

Non-hereditary Maximum Parsimony trees

Mareike Fischer

Received: 3 August 2010 / Revised: 25 July 2011
© Springer-Verlag 2011

Abstract In this paper, we investigate a conjecture by Arndt von Haeseler concerning the Maximum Parsimony method for phylogenetic estimation, which was published by the Newton Institute in Cambridge on a list of open phylogenetic problems in 2007. This conjecture deals with the question whether Maximum Parsimony trees are hereditary. The conjecture suggests that a Maximum Parsimony tree for a particular (DNA) alignment necessarily has subtrees of all possible sizes which are most parsimonious for the corresponding subalignments. We answer the conjecture affirmatively for binary alignments on 5 taxa but also show how to construct examples for which Maximum Parsimony trees are not hereditary. Apart from showing that a most parsimonious tree cannot generally be reduced to a most parsimonious tree on fewer taxa, we also show that compatible most parsimonious quartets do not have to provide a most parsimonious supertree. Last, we show that our results can be generalized to Maximum Likelihood for certain nucleotide substitution models.

Keywords Phylogenetics · Maximum Parsimony · Maximum Likelihood · Jukes-Cantor model

Mathematics Subject Classification (2000) 90C35 · 92E99 · 94C15

1 Introduction

Tree reconstruction methods for inferring phylogenetic trees are used to interpret the ever-growing amount of available genetic sequence data. Unsurprisingly, such methods

M. Fischer (✉)
Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories,
University of Vienna, Medical University of Vienna, University of Veterinary Medicine Vienna,
Dr. Bohr Gasse 9, 1030 Vienna, Austria
e-mail: email@mareikefischer.de

have therefore been widely discussed in the last decades (e.g., [Felsenstein 1978, 2004](#); [Semple and Steel 2003](#); [Yang 2006](#)). One of the most frequently used tree reconstruction methods is the so-called Fitch parsimony ([Fitch 1971](#)) or *Maximum Parsimony* method (MP). Two of the reasons for its popularity are its simplicity compared to other methods such as Maximum Likelihood as well as its purely combinatorial basic principle. The latter makes MP a method that can be applied to any data alignment without any assumptions on the way the data has been generated, which means for the DNA that no assumptions on the probability of a nucleotide substitution have to be made (which is why MP is often said to be ‘model-free’). Despite this simplicity, not all aspects of MP are to-date understood. One of the questions that remained unsolved for quite some time is whether MP trees are hereditary, i.e. if for an MP tree of an alignment on m taxa we can find a subtree of this tree of size k (for all $k = 4, \dots, m - 1$) which is most parsimonious for the corresponding subalignment. This problem was submitted by Professor Arndt von Haeseler to the Isaac Newton Institute’s list of open phylogenetic problems in 2007 ([von Haeseler 2007](#)) as well as to the ‘Penny Ante’ list of the Annual New Zealand Phylogenetics Meeting in Kaikoura in 2009 ([von Haeseler 2009](#)).

The importance of the question is manifold. Biologically, MP trees with no MP subtrees seem quite counterintuitive as one would expect the MP tree to be related to MP trees on fewer taxa. Particularly when outgroups are included in a DNA analysis, one would want the topology of the rest of the tree to be independent of the outgroup, so the topology of the ‘best’ tree should be independent of the presence or absence of the outgroup. Moreover, if the conjecture was true, there would be sequences of MP trees, starting from four taxa and growing one new leaf at a time, leading to each MP tree of the whole alignment under consideration, so the big trees would ‘inherit’ their MP property from smaller trees. Mathematically, such a property would be particularly interesting with regards to inductive proofs or dynamic programming. Last but not least, the question of whether or not MP trees are generally hereditary arises naturally, as one often observes in real-world data that these trees in fact have that property (results not shown). Also, it is known that distance-based tree reconstruction methods are in fact hereditary in some sense: there, perfect additivity (‘tree-likeness’) leads to the reconstructibility of the tree, but even a relaxation of the so-called four-point condition still returns the correct tree ([Bandelt and Dress 1986](#)), which does not change if taxa are removed. The same holds for Maximum Parsimony in the ideal case—i.e. in a homoplasy-free setting, which we formally show in Lemma 1. So the question whether MP trees are hereditary is in fact motivated by the question whether the assumption of homoplasy-freeness can be relaxed or whether such a relaxation destroys heridity. However, we show in this paper that the conjecture is not in general true, but does hold in some special cases like in the case when the alignment is homoplasy-free or when the alignment is binary and there are only 5 taxa.

While the above-mentioned aspects of heredity basically refer to reducing large MP trees to smaller ones, we also consider the opposite scenario: we show that even if an alignment has only unique MP quartet trees for all 4-taxa subalignments and even if all these quartets are compatible with one another, the supertree comprising all these quartets is not necessarily an MP tree for the original alignment. This means that MP quartets cannot generally be combined into larger MP trees.

Last, we investigate the impact these findings concerning MP have on Maximum Likelihood (ML) under the (generalized) Jukes-Cantor model (also known as N_r -model). In this analysis, we use the strong relationship of MP and ML as described in (Tuffley and Steel 1997) and conclude that the cases that are problematic for MP also turn out to be problematic for ML under the N_r -model, even if there is a common mechanism of site evolution.

2 Notation and model assumptions

Recall that an *unrooted binary phylogenetic tree on a leaf set X* is a tree $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ with leaf set (also called taxon set) $X = \{1, \dots, m\} \subset V(\mathcal{T})$ with only vertices of degree 1 (leaves) or 3 (internal vertices). In this paper, when there is no ambiguity we often just write ‘phylogenetic tree’ or ‘tree’ when referring to an unrooted binary phylogenetic tree. Also, when referring to a tree on a leaf set X with $|X| = m$, we write *m -taxa tree* for short. Sometimes it makes sense to denote trees in the so-called *Newick Format* (Felsenstein et al 2000), which basically is a list of the taxa under consideration such that each clade of the tree is represented by brackets. For instance, the tree $((1, 2), (3, 4))$ consists of one clade employing taxa 1 and 2 and another one employing taxa 3 and 4. Note that for unrooted trees, the notations $((1, 2), (3, 4))$, $(1, 2, (3, 4))$ and $((1, 2), 3, 4)$ are equivalent.

Furthermore, recall that a *character f* is a function $f : X \rightarrow \mathcal{C}$ for some set $\mathcal{C} := \{c_1, c_2, c_3, \dots, c_r\}$ of r *character states* ($r \in \mathbb{N}$). An *extension* of f to $V(\mathcal{T})$ is a map $g : V(\mathcal{T}) \rightarrow \mathcal{C}$ such that $g(i) = f(i)$ for all i in X . For such an extension g of f , we denote by $l_{\mathcal{T}}(g)$ the number of edges $e = \{u, v\}$ in \mathcal{T} on which a substitution occurs, i.e. where $g(u) \neq g(v)$. The *parsimony score* of f on \mathcal{T} , denoted by $l_{\mathcal{T}}(f)$, is obtained by minimizing $l_{\mathcal{T}}(g)$ over all possible extensions g . Given a tree \mathcal{T} and a character f on the same taxon set, one can easily calculate the parsimony score of f on \mathcal{T} with the famous Fitch algorithm (Fitch 1971). The parsimony score of a sequence of characters $S := f_1 f_2 \dots f_n$ on a tree \mathcal{T} is then given by $l_{\mathcal{T}}(S) = \sum_{i=1}^n l_{\mathcal{T}}(f_i)$. Then, a most parsimonious or Maximum Parsimony tree (MP tree for short) for a sequence of characters S is a tree which minimizes $l_{\mathcal{T}}(S)$. So a Maximum Parsimony tree is a tree which minimizes the number of evolutionary events like substitutions necessary to explain a given alignment. This is also often referred to as the principle of Ockham’s razor, which says that a simpler explanation is preferable to a more complex one—so in the parsimony case, Maximum Parsimony tries to explain the alignment data with as few substitution events as possible. Recall that theoretically the Fitch algorithm can be applied to all trees in the tree space under consideration, such that at least one most parsimonious tree will be found. However, as MP is known to be NP-hard, this might not be possible in polynomial time.

Note that S cannot only be viewed columnwise as a sequence of characters, but also rowwise as aligned (DNA) species data. In this paper, we therefore use the terms ‘sequence of characters’ and ‘alignment’ synonymously when there is no ambiguity. Moreover, for a taxon $k \in X$ we denote by $f - k$ and $S - k$ the restriction of f and S on the set $X - k$, respectively. Similarly, we denote by $\mathcal{T} - \{k_1, \dots, k_i\}$

the binary phylogenetic tree that results from tree \mathcal{T} by deleting taxa k_1, \dots, k_i (for some integer $i \geq 1$) and all nodes of degree 2 that might be caused by this deletion.

A character f is said to be *homoplasy-free* on a tree \mathcal{T} if $l_{\mathcal{T}}(f) = |f| - 1$, where $|f|$ denotes the number of character states employed by f . A sequence S of characters is called homoplasy-free when all its characters have that property. This means that a sequence S is homoplasy-free if there exists a binary phylogenetic tree \mathcal{T} such that all characters in S are homoplasy free on \mathcal{T} . Note that if a character or an alignment is homoplasy-free on a certain tree, this tree minimizes its parsimony score and is therefore most parsimonious for this character or alignment, respectively.

Recall that a character f on a leaf set X is said to be *informative* (with respect to parsimony) if at least two distinct character states occur more than once on X . Otherwise f is called *non-informative*. Note that for a non-informative character f , $l_{\mathcal{T}_i}(f) = l_{\mathcal{T}_j}(f)$ for all trees $\mathcal{T}_i, \mathcal{T}_j$ on the same set X of leaves. In this paper, we refer to a character always with its underlying taxon clustering pattern in mind, i.e. for instance we do not distinguish between $AACC$, $CCAA$ and $CCGG$, and so on. Note that when we explicitly name a character, e.g. $f = AACC$, we write it like a word, whereas in a character sequence $S = f_1 \dots f_n$, the characters are written as columns.

Next we describe the fully symmetric r -state model (Neyman 1971), also known as the N_r -model, where r denotes the number of character states involved, which underlies the Tuffley and Steel equivalence result concerning MP and ML (Tuffley and Steel 1997).

Consider a binary phylogenetic tree \mathcal{T} on a leaf set X arbitrarily rooted at one of its vertices (note that because of the symmetry of the N_r -model described in the following, this vertex can even be a leaf). The N_r -model assumes that a state is assigned to the root from the uniform distribution on the set of states. The state then evolves away from the root as follows. The model assumes equal rates of substitutions between any two distinct character states. For any edge $e = \{u, v\} \in E(\mathcal{T})$, where u is the vertex closer to the root, let p_e denote the conditional probability $P(v = c_i | u = c_j)$, where $c_i \neq c_j$ are two distinct character states. The probability p_e is equal for all pairs of distinct states c_i and c_j . Therefore, the probability that a substitution (c_j to a state different from c_j) occurs on the edge e is $(r - 1)p_e$. Let q_e be the conditional probability $P(v = c_i | u = c_i)$, i.e. the probability that no substitution occurs on edge e . In the N_r -model, we have $0 \leq p_e \leq \frac{1}{r}$ for all $e \in E(\mathcal{T})$, and $(r - 1)p_e + q_e = 1$. Moreover, the N_r -model assumes that substitutions on different edges are independent. Note that for $r = 4$, the N_r -model coincides with the well-known Jukes-Cantor model (Jukes and Cantor 1969).

Let \mathcal{T} be a phylogenetic X -tree and let f be a character on its leaf set X . Let the substitution probabilities assigned to the edges of \mathcal{T} under the N_r -model be collectively denoted by $\bar{p} := (p_e : e \in E(\mathcal{T}))$. Then we denote by $P(f|\mathcal{T}, \bar{p})$ the probability of observing character f given tree \mathcal{T} and the parameter values \bar{p} . Note that $P(f|\mathcal{T}, \bar{p})$ does not depend on the root position since the model is symmetric. The maximum value of this probability for fixed f and \mathcal{T} as \bar{p} ranges over all possibilities is denoted by $\max P(f|\mathcal{T})$, i.e. $\max P(f|\mathcal{T}) := \max_{\bar{p}} P(f|\mathcal{T}, \bar{p})$.

Now let $S := f_1 \dots f_n$ be a sequence of characters. When we refer to a sequence of characters under the N_r -model with *no common mechanism*, this means that the substitution probabilities on edges may be different for different characters in S without any correlation between the characters. We suppose that for each character f_i in the sequence and for each edge e of the tree, there is a parameter $p_{e,i}$ that gives the substitution probability for f_i on edge e . When there is no common mechanism, the parameters $p_{e,i}$ can all be chosen independently in order to maximize the likelihood. For $i = 1, \dots, n$, let $\bar{p}_i := (p_{e,i} : e \in E(\mathcal{T}))$ be the vectors of substitution probabilities. We denote the model parameters $(\bar{p}_i, i = 1, \dots, n)$ collectively as Θ and refer to $P(S|\mathcal{T}, \Theta)$ as the probability of observing sequence S given the phylogenetic tree \mathcal{T} and model parameters Θ . We then define the likelihood of the tree \mathcal{T} and the model parameters Θ given the sequence S , which we refer to as the *likelihood function*, as $L(\mathcal{T}, \Theta|S) := P(S|\mathcal{T}, \Theta)$. The Maximum Likelihood method of phylogenetic tree reconstruction involves optimizing the likelihood function in two steps as described in (Semple and Steel 2003). We first maximize $P(S|\mathcal{T}, \Theta)$ over the space of model parameters Θ . We define:

$$\max P(S|\mathcal{T}) := \max_{\Theta} P(S|\mathcal{T}, \Theta).$$

We then choose a tree \mathcal{T} that maximizes $\max P(S|\mathcal{T})$. We call such a tree a Maximum Likelihood tree (ML-tree) of S . Thus, an ML-tree of a sequence S is $\operatorname{argmax}_{\mathcal{T}} (\max P(S|\mathcal{T}))$. Note that under the assumption of no common mechanism, i.e. the characters in an alignment are regarded independent of one another, we have:

$$\max P(S|\mathcal{T}) = \prod_{i=1}^n \max_{\bar{p}_i} P(f_i|\mathcal{T}, \bar{p}_i).$$

3 Results

3.1 Heredity part I: Inferring small MP trees from larger ones

As explained in Sect. 1, we analyze cases in which MP trees are or are not hereditary. In particular, we examine whether a most parsimonious tree on an alignment needs to be related to most parsimonious trees on subalignments. In fact, the conjecture under investigation suggests a sequence of MP trees of sizes leading from the number of taxa considered down to four, where small MP trees are subtrees of the larger ones. Note that as there is only one unrooted tree on one, two and three taxa, respectively, the conjecture does not consider these cases as subtrees of these sizes are unique and therefore always most parsimonious.

We now formulate the conjecture mathematically.

Conjecture 1 (Conjecture PC5 from the Isaac Newton Institute’s ‘Phylogenetics: Challenges and Conjectures’ list 2007) Let $S := f_1 f_2 \dots f_n$ be a sequence of characters (‘alignment’) on the set X of taxa, where $|X| = m$, and let \mathcal{T} be a Maximum Parsimony tree for S . Then, for each $k = 4, \dots, m - 1$, there exists a subset Y of X

of size k so that $T|_Y$ is an MP tree for $S|_Y$ (where $S|_Y$ is the sequence S of characters restricted to the taxa in Y and $T|_Y$ is the tree T restricted to the taxa in Y).

While finding an MP tree is generally NP-hard (Foulds and Graham 1982), if this conjecture was true it might be relevant for dynamic programming approaches concerning certain instances of parsimony. Note that the conjecture does not state which particular subtree would be most parsimonious—so the conjecture is not in conflict with the NP-hardness of Maximum Parsimony and therefore could be valid. Moreover, mathematically a statement like that given in the conjecture would be useful to investigate theoretical properties of MP using inductive proofs, as the inductive step in such proofs requires knowledge on smaller instances of the problem under investigation.

In the following, we will present two special cases of the conjecture, namely the case in which the given alignment is homoplasy-free as well as the case where the alignment is on 5 taxa and employs only binary characters. In these cases, the conjecture is true. Moreover, we afterwards analyze more general cases where the conjecture fails.

We need the following lemma in order to prove a first positive result concerning Conjecture 1.

Lemma 1 *Let T be an unrooted binary phylogenetic X -tree for a taxon set X with $|X| = m$ and let $f = c_1, \dots, c_m$ be a homoplasy-free character on T . Let $k \in X$ be a taxon. Then, $f - k$ is homoplasy-free on $T - k$.*

Proof By definition of homoplasy-free, $l_T(f) = |f| - 1$. For any taxon $k \in X$, note that the parsimony score of the character $f - k$ on $T - k$ can be calculated as follows:

$$\begin{aligned} l_{T-k}(f - k) &= \begin{cases} l_T(f) = |f| - 1 & \text{if } c_k \text{ is not unique in } f \\ l_T(f) - 1 = |f| - 2 & \text{else,} \end{cases} \\ &= \begin{cases} |f - k| - 1 & \text{if } |f - k| = |f|, \\ |f - k| - 1 & \text{if } |f - k| = |f| - 1. \end{cases} \end{aligned}$$

So altogether $l_{T-k}(f - k) = |f - k| - 1$. Thus, $f - k$ is homoplasy-free on $T - k$. \square

Now we are in the position to state the first heredity result.

Theorem 1 *Conjecture 1 is true if S is homoplasy-free.*

Proof Let S be a homoplasy-free alignment with MP tree T and taxon set $X = \{1, \dots, m\}$. Then, by Lemma 1, for any taxon $k \in X$ we conclude that the restriction $S - k$ of S on $X - k$ is homoplasy-free on the corresponding restriction $T - k$ of T . As explained in Sect. 2, homoplasy-free alignments are parsimoniously best possible, i.e. because $S - k$ is homoplasy-free on $T - k$, $T - k$ is an MP tree for $S - k$. We repeat this argument to derive the desired sequence of MP trees from $m - 1$ taxa down to 4 taxa. This completes the proof. \square

An example for heredity of homoplasy-free alignments is depicted in Fig. 1.

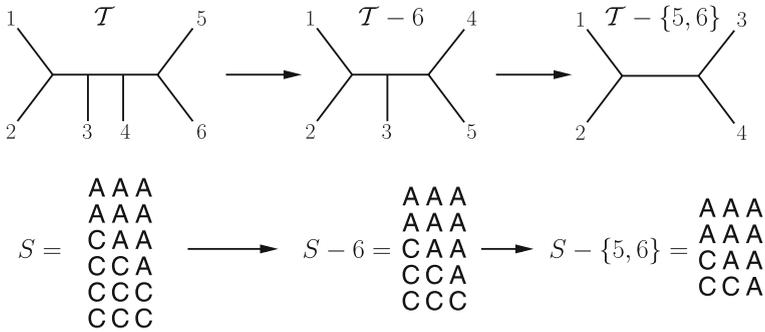
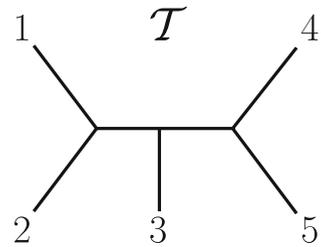


Fig. 1 Illustration of Theorem 1. Alignment S is homoplasy-free on tree \mathcal{T} , so all subalignments are homoplasy-free on the corresponding subtrees

Fig. 2 Tree $\mathcal{T} = ((1, 2), 3, (4, 5))$, which is an MP tree for some given alignment S



In order to investigate Conjecture 1 for general alignments, we now describe the idea underlying the following results.

Main idea. If for m taxa there exist p distinct characters (or, more precisely, character patterns, cf. Sect. 2) f_1, \dots, f_p , the parsimony score of an alignment S on an m -taxa tree \mathcal{T} can be expressed as $\sum_{i=1}^p x_i l_{\mathcal{T}}(f_i)$, where x_i denotes the number of times the character f_i occurs in S (note that this implies $|S| = \sum_{i=1}^p x_i$). So the fact that a tree \mathcal{T} is parsimoniously better than another tree $\hat{\mathcal{T}}$ concerning some alignment S can be expressed in terms of the inequality $\sum_{i=1}^p x_i l_{\mathcal{T}}(f_i) < \sum_{i=1}^p x_i l_{\hat{\mathcal{T}}}(f_i)$. The same can be done for subalignments and the corresponding subtrees, so that altogether Conjecture 1 leads to a system of inequalities that need to be fulfilled by a potential counterexample. Such systems can then be tackled with the help of computer algebra systems.

We now use the idea explained above to prove the following statement on 5 taxa.

Theorem 2 *Conjecture 1 is true in the case where $|X| = m = 5$ and $S = f_1 \dots f_n$ is binary, i.e. f_1, \dots, f_n are 2-state characters. In particular, if a tree $\mathcal{T} = ((1, 2), 3, (4, 5))$ as depicted in Fig. 2 is an MP tree for such an alignment S , then the tree $\mathcal{T} - 3 = ((1, 2)(4, 5))$ as depicted in Fig. 3 is an MP tree for the alignment $S - 3$, which results from S when taxon 3 is deleted.*

Proof Let f_1, \dots, f_{10} be the ten binary parsimoniously informative characters on 5 taxa as depicted in Table 1. Let S be a binary alignment on 5 taxa. Without loss of generality, we assume that tree $\mathcal{T} = ((1, 2), 3, (4, 5))$ as depicted in Fig. 2 is

Fig. 3 If $\mathcal{T} = ((1, 2), 3, (4, 5))$ is an MP tree for some given binary alignment S , tree $\mathcal{T} - 3 = ((1, 2)(4, 5))$, is an MP tree for $S - 3$

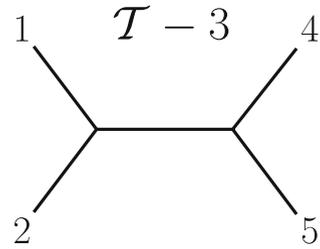


Table 1 Overview of all parsimoniously informative characters on 5 taxa. The proof of Theorem 2 employs only the binary characters f_1, \dots, f_{10} , whereas the proof of Proposition 1 uses all 25 characters

Parsimoniously informative characters on five taxa					
Binary	$f_1 := \text{AACCC}$	$f_2 := \text{ACACC}$	$f_3 := \text{ACCAC}$	$f_4 := \text{ACCCA}$	$f_5 := \text{ACCAA}$
	$f_6 := \text{ACACA}$	$f_7 := \text{ACAAC}$	$f_8 := \text{AACCA}$	$f_9 := \text{AACAC}$	$f_{10} := \text{AAACC}$
Ternary	$f_{11} := \text{AACCT}$	$f_{12} := \text{AACTC}$	$f_{13} := \text{AATCC}$	$f_{14} := \text{AACTT}$	$f_{15} := \text{ACATC}$
	$f_{16} := \text{ATACC}$	$f_{17} := \text{ACCAT}$	$f_{18} := \text{ACTAC}$	$f_{19} := \text{ATCAC}$	$f_{20} := \text{ACCTA}$
	$f_{21} := \text{ACTCA}$	$f_{22} := \text{ATCCA}$	$f_{23} := \text{TAACC}$	$f_{24} := \text{TACAC}$	$f_{25} := \text{TACCA}$

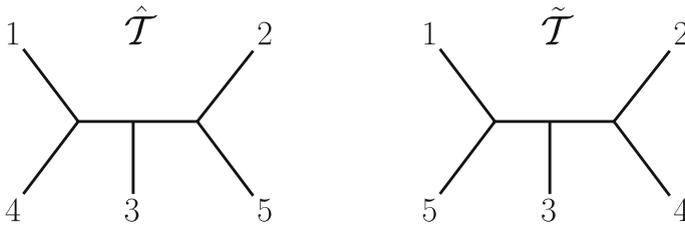


Fig. 4 If tree \mathcal{T} depicted in Fig. 2 is an MP tree for an alignment S , the trees $\hat{\mathcal{T}} = ((1, 4), 3, (2, 5))$ and $\tilde{\mathcal{T}} = ((1, 5), 3, (2, 4))$ cannot have better parsimony scores for S than \mathcal{T} , which leads to Inequality (1)

most parsimonious for S (otherwise we re-label the leaves). This particularly implies that

$$l_{\mathcal{T}}(S) \leq l_{\hat{\mathcal{T}}}(S) \text{ as well as } l_{\mathcal{T}}(S) \leq l_{\tilde{\mathcal{T}}}(S), \tag{1}$$

where $\hat{\mathcal{T}} = ((1, 4), 3, (2, 5))$ and $\tilde{\mathcal{T}} = ((1, 5), 3, (2, 4))$ are the trees depicted in Fig. 4. Note that we may ignore non-informative characters as they have the same score on all trees. Therefore, we can think of S as a combination of characters f_1, \dots, f_{10} , which occur x_1, \dots, x_{10} times in S , respectively. We then rewrite Inequality 1 as follows:

$$\sum_{i=1}^{10} x_i l_{\mathcal{T}}(f_i) \leq \sum_{i=1}^{10} x_i l_{\hat{\mathcal{T}}}(f_i) \quad \text{and} \quad \sum_{i=1}^{10} x_i l_{\mathcal{T}}(f_i) \leq \sum_{i=1}^{10} x_i l_{\tilde{\mathcal{T}}}(f_i). \tag{2}$$

Calculating the parsimony scores for f_1, \dots, f_{10} on trees $\mathcal{T}, \hat{\mathcal{T}}$ and $\tilde{\mathcal{T}}$, respectively, we get $l_{\mathcal{T}}(f_1) = l_{\mathcal{T}}(f_{10}) = l_{\hat{\mathcal{T}}}(f_3) = l_{\tilde{\mathcal{T}}}(f_7) = l_{\tilde{\mathcal{T}}}(f_4) = l_{\hat{\mathcal{T}}}(f_6) = 1$, and all other parsimony scores are equal to 2. We now rewrite Inequality 2 using these scores:

$$\begin{aligned}
 &x_1 + 2x_2 + 2x_3 + 2x_4 + 2x_5 + 2x_6 + 2x_7 + 2x_8 + 2x_9 + x_{10} \\
 &\leq x_1 + 2x_2 + x_3 + 2x_4 + 2x_5 + 2x_6 + x_7 + 2x_8 + 2x_9 + 2x_{10} \\
 &\Leftrightarrow x_3 + x_7 \leq x_1 + x_{10}
 \end{aligned} \tag{3}$$

and

$$\begin{aligned}
 &x_1 + 2x_2 + 2x_3 + 2x_4 + 2x_5 + 2x_6 + 2x_7 + 2x_8 + 2x_9 + x_{10} \\
 &\leq 2x_1 + 2x_2 + 2x_3 + x_4 + 2x_5 + x_6 + 2x_7 + 2x_8 + 2x_9 + 2x_{10} \\
 &\Leftrightarrow x_4 + x_6 \leq x_1 + x_{10}
 \end{aligned} \tag{4}$$

Now we assume that the subtree $\mathcal{T} - 3 = ((12), (45))$ of \mathcal{T} is not most parsimonious. This implies that at least one of the two alternative trees, namely $\hat{\mathcal{T}} - 3 = ((14), (25))$ or $\tilde{\mathcal{T}} - 3 = ((15), (24))$, must be strictly better than $\mathcal{T} - 3$ in the sense of parsimony. Using the above argument, we get

$$l_{\mathcal{T}-3}(S - 3) > l_{\hat{\mathcal{T}}-3}(S - 3) \quad \text{or} \quad l_{\mathcal{T}-3}(S - 3) > l_{\tilde{\mathcal{T}}-3}(S - 3). \tag{5}$$

Calculating the parsimony scores of $f_1 - 3, \dots, f_{10} - 3$ on trees $\mathcal{T} - 3, \hat{\mathcal{T}} - 3$ and $\tilde{\mathcal{T}} - 3$, respectively, we get

$$\begin{aligned}
 l_{\mathcal{T}-3}(f_3 - 3) &= l_{\mathcal{T}-3}(f_4 - 3) = l_{\mathcal{T}-3}(f_6 - 3) = l_{\mathcal{T}-3}(f_7 - 3) = l_{\hat{\mathcal{T}}-3}(f_1 - 3) \\
 &= l_{\hat{\mathcal{T}}-3}(f_4 - 3) = l_{\hat{\mathcal{T}}-3}(f_6 - 3) = l_{\hat{\mathcal{T}}-3}(f_{10} - 3) = l_{\tilde{\mathcal{T}}-3}(f_1 - 3) \\
 &= l_{\tilde{\mathcal{T}}-3}(f_3 - 3) = l_{\tilde{\mathcal{T}}-3}(f_7 - 3) = l_{\tilde{\mathcal{T}}-3}(f_{10} - 3) = 2
 \end{aligned}$$

and all other parsimony scores are equal to 1. We now rewrite Inequality 5 using these scores:

$$\begin{aligned}
 &x_1 + x_2 + 2x_3 + 2x_4 + x_5 + 2x_6 + 2x_7 + x_8 + x_9 + x_{10} > 2x_1 + x_2 + x_3 \\
 &\quad + 2x_4 + x_5 + 2x_6 + x_7 + x_8 + x_9 + 2x_{10} \\
 &\Leftrightarrow x_3 + x_7 > x_1 + x_{10}
 \end{aligned} \tag{6}$$

or

$$\begin{aligned}
 &x_1 + x_2 + 2x_3 + 2x_4 + x_5 + 2x_6 + 2x_7 + x_8 + x_9 + x_{10} > 2x_1 + x_2 \\
 &\quad + 2x_3 + x_4 + x_5 + x_6 + 2x_7 + x_8 + x_9 + 2x_{10} \\
 &\Leftrightarrow x_4 + x_6 > x_1 + x_{10}
 \end{aligned} \tag{7}$$

As either Inequality 6 or 7 must hold, this contradicts either (3) or (4). Therefore, $\mathcal{T} - 3$ is an MP tree for $S - 3$. □

Next we show that the result presented in Theorem 2 cannot be generalized to r -state characters for $r > 2$. In fact, not only is it possible that the particular subtree $((12), (45))$ of a most parsimonious tree $((12), 3, (45))$ is not most parsimonious for

the corresponding subalignment. It is even possible that the most parsimonious tree does not have any most parsimonious 4-taxa subtree at all.

Proposition 1 *Conjecture 1 is not generally true for multistate characters, even if the tree under consideration is the only MP tree.*

Proof We construct an explicit example employing three character states. Let f_1, \dots, f_{25} be as depicted in Table 1. Note that there is no parsimoniously informative 5-taxa character employing more than three states, so f_1, \dots, f_{25} is the complete list of parsimoniously informative characters on 5 taxa. Now if a tree $\mathcal{T} = ((12), 3, (45))$ shall be the unique MP tree for an alignment S and none of the 4-taxa subtrees of \mathcal{T} , i.e. $\mathcal{T} - 1 = ((23), (45)), \mathcal{T} - 2 = ((13), (45)), \mathcal{T} - 3 = ((12), (45)), \mathcal{T} - 4 = ((12), (35))$ and $\mathcal{T} - 5 = ((12), (34))$, shall be most parsimonious for the corresponding subalignments of S , this can be expressed with the help of the following system of inequalities (as in the proof of Theorem 2 we can ignore non-informative characters without loss of generality):

$$\text{for all } \hat{\mathcal{T}} \neq \mathcal{T} \text{ we have } \sum_{i=1}^{25} x_i l_{\mathcal{T}}(f_i) \leq \sum_{i=1}^{25} x_i l_{\hat{\mathcal{T}}}(f_i)$$

and for each $j = 1, \dots, 5$ there exists some tree $\tilde{\mathcal{T}}^j \neq \mathcal{T}$ such that $\sum_{i=1}^{25} x_i l_{\mathcal{T}-j}(f_i - j) > \sum_{i=1}^{25} x_i l_{\tilde{\mathcal{T}}^j}(f_i - j)$.

One possible solution to this system of inequalities is $x_4 = 1, x_6 = 2, x_{11} = 1, x_{13} = 2, x_{23} = 1$ and all other $x_i = 0$. This gives the following alignment:

$$S := \begin{cases} \text{AAAAAAA} \\ \text{CCCAAAC} \\ \text{CAACCCC} \\ \text{CCCCTTT} \\ \text{AAATTTT} \end{cases}$$

So S is an alignment with unique MP tree $\mathcal{T} = ((12), 3, (45))$ and no most parsimonious 4-taxa subtree, which can be verified by examining all five distinct characters employed by S on all 15 trees on 5 taxa and their corresponding 4-taxa subtrees. \square

So for 5 taxa, the question whether or not Conjecture 1 holds depends on the number of character states the alignment employs: for two character states it holds, whereas it fails for three or more states. Next we show that this distinction cannot be generalized to more than 5 taxa. In fact, we use our approach of solving inequality systems in order to generate an alignment S on six taxa with unique MP tree $\mathcal{T} = (((12), 3), (4, (5, 6)))$, which has neither a most parsimonious 5-taxa subtree nor a most parsimonious 4-taxa subtree for the corresponding subalignments of S .

Proposition 2 *Conjecture 1 is not generally true for more than 5 taxa, even if only binary characters are employed and the tree under consideration is the only MP tree. In fact, a unique MP tree might not have any (non-trivial) most parsimonious subtree at all.*

Table 2 Overview of all parsimoniously informative binary characters on six taxa

Parsimoniously informative binary characters on six taxa				
$f_1 :=$ AACCCC	$f_2 :=$ ACACCC	$f_3 :=$ ACCACC	$f_4 :=$ ACCCAC	$f_5 :=$ ACCCCA
$f_6 :=$ ACCAAA	$f_7 :=$ ACACAA	$f_8 :=$ ACAACA	$f_9 :=$ ACAAAC	$f_{10} :=$ AACCAA
$f_{11} :=$ AACACA	$f_{12} :=$ AACAAC	$f_{13} :=$ AAACCA	$f_{14} :=$ AAACAC	$f_{15} :=$ AAAACC
$f_{16} :=$ AAACCC	$f_{17} :=$ AACACC	$f_{18} :=$ AACCAC	$f_{19} :=$ AACCCA	$f_{20} :=$ ACAACC
$f_{21} :=$ ACACAC	$f_{22} :=$ ACACCA	$f_{23} :=$ ACCAAC	$f_{24} :=$ ACCACA	$f_{25} :=$ ACCCAA

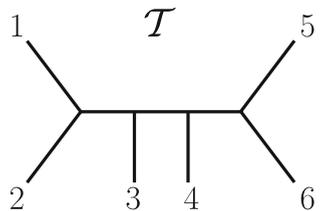
Proof Consider all informative binary characters on six taxa as shown in Table 2. Now we construct an example analogously to the construction shown in the proof of Proposition 1 and find that the alignment S employing two copies of f_3 , five copies of f_4 , four copies of f_5 , one copy of f_7 , nine copies of f_8 , six copies of f_9 , eleven copies of f_{11} , nine copies of f_{12} , three copies of f_{13} , two copies of f_{14} , seven copies of f_{16} , one copy of f_{19} , four copies of f_{20} and six copies of f_{25} has the desired properties. This alignment is depicted in Fig. 5. It has a unique MP tree, namely $\mathcal{T} = (((1, 2), 3), (4, (5, 6)))$ as depicted in Fig. 6. The 5-taxa and 4-taxa MP trees for S are depicted in Figs. 7 and 8, respectively. We recommend the program ‘Penny’ from the free Phylip-package (Felsenstein 2005) for running our example or creating others, as this program is able to run an exhaustive search through the tree space for binary character sequences. \square

Note that while the example presented in the proof of Theorem 2 concerns tree $\mathcal{T} = (((1, 2), 3), (4, (5, 6)))$, analogous examples can be constructed for the other tree shape on six taxa, namely $(((1, 2), (3, 4)), (5, 6))$ (example not shown). We conclude that in general, MP trees for an alignment do not have to be related to MP trees on subalignments. This surprising result shows once again that MP, while being a simple combinatorial algorithm, is more complicated than one might intuitively think. As

AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
 CCCCCCCCC CCCCCCCCC CCCCCCAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
 CCCCCCCCC CAAAAAAAAA AAAAAAAAAACC CCCCCCCCC CCCCCCAA AAAAAAAAAA CAAAAACCCCC
 AACCCCCC CCAAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAACC CCCCCCCCC CAAAAACCCCC
 CAAAAACC CACCCCCC CAAAAACC CCCCCC CAAAAACC CAAAAACC CAAAAACC CAAAAACC CAAAAACC CAAAAACC CAAAAACC
 CCCCCCAA AAAAAAAAAA ACCCCCCAA AAAAAAAAAACC CCCCCCAA ACCCCCCC ACCCCCCC ACCCCCCC

Fig. 5 Alignment S as defined in the proof of Proposition 2 has a unique MP tree \mathcal{T} (cf. Fig. 6), which has no MP subtrees. S consists of 70 characters, where every 10th character is highlighted in bold. Note that the alignment could also be written in terms of the f_i ’s: $f_3 f_3 f_4 f_4 f_4 f_4 \dots$ as described in the above proof

Fig. 6 Tree $\mathcal{T} = (((1, 2), 3), (4, (5, 6)))$ is the unique MP tree for S shown in Fig. 5 but has no most parsimonious subtrees



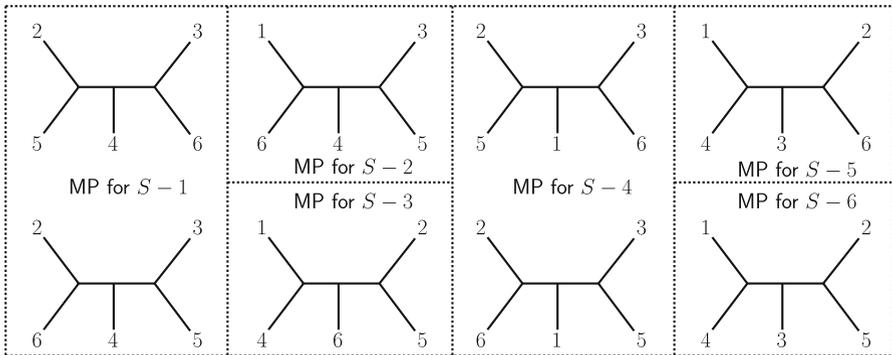


Fig. 7 Illustrations of all 5-taxa MP trees for the corresponding subalignments of alignment S as defined in the proof of Proposition 2. None of these trees is a subtree of \mathcal{T} shown in Fig. 6, which is the unique MP tree of S

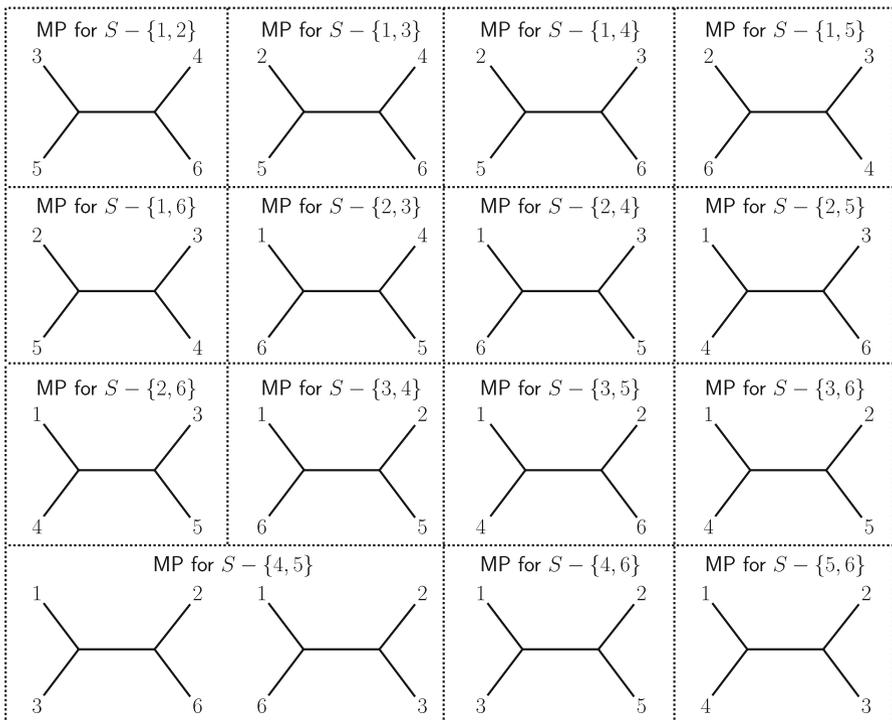


Fig. 8 Illustrations of all 4-taxa MP trees for the corresponding subalignments of alignment S as defined in the proof of Proposition 2. None of these trees is a subtree of \mathcal{T} shown in Fig. 6, which is the unique MP tree of S

explained above, the existence of such instances is *not* immediately clear due to the NP-hardness of finding the set of most parsimonious trees in the tree space. It rather adds another complicated aspect to an already hard problem.

3.2 Heredity part II: Constructing large MP trees from smaller ones

In the previous section, we showed that Maximum Parsimony trees are not in general hereditary in the sense that MP trees for a set of taxa do not necessarily lead to MP trees for subsets of this taxon set. In the present section, we approach a different aspect of heredity: Given an alignment, is it possible to combine small compatible MP trees, e.g. quartets, to derive an MP tree for the entire set of taxa? Intuitively, one might think that it is quite likely that this is true, as the assumption of compatibility of the MP-quartets is a strong condition. Moreover, one might think that if additionally the MP-quartets are all unique, which is another strong condition, it is even more likely for such a statement to hold. However, in this section we present a counterexample which shows that even under these seemingly ideal conditions it may be impossible to infer large MP trees from the smaller ones.

Proposition 3 *There exists an alignment S on a taxon set X such that the following conditions hold:*

1. *the MP tree \mathcal{T} of S is unique and*
2. *all most parsimonious quartet trees, i.e. MP trees on taxa sets $\{x_1, x_2, x_3, x_4\} \subseteq X$, on the corresponding subalignments of S are compatible with an X -tree $\hat{\mathcal{T}}$ and*
3. *all such quartet trees are unique and*
4. *$\mathcal{T} \neq \hat{\mathcal{T}}$.*

Proof We prove the proposition by providing an explicit counterexample. Consider the following binary alignment:

$$S := f_1 f_1 f_2 f_2 f_2 f_8 f_8 f_8 f_9 f_9 f_9 f_{10} f_{10} f_{10} = \begin{cases} \text{AAAAAAAAAAAAAAAA} \\ \text{AACCCAAAAAAAAAAA} \\ \text{CCAAACCCCCCAAAA} \\ \text{CCCCCCCCAAACCCC} \\ \text{CCCCCAAACCCCCC} \end{cases}$$

So S consists of two copies of f_1 , three copies of f_2 , three copies of f_8 , three copies of f_9 and four copies of f_{10} , where f_1, \dots, f_{10} are defined as in the proof of Theorem 2.

Using an exhaustive search through the space of all 15 trees on 5 taxa, which for binary alignments is provided e.g. by the ‘Penny’ program of the Phylip package (Felsenstein 2005), we find that S has the unique MP tree $\mathcal{T} = ((1, 3), 2, (4, 5))$ depicted in Fig. 9. The 4-taxa subalignments of S are depicted in Table 3. Now, these subalignment $S - 1, \dots, S - 5$ each have unique MP quartet trees, which are also depicted in Table 3. Note that trees $\hat{\mathcal{T}} - 1, \hat{\mathcal{T}} - 2, \hat{\mathcal{T}} - 3, \hat{\mathcal{T}} - 4$ and $\hat{\mathcal{T}} - 5$ are all compatible with tree $\hat{\mathcal{T}} := ((1, 2), 3, (4, 5))$ depicted in Fig. 9, whereas $\hat{\mathcal{T}} - 4$ and $\hat{\mathcal{T}} - 5$ are incompatible with \mathcal{T} . So the unique and compatible MP quartets cannot be combined to give the unique MP tree for the whole alignment. This completes the proof. □

While Sect. 3.1 shows that in general large MP trees cannot be used to infer smaller MP trees on subsets of the taxon set, the above example shows that the opposite is also impossible, even under strong compatibility conditions. So MP is a phylogenetic

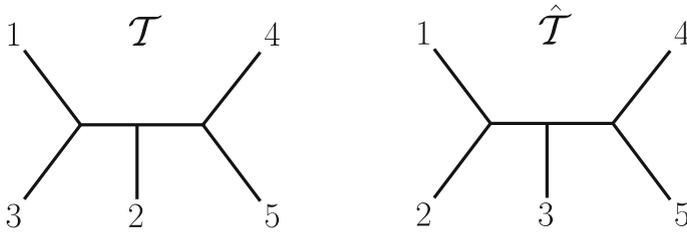


Fig. 9 Tree \mathcal{T} is the unique MP tree of S as defined in the proof of Proposition 3, but tree $\hat{\mathcal{T}} = ((1, 2), 3, (4, 5))$ is the only tree that is compatible with all unique MP quartet trees

Table 3 Overview of the 4-taxa subalignments of alignment S and their corresponding unique MP trees as described in the proof of Proposition 3.

$S - 1$	AACCCAAAAAAAAA CCAAACCCCCAAAA CCCCCCCCAAACCC CCCCCAAACCCCCC	$\hat{\mathcal{T}} - 1$ 	$S - 2$	AAAAAAAAAAAAAAAA CCAAACCCCCAAAA CCCCCCCCAAACCC CCCCCAAACCCCCC	$\hat{\mathcal{T}} - 2$
$S - 3$	AAAAAAAAAAAAAAAA AACCCAAAAAAAAA CCCCCCCCAAACCC CCCCCAAACCCCCC	$\hat{\mathcal{T}} - 3$ 	$S - 4$	AAAAAAAAAAAAAAAA AACCCAAAAAAAAA CCAAACCCCCAAAA CCCCCAAACCCCCC	$\hat{\mathcal{T}} - 4$
$S - 5$	AAAAAAAAAAAAAAAA AACCCAAAAAAAAA CCAAACCCCCAAAA CCCCCCCCAAACCC	$\hat{\mathcal{T}} - 5$ 			

tree inference method that may find that unique ‘best’ trees are unrelated to ‘best’ trees on subsets or supersets of the taxa under consideration. As this is somewhat counterintuitive, naturally the question arises whether this problem only occurs with MP or also affects other methods.

Remark 1 By [Tuffley and Steel \(1997\)](#), the results of Propositions 2 and 3 can be directly generalized to Maximum Likelihood under the N_r -model. If we assume no common mechanism, all examples provided in the preceding sections immediately lead to analogous results for Maximum Likelihood by Theorem 5 of [Tuffley and Steel \(1997\)](#). If a common mechanism is assumed, the examples provided in these sections may need to be modified in the sense of adding constant characters. These extra characters do not change the MP tree as constant characters are non-informative, but they will make ML agree with MP according to Theorem 7 of [Tuffley and Steel \(1997\)](#). So in both cases, we derive alignments for which the ML tree is not hereditary.

4 Discussion

In this paper, we presented various examples of non-hereditary Maximum Parsimony and Maximum Likelihood trees together with an idea of how to construct them as solutions to systems of inequalities. The results show that there are alignments for which the ‘best’ tree with respect to one of these phylogenetic tree inference methods does not have to be related to the ‘best’ tree on fewer taxa. Also, even if a tree is constructed from uniquely ‘best’ and compatible quartet trees, it might not coincide with the ‘best’ tree on all taxa. On the one hand, these facts might help to understand why tree reconstruction for MP and ML is hard (Foulds and Graham 1982; Roch 2006; Chor and Tuller 2006). On the other hand, the result is surprising and gives rise to new questions, e.g. whether or not one should include outgroups when inferring trees, as these might change the tree topology of the optimal tree. Naturally, it would also be interesting to know whether similar heredity problems occur with other methods, such as e.g. ML under more complicated models of nucleotide substitution. We conjecture that other methods are also not hereditary in the sense described in this paper.

Acknowledgments Some of the results underlying this paper were achieved during a summer studentship funded by the New Zealand Marsden Fund and hosted by the Allan Wilson Centre for Molecular Ecology and Evolution. Support from the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) to this work is greatly appreciated. I also wish to thank Mike Steel, Arndt von Haeseler and Anne Kupczok for helpful discussions on the topic and Benny Chor for asking a question that led to the results presented in Sect. 3.2. Also, I wish to thank Arndt von Haeseler for putting his conjecture on the Isaac Newton Institute’s list of open phylogenetic problems. Last but not least, I want to thank two anonymous reviewers for helpful suggestions on the manuscript.

References

- Bandelt HJ, Dress A (1986) Reconstructing the shape of a tree from observed dissimilarity data. *Adv Appl Math* 7:309–343
- Chor B, Tuller T (2006) Finding a maximum likelihood tree is hard. *J ACM* 53:722–744
- Felsenstein J (1978) Cases in which parsimony or compatibility will be positively misleading. *Syst Zool* 27:401–410
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Massachusetts
- Felsenstein J (2005) *Phylip (phylogeny inference package version 3.6)*. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle
- Felsenstein J, Archie J, Day W, Maddison W, Meacham C, Rohlf F, Swofford D (2000) The newick tree format. <http://evolution.genetics.washington.edu/phylip/newicktree.html>
- Fitch W (1971) Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Zool* 20(4):406–416
- Foulds L, Graham R (1982) The steiner problem in phylogeny is np-complete. *Adv Appl Math* 3:43–49
- Jukes T, Cantor C (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism*. Academic Press, New York, pp 21–132
- Neyman J (1971) Molecular studies of evolution: A source of novel statistical problems. In: *Statistical Decision Theory and Related Topics*. Academic Press, New York, pp 1–27
- Roch S (2006) A short proof that phylogenetic tree reconstruction by maximum likelihood is hard. *IEEE/ACM Trans Comp Biol Bioinform* 3:92–94
- Semple C, Steel M (2003) *Phylogenetics*. Oxford University Press
- Tuffley C, Steel M (1997) Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull Math Biol* 59:581–607

- von Haeseler A (2007) Hereditary maximum parsimony trees. In: Steel M (ed) Phylogenetics: challenges and conjectures, Isaac Newton Institute for Mathematical Sciences. <http://www.newton.ac.uk/programmes/PLG/conj.pdf>
- von Haeseler A (2009) Hereditary maximum parsimony trees. In: Steel M (ed) Penny Ante List of open problems in phylogenetics, Biomathematics Research Centre of the Allan Wilson Centre for Molecular Ecology and Evolution. <http://www.math.canterbury.ac.nz/bio/events/kaikoura09/penny.shtml>
- Yang Z (2006) Computational Molecular Evolution. Oxford University Press, Oxford