number of taxa provides an upper bound on the $d_{M P}$ of any instance. Thus, $d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \leq\left|X^{\prime}\right|$. By choice of $k$, we have $\left|X^{\prime}\right|<\alpha(k+1)$. Then we have

$$
\begin{aligned}
d_{M P}\left(T_{1}, T_{2}\right) / \alpha & =d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) / \alpha \\
& \leq\left|X^{\prime}\right| / \alpha \\
& <\alpha(k+1) / \alpha=k+1 \\
& \leq d_{M P \chi}\left(T_{1}, T_{2}\right)+1 \\
& \leq(1+1 / r) d_{M P \chi}\left(T_{1}, T_{2}\right)
\end{aligned}
$$

Thus $d_{M P}\left(T_{1}, T_{2}\right) /(1+1 / r) \alpha \leq d_{M P}\left(T_{1}, T_{2}\right)$, as required.

### 4.2. Bounding the distance between $d_{T B R}$ and $d_{M P}$

Tree Bisection and Reconnection (TBR) distance, denoted $d_{T B R}$, is a distance measure defined on two unrooted binary phylogenetic trees $T_{1}, T_{2}$. It is defined as the minimum number of "TBR-moves" required to transform $T_{1}$ into $T_{2}$ (or viceversa): it is a metric [1]. Informally, a TBR-move consists of deleting an edge of a tree and then reconnecting the two resulting components via a new edge. This definition is motivated by the way software for constructing phylogenetic trees heuristically navigates through tree space in search of better trees [36]. However, for algorithmic and analytical purposes $d_{T B R}$ is most interesting because of its equivalence to the agreement forest abstraction. An agreement forest of $T_{1}$ and $T_{2}$ on the same set of taxa $X$ is a partition of $X$ into non-empty sets $S_{1}, S_{2} \ldots, S_{t}$ called blocks, such that: (1) for each $i, T_{1}\left|s_{i}=T_{2}\right| s_{i}$; (2) $S_{1}, S_{2}, \ldots, S_{t}$ are spanning-disjoint in $T_{1}$ and in $T_{2}$. An (unrooted) maximum agreement forest is an agreement forest with a minimum number of blocks, and $d_{T B R}\left(T_{1}, T_{2}\right)$ is equal to this minimum, minus 1 [1]. A maximum agreement forest for the two trees in Fig. 1 consists of three blocks $\{a, b\},\{f, g\}$ and $\{c, d, e\}$, so here $d_{T B R}$ is 2 .

The characterization of $d_{T B R}$ via agreement forests is significant, because agreement forests have opened the door to a large number of positive FPT and approximation results in the phylogenetics literature, and they have also attracted attention
from outside phylogenetics. We refer to $[41,16,39,14,10,35,4]$ for recent results. Moreover, a number of other problems have been shown to be FPT when parameterized by $d_{T B R}$, by leveraging properties of the $d_{T B R}$ kernel [23] and/or showing that, via agreement forests, the treewidth of a certain auxiliary graph structure is bounded by a function of $d_{T B R}$ [28] (see the next section). $d_{T B R}$ is a lower bound on many phylogenetic dissimilarity measurements [28], which helps to prove FPT results for these larger parameters, but what about $d_{M P}$ ? It has previously been shown that $d_{M P}\left(T_{1}, T_{2}\right) \leq d_{T B R}\left(T_{1}, T_{2}\right)$ for any pair of trees $T_{1}, T_{2}$ [18,31]. However, the possibility remained that $d_{M P}$ could be arbitrarily smaller than $d_{T B R}$, and this hinders our ability to bind $d_{M P}$ to other phylogenetic parameters. Our contribution is to show that $d_{M P}$ and $d_{T B R}$ are in fact within a constant factor of each other: $d_{T B R}\left(T_{1}, T_{2}\right) \leq 2 \alpha d_{M P}\left(T_{1}, T_{2}\right)$.

To show this, we use the fortunate fact that Reduction Rules 1 and 2, which we used to prove the kernel bound for DMP, preserve $d_{T B R}$ as well as $d_{M P}$ for $d_{T B R}$. The following theorem is, modulo a small modification, due to [1].

Theorem 4. Let $\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ be a pair of phylogenetic trees on $X^{\prime}$ derived from $\left(T_{1}, T_{2}\right)$ by an application of Reduction Rule 1 or 2 . Then

$$
d_{T B R}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{T B R}\left(T_{1}, T_{2}\right)
$$

Proof. Theorem 3.4 of [1] shows that $d_{T B R}$ is preserved under reduction rules similar to Reduction Rules 1 and 2 , except that common chains are reduced to length 3 instead of 4 . For a pair of trees $T_{1}, T_{2}$ on $X$, let ( $T_{1}^{\prime \prime}, T_{2}^{\prime \prime}$ ) with leaf set $X^{\prime \prime}$ be the instance derived from $\left(T_{1}, T_{2}\right)$ by exhaustively applying these reduction rules. Also let ( $T_{1}^{\prime}, T_{2}^{\prime}$ ) with leaf set $X^{\prime}$ be the instance derived from ( $T_{1}, T_{2}$ ) by exhaustively applying Reduction Rules 1 and 2 . Observe that we may assume $X^{\prime \prime} \subseteq X^{\prime} \subseteq X$, since any leaf deleted in an application of Reduction Rule 1 or 2 can also be deleted by an application of one of the reduction rules in [1]. Furthermore by Lemma 2.1 of [1], $d_{T B R}$ distance is non-increasing on subtrees induced by subsets of $X$, which implies that

$$
d_{T B R}\left(T_{1}^{\prime \prime}, T_{2}^{\prime \prime}\right) \leq d_{T B R}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \leq d_{T B R}\left(T_{1}, T_{2}\right)
$$

As Theorem 3.4 of [1] states that $d_{T B R}\left(T_{1}^{\prime \prime}, T_{2}^{\prime \prime}\right)=d_{T B R}\left(T_{1}, T_{2}\right)$, the chain of inequalities becomes a chain of equalities and hence $d_{T B R}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)=d_{T B R}\left(T_{1}, T_{2}\right)$.

Theorem 5. For any pair of phylogenetic trees $T_{1}, T_{2}$ such that $T_{1} \neq T_{2}$, whence $d_{M P}\left(T_{1}, T_{2}\right) \geq 1$,

$$
1 \leq \frac{d_{T B R}\left(T_{1}, T_{2}\right)}{d_{M P}\left(T_{1}, T_{2}\right)} \leq 2 \alpha
$$

Proof. Let $\left(T_{1}^{\prime}, T_{2}^{\prime}\right)$ be the pair of trees derived from $\left(T_{1}, T_{2}\right)$ by exhaustively applying Reduction Rules 1 and 2 , and let $X^{\prime}$ be the leaf set of $T_{1}^{\prime}$ and $T_{2}^{\prime}$. It is well-known that $d_{T B R}\left(T_{1}^{\prime}, T_{2}^{\prime}\right) \leq\left|X^{\prime}\right|-3$ [1]. Then by Theorems 1,2 and 4,

$$
\begin{aligned}
d_{T B R}\left(T_{1}, T_{2}\right) & =d_{T B R}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)<\left|X^{\prime}\right| \\
& <\alpha\left(d_{M P}\left(T_{1}^{\prime}, T_{2}^{\prime}\right)+1\right) \\
& \leq 2 \alpha d_{M P}\left(T_{1}, T_{2}\right)
\end{aligned}
$$

Using $d_{M P}\left(T_{1}, T_{2}\right) \leq d_{T B R}\left(T_{1}, T_{2}\right)$ [31, Lemma 2.1], we have

$$
d_{M P}\left(T_{1}, T_{2}\right) \leq d_{T B R}\left(T_{1}, T_{2}\right) \leq 2 \alpha d_{M P}\left(T_{1}, T_{2}\right)
$$

which, dividing by $d_{M P}\left(T_{1}, T_{2}\right)$, proves the theorem.

### 4.3. The treewidth of the display graph

Let $G=(V, E)$ be an undirected graph. A tree decomposition of $G$ consists of a multi-set of bags, $B=\left\{B_{1}, \ldots, B_{t}\right\}$ where each $B_{i} \subseteq V$, and a tree $T$ whose nodes are in bijection with $B$, such that: (1) Every vertex $v \in V$ is in at least one bag; (2) for every edge $\{u, v\}$, at least one bag contains both $u$ and $v$, and (3) for every vertex $v \in V$, the bags of $T$ that contain $v$ induce a connected subtree of $T$. The width of the tree decomposition is equal to the size of its largest bag, minus one, and the treewidth of $G$ is the minimum width, ranging over all tree decompositions $T$ of $G$ [7]. Treewidth derives its importance in combinatorial optimization from the fact that many NP-hard problems on graphs become fixed parameter tractable when parameterized by the treewidth of the graph [6].

Given two phylogenetic trees $T_{1}, T_{2}$ on $X$, where $|X| \geq 3$, the display graph of $T_{1}$ and $T_{2}$, denoted $D\left(T_{1}, T_{2}\right)$, is the graph obtained by identifying the leaves of $T_{1}$ and $T_{2}$ with the same label. A sequence of articles have studied the treewidth of display graphs, expressed as a function of various phylogenetic parameters, and used this to prove FPT results for a number of NP-hard phylogenetics problems using Courcelle's Theorem [11,28,21] and explicit dynamic programming algorithms running over tree decompositions of the display graph [5]. However, the question remained whether the treewidth of the display graph, denoted by $t w\left(D\left(T_{1}, T_{2}\right)\right)$ could be bounded by a function of $d_{M P}\left(T_{1}, T_{2}\right)$ [23].

The answer is emphatically yes: here we show, by leveraging the fact that $d_{M P}$ and $d_{T B R}$ are within a constant factor of each other, that the display graph has treewidth bounded by a linear function of $d_{M P}\left(T_{1}, T_{2}\right)$.

Theorem 6. For two phylogenetic trees $T_{1}, T_{2}$ on $X$,

$$
t w\left(D\left(T_{1}, T_{2}\right)\right) \leq 2 \alpha d_{M P}\left(T_{1}, T_{2}\right)+2
$$

Proof. It was shown in [28] that $t w\left(D\left(T_{1}, T_{2}\right)\right) \leq d_{T B R}\left(T_{1}, T_{2}\right)+2$. As Theorem 5 shows $d_{T B R}\left(T_{1}, T_{2}\right) \leq 2 \alpha d_{M P}\left(T_{1}, T_{2}\right)$ the theorem follows.

Note that Theorem 7.2 of [27] shows an infinite family of trees where the treewidth of the display graph is 3 but $d_{M P}$ is unbounded.

## 5. Conclusion

A natural question is how far the analysis can be tightened, or changed, to improve the existing bound on the size of the kernel. In any case, it can be shown that for these two reduction rules a bound smaller than $20 k-12$ is not possible. That is because the family of fully-reduced instances described in [24] have exactly $15 k-9$ taxa, where in this specific case $k=d_{T B R}=d_{M P}$. By replacing the length-3 chains with length-4 chains in this family we obtain the bound $20 k-12$. We expect that, in practice, the achieved reduction on realistic trees will be far superior to the bounds proven in this paper.

From the perspective of algorithm design it would be useful to design an explicit algorithm with FPT runtime that does not rely on kernelization; for example, by branching or by dynamic programming over an appropriately defined decomposition. Similarly, in the quest for small constant approximation factors it would be interesting to design polynomial-time approximation algorithms that do not rely on kernelization. It is unlikely that through kernelization we will be able to achieve such truly small constant ratios.

The precise relationship between $d_{M P}$ and $d_{T B R}$ remains intriguing. Although we have now established that they are within a constant factor of each other, we are still a long way from proving or disproving the conjecture that $d_{M P} \geq(1 / 2) d_{T B R}$ [23]. An infinite family of examples is known where $d_{M P}=(1 / 2) d_{T B R}+o(1)$ [31, Theorem 7.1], and small examples are known where $d_{M P}=(1 / 2) d_{T B R}$ (see e.g. Fig. 1 , based on [23, Figure 5$]$ ), so $d_{M P} \geq(1 / 2) d_{T B R}$ would be the best possible bound.

On a slightly different note, recent publications have reduced the $d_{T B R}$ kernel size from $28 k$ to $15 k-9$ [24], and then to $11 k-9$ [25]. The $11 k-9$ kernel augments the two reduction rules discussed in this article, with five new reduction rules. Which of these new reduction rules work (possibly in a modified form) for $d_{M P}$, and how might this help us obtain a smaller linear kernel for $d_{M P}$ ?

Finally, we note that there are several slight variations of $d_{M P}$ in the literature. These include the "asymmetric" version $d_{A M P}\left(T_{1}, T_{2}\right)$, defined as $\max _{\chi}\left(l_{\chi}\left(T_{1}\right)-l_{\chi}\left(T_{2}\right)\right)$, in which $T_{1}$ is required to have the higher parsimony score, and the "restricted states" version $d_{M P}^{2}\left(T_{1}, T_{2}\right):=\max _{\chi} d_{M P \chi}\left(T_{1}, T_{2}\right)$, where the maximum is taken over all characters with at most 2 states [28,22]. Many of the results in this article will go through for $d_{A M P}\left(T_{2}, T_{1}\right)$, as the characters we construct consistently give a larger score to $T_{2}$. It is less obvious how our results impact on $d_{M P}^{2}$. In particular, it is not immediately clear whether the reduction rules described in [23] go through for $d_{M P}^{2}$, or how one would prove an analogue of Lemma 5 for $d_{M P}^{2}$. Relatedly, it is unclear how much smaller $d_{M P}^{2}$ can be than $d_{M P}$ itself. Specifically, how important are additional states when attempting to maximize the parsimony distance between trees? It is known that $7 d_{M P}-5$ states are sufficient to obtain a character that witnesses $d_{M P}$ [8], but it is unclear what happens below this bound.

## CRediT authorship contribution statement

Mark Jones: Conceptualization, Methodology, Writing - original draft, Writing - review \& editing. Steven Kelk: Conceptualization, Methodology, Writing - original draft, Writing - review \& editing. Leen Stougie: Conceptualization, Methodology, Writing - original draft, Writing - review \& editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Programme Networks 024.002.003 and Klein Grant OCENW.KLEIN.125.

|  |  | $d_{2}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $d_{1}$ | 2 | - | 34 | 43 | 52 | 61 | 70 | 79 | 88 |
|  | 3 | 34 | 43 | 52 | 61 | 70 | 79 | 88 | 97 |
|  | 4 | 43 | 52 | 61 | 70 | 79 | 88 | 97 | 106 |
|  | 5 | 52 | 61 | 70 | 79 | 88 | 97 | 106 | 115 |
|  | 6 | 61 | 70 | 79 | 88 | 97 | 106 | 115 | 124 |
|  | 7 | 70 | 79 | 88 | 97 | 106 | 115 | 124 | 133 |
|  | 8 | 79 | 88 | 97 | 106 | 115 | 124 | 133 | 142 |
|  | 9 | 88 | 97 | 106 | 115 | 124 | 133 | 142 | 151 |

Fig. A.5. Values for $c=9\left(d_{1}+d_{2}\right)-11$.

|  |  | $d_{2}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $d_{1}$ | 2 | - | 14 | 9 | 8 | 7 | 6 | 6 | 6 |
|  | 3 | 15 | 7 | 6 | 5 | 5 | 5 | 4 | 4 |
|  | 4 | 10 | 6 | 5 | 4 | 4 | 4 | 4 | 4 |
|  | 5 | 9 | 5 | 5 | 4 | 4 | 4 | 4 | 4 |
|  | 6 | 8 | 5 | 4 | 4 | 4 | 4 | 3 | 3 |
|  | 7 | 7 | 5 | 4 | 4 | 4 | 3 | 3 | 3 |
|  | 8 | 7 | 5 | 4 | 4 | 4 | 3 | 3 | 3 |
|  | 9 | 7 | 5 | 4 | 4 | 3 | 3 | 3 | 3 |

Fig. A.6. Values for $t^{\prime}=\left\lceil\frac{2 d_{1} d_{2}+d_{1}}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil$.

|  |  | $d_{2}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $d_{1}$ | 2 | - | 952 | 774 | 832 | 854 | 840 | 948 | 1056 |
|  | 3 | 1020 | 602 | 624 | 610 | 700 | 790 | 704 | 776 |
|  | 4 | 860 | 624 | 610 | 560 | 632 | 704 | 776 | 848 |
|  | 5 | 936 | 610 | 700 | 632 | 704 | 776 | 848 | 920 |
|  | 6 | 976 | 700 | 632 | 704 | 776 | 848 | 690 | 744 |
|  | 7 | 980 | 790 | 704 | 776 | 848 | 690 | 744 | 798 |
|  | 8 | 1106 | 880 | 776 | 848 | 920 | 744 | 798 | 852 |
|  | 9 | 1232 | 970 | 848 | 920 | 744 | 798 | 852 | 906 |

Fig. A.7. Values for $\alpha=2 \cdot\left(9\left(d_{1}+d_{2}\right)-11\right) \cdot\left\lceil\frac{2 d_{1} d_{2}+d_{1}}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil$. Observe that the minimum is achieved at $d_{1}=4, d_{2}=5$.

## Appendix A. Finding optimal $\boldsymbol{d}_{\mathbf{1}}, \boldsymbol{d}_{\mathbf{2}}$

For the sake of completeness, we here argue that the choice of $d_{1}=4, d_{2}=5$ gives the minimum value of $\alpha=2 \cdot(9)\left(d_{1}+\right.$ $\left.\left.d_{2}\right)-11\right) \cdot\left\lceil\frac{2 d_{1} d_{2}+d_{1}}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil$ in Theorem 1. Let $c=9\left(d_{1}+d_{2}\right)-11$ and $t^{\prime}=\left\lceil\frac{2 d_{1} d_{2}+d_{1}}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil$, so that $\alpha=2 c t^{\prime}$. For $d_{1}=4, d_{2}=5$, we have $c=81-11=70$ and $t^{\prime}=\left\lceil\frac{44}{11}\right\rceil=4$, and so $\alpha=2 \cdot 70 \cdot 4=560$. Figs. A.5, A. 6 and A. 7 gives the possible values of $c, t^{\prime}$ and $\alpha$ respectively, for $d_{1}, d_{2}$ taking values between 2 and 9 (recall that $d_{1}, d_{2}$ must be at least 2 , as Lemma 6 requires $d_{1} d_{2}-d_{1}-d_{2}>0$ ).

By inspection of Fig. A.7, it is easy to see that the minimum possible value of $\alpha$ for $2 \leq d_{1}, d_{2} \leq 9$ is 560 . For larger values of $d_{1}, d_{2}$, we argue as follows: Observe that $t^{\prime}=\left\lceil\frac{2 d_{1} d_{2}+d_{1}}{d_{1} d_{2}-d_{1}-d_{2}}\right\rceil$ is at least 3 for any $d_{1}, d_{2}$, as

$$
\frac{2 d_{1} d_{2}+d_{1}}{d_{1} d_{2}-d_{1}-d_{2}}>\frac{2 d_{1} d_{2}}{d_{1} d_{2}}=2
$$

If one of $d_{1}, d_{2}$ is at least 10 , then $c=9\left(d_{1}+d_{2}\right)-11 \geq 9(10+2)-11=97$. But then for such values we would have $\alpha=2 c t^{\prime} \geq 2 \cdot 97 \cdot 3=582$. Thus, the smallest value of $\alpha$ is in fact 560 , achieved for $d_{1}=4, d_{2}=5$.

## References

[1] B.L. Allen, M. Steel, Subtree transfer operations and their induced metrics on evolutionary trees, Ann. Comb. 5 (2001) 1-15.
[2] N. Alon, B. Chor, F. Pardi, A. Rapoport, Approximate maximum parsimony and ancestral maximum likelihood, IEEE/ACM Trans. Comput. Biol. Bioinform. 7 (1) (January 2010) 183-187.
[3] answer Anonymous, How many vertices of degree 3 or more can a tree have at most?, Mathematics Stack Exchange, https://math.stackexchange.com/ q/388948 (version: 2013-05-12).
[4] R. Atkins, C. McDiarmid, Extremal distances for subtree transfer operations in binary trees, Ann. Comb. 23 (1) (2019) 1-26.
[5] J. Baste, C. Paul, I. Sau, C. Scornavacca, Efficient FPT algorithms for (strict) compatibility of unrooted phylogenetic trees, Bull. Math. Biol. 79 (4) (2017) 920-938.
[6] H.L. Bodlaender, A tourist guide through treewidth, Acta Cybern. 11 (1-2) (1994) 1.
[7] H.L. Bodlaender, A linear-time algorithm for finding tree-decompositions of small treewidth, SIAM J. Comput. 25 (1996) $1305-1317$.
[8] O. Boes, M. Fischer, S. Kelk, A linear bound on the number of states in optimal convex characters for maximum parsimony distance, IEEE/ACM Trans. Comput. Biol. Bioinform. 14 (2) (2016) 472-477.
[9] M.L. Bonet, K. St John, On the complexity of uSPR distance, IEEE/ACM Trans. Comput. Biol. Bioinform. 7 (3) (2010) 572-576.
[10] M. Bordewich, C. Scornavacca, N. Tokac, M. Weller, On the fixed parameter tractability of agreement-based phylogenetic distances, J. Math. Biol. 74 (1-2) (2017) 239-257.
[11] D. Bryant, J. Lagergren, Compatibility of unrooted phylogenetic trees is FPT, Theor. Comput. Sci. 351 (3) (2006) 296-302.
[12] L. Bulteau, M. Weller, Parameterized algorithms in bioinformatics: an overview, Algorithms 12 (12) (2019) 256.
[13] J. Chen, J-H. Fan, S-H. Sze, Parameterized and approximation algorithms for maximum agreement forest in multifurcating trees, Theor. Comput. Sci. 562 (2015) 496-512.
[14] J. Chen, F. Shi, J. Wang, Approximating maximum agreement forest on multiple binary trees, Algorithmica 76 (4) (2016) $867-889$.
[15] M. Cygan, F. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, S. Saurabh, Parameterized Algorithms, 1st edition, Springer Publishing Company, Incorporated, 2015.
[16] R. Downey, M. Fellows, Fundamentals of Parameterized Complexity, vol. 4, Springer, 2013.
[17] J. Felsenstein, Inferring Phylogenies, Sinauer Associates, Incorporated, 2004.
[18] M. Fischer, S. Kelk, On the maximum parsimony distance between phylogenetic trees, Ann. Comb. 20 (1) (2016) 87-113.
[19] W. Fitch, Toward defining the course of evolution: minimum change for a specific tree topology, Syst. Biol. 20 (4) (1971) 406-416.
[20] D. Huson, R. Rupp, C. Scornavacca, Phylogenetic Networks: Concepts, Algorithms and Applications, Cambridge University Press, 2011.
[21] R. Janssen, M. Jones, S. Kelk, G. Stamoulis, T. Wu, Treewidth of display graphs: bounds, brambles and applications, J. Graph Algorithms Appl. 23 (4) (2019).
[22] S. Kelk, M. Fischer, On the complexity of computing MP distance between binary phylogenetic trees, Ann. Comb. 21 (2017) 573-604.
[23] S. Kelk, M. Fischer, V. Moulton, T. Wu, Reduction rules for the maximum parsimony distance on phylogenetic trees, Theor. Comput. Sci. 646 (2016) 1-15.
[24] S. Kelk, S. Linz, A tight kernel for computing the tree bisection and reconnection distance between two phylogenetic trees, SIAM J. Discrete Math. 33 (3) (2019) 1556-1574.
[25] S. Kelk, S. Linz, New reduction rules for the tree bisection and reconnection distance, Ann. Comb. 24 (2020) 475-502.
[26] S. Kelk, G. Stamoulis, A note on convex characters, Fibonacci numbers and exponential-time algorithms, Adv. Appl. Math. 84 (2017) 34-46.
[27] S. Kelk, G. Stamoulis, T. Wu, Treewidth distance on phylogenetic trees, Theor. Comput. Sci. 731 (2018) 99-117.
[28] S. Kelk, L.J.J. van Iersel, C. Scornavacca, M. Weller, Phylogenetic incongruence through the lens of monadic second order logic, J. Graph Algorithms Appl. 20 (2) (2016) 189-215.
[29] F. Liers, A. Martin, S. Pape, Binary Steiner trees: structural results and an exact solution approach, Discrete Optim. 21 (2016) 85-117.
[30] S. Moran, S. Snir, Convex recolorings of strings and trees: definitions, hardness results and algorithms, J. Comput. Syst. Sci. 74 (5) (2008) 850-869.
[31] V. Moulton, T. Wu, A parsimony-based metric for phylogenetic trees, Adv. Appl. Math. 66 (2015) 22-45.
[32] L. Nakhleh, Computational approaches to species phylogeny inference and gene tree reconciliation, Trends Ecol. Evol. 28 (12) (2013) $719-728$.
[33] D. Robinson, L. Foulds, Comparison of phylogenetic trees, Math. Biosci. 53 (1-2) (1981) 131-147.
[34] C. Semple, M. Steel, Phylogenetics, Oxford University Press, 2003.
[35] F. Shi, J. Chen, Q. Feng, J. Wang, A parameterized algorithm for the maximum agreement forest problem on multiple rooted multifurcating trees, J. Comput. Syst. Sci. 97 (2018) 28-44.
[36] K. St John, The shape of phylogenetic treespace, Syst. Biol. 66 (1) (2017) e83-e94.
[37] M. Steel, Phylogeny: Discrete and Random Processes in Evolution, SIAM, 2016.
[38] L. van Iersel, M. Jones, S. Kelk, A third strike against perfect phylogeny, Syst. Biol. 68 (5) (2019) 814-827.
[39] L. van Iersel, S. Kelk, N. Lekic, C. Whidden, N. Zeh, Hybridization number on three rooted binary trees is EPT, SIAM J. Discrete Math. 30 (3) (2016) 1607-1631.
[40] T. Warnow, Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation, Cambridge University Press, 2017.
[41] C. Whidden, R.G. Beiko, N. Zeh, Fixed-parameter algorithms for maximum agreement forests, SIAM J. Comput. 42 (4) (2013) 1431-1466.
[42] C. Whidden, N. Zeh, R.G. Beiko, Supertrees based on the subtree prune-and-regraft distance, Syst. Biol. 63 (4) (2014) 566-581.


[^0]:    * Corresponding author at: Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Delft, the Netherlands.

    E-mail address: M.E.L.Jones@tudelft.nl (M. Jones).
    https://doi.org/10.1016/j.jcss.2020.10.003
    0022-0000/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license
    (http://creativecommons.org/licenses/by/4.0/).

[^1]:    ${ }^{1}$ The proof of this observation is based on an argument in [3].

