Molecular Phylogenetics and Evolution 80 (2014) 165–168

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

The most parsimonious tree for random data

Mareike Fischer [a], Michelle Galla [a], Lina Herbst [a], Mike Steel [b,*]

[a] Department for Mathematics and Computer Science, Ernst-Moritz-Arndt University, Greifswald, Germany
[b] Allan Wilson Centre, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Article history:
Received 2 June 2014
Revised 18 July 2014
Accepted 19 July 2014
Available online 29 July 2014

Keywords:
Tree
Maximum parsimony
Random data
Central limit theorem

ABSTRACT

Applying a method to reconstruct a phylogenetic tree from random data provides a way to detect whether that method has an inherent bias towards certain tree 'shapes'. For maximum parsimony, applied to a sequence of random 2-state data, each possible binary phylogenetic tree has exactly the same distribution for its parsimony score. Despite this pleasing and slightly surprising symmetry, some binary phylogenetic trees are more likely than others to be a most parsimonious (MP) tree for a sequence of k such characters, as we show. For $k = 2$, and unrooted binary trees on six taxa, any tree with a caterpillar shape has a higher chance of being an MP tree than any tree with a symmetric shape. On the other hand, if we take any two binary trees, on any number of taxa, we prove that this bias between the two trees vanishes as the number of characters k grows. However, again there is a twist: MP trees on six taxa for $k = 2$ random binary characters are more likely to have certain shapes than a uniform distribution on binary phylogenetic trees predicts. Moreover, this shape bias appears, from simulations, to be more pronounced for larger values of k.

1. Introduction

The 'shape' of reconstructed evolutionary trees is of interest to evolutionary biologists, as it should provide some insight into the processes of speciation and extinction (Aldous, 2001; Aldous et al., 2011; Hey, 1992; Holton et al., 2014; Lambert et al., 2013; Stadler, 2013). In this paper, 'shape' refers just to the discrete shape of the tree (i.e. we ignore the branch lengths); the advantages of this are that it simplifies the analysis, and it also confers a certain robustness (i.e. the resulting probability distribution on discrete shapes is often independent of the fine details of an underlying speciation/extinction model (Aldous, 1995; Lambert et al., 2013)). For example, if all speciation (and extinction) events affect all taxa at any given epoch in the same way, then we should expect the shape of a reconstructed tree to be that predicted by the discrete 'Yule–Harding' model (Aldous, 2001; Harding, 1971; Lambert et al., 2013). In fact, a general trend (see e.g. Aldous, 2001) is that the shape of phylogenetic trees reconstructed from biological data tends to be a little less balanced than this model predicts, but is more balanced than what would be obtained under a uniform model in which each binary phylogenetic tree has the same probability (this model is sometimes also called the 'Proportional-to-Distinguishable-Arrangements' (PDA) model (Rosen, 1978)).

There are, however, other factors which can lead to biases in tree shape. One is non-random sampling of the taxa on which to construct a tree (influenced, for example, by the particular interests of the biologists or the application of a certain strategy to sample taxa). Another cause of possible bias is that a tree reconstruction method may itself have an inherent preference towards certain tree shapes. A way to test this latter possibility is to apply the tree reconstruction method to data that contain no phylogenetic signal at all, in particular, purely random data, where each character is generated independently by a process that assigns states to the taxa uniformly (e.g. by the toss of a fair coin in the case of two states). For some methods, such as 'TreePuzzle', such data leads to very balanced trees (similar to the Yule–Harding model (Vinh et al., 2010; Zhu et al., 2013)). However, other methods, such as maximum likelihood and maximum parsimony, lead to less balanced trees, that are closer in shape to the uniform model, as recently reported in Holton et al. (2014). In the case of maximum parsimony, the two-state symmetric model has the even-handed property that every binary tree has exactly the same distribution of its parsimony score on k randomly generated characters. Thus, it might be supposed that the maximum parsimony (MP) tree for such a sequence of characters would also follow a uniform distribution. However, while this holds in special cases, it does not hold in general, as we show below.

* Corresponding author. Fax: +64 33642587.
E-mail address: mike.steel@canterbury.ac.nz (M. Steel).

http://dx.doi.org/10.1016/j.ympev.2014.07.010

1.1. Trees and parsimony: definitions and basic properties

In phylogenetics, graphs, especially trees, are used to describe the ancestral relationships among different species. A main goal of phylogenetics is to infer an evolutionary tree from data available from present-day species. In graph theory, a tree $T = (V, E)$ consists of a connected graph with no cycles. Certain leaf-labelled trees ('phylogenetic trees') are widely used where the set of extant species label the leaves and the remaining vertices represent ancestral speciation events (Felsenstein, 2004). There are different methods of reconstructing a phylogenetic tree. One of the most famous tree reconstruction methods is maximum parsimony. For a given tree and discrete character data, the parsimony score can be found in polynomial time by using the Fitch–Hartigan algorithm (Fitch, 1971; Hartigan, 1973). The parsimony score counts the number of changes (mutations) required on the tree to describe the data. This problem of finding the optimal parsimony score for a given tree is often called the 'small parsimony' problem. The 'big parsimony' problem aims at finding the most parsimonious tree ('MP tree') amongst all possible trees. This problem has been proven to be NP-hard (Foulds and Graham, 1982).

In this paper, we assume that each taxon from the leaf set X of the tree is assigned a binary state (0 or 1) independently, and with equal probability (the case where the two states have different probabilities is less interesting, since then the distribution of the parsimony score of a fixed binary tree is easily seen to depend on the shape of the tree, even for a single character). This process is then repeated (also independently) to generate a sequence of characters (defined formally below). For binary trees with random data, we are interested in the probability that a tree is an MP tree, and also what happens when the length of the sequences or the number of leaves gets larger. In particular, we wish to determine whether each tree is equally likely to be selected as an MP tree.

Definition 1 (*Binary phylogenetic trees*). An *(unrooted) binary phylogenetic X-tree* is a tree T with leaf set X and with every interior (i.e. non-leaf) vertex of degree exactly three. We will let $UB(X)$ be the set of unrooted binary phylogenetic X-trees. When $X = [n] = \{1, \dots, n\}$, we will write $UB(n)$.

Definition 2. [Character, extension, parsimony score]

- A *character on X* over a finite set R of character states is any function f from X into R; $f : X \to R$. In this paper we will consider two-state characters; $f : X \to \{0, 1\}$.
- A function $\bar{f} : V \to R$ such that $\bar{f}|_X = f$ is said to be an *extension* of f since it describes an assignment of states to all vertices of T that agrees with the states that f stipulates at the leaves.
- Let $ch(\bar{f}, T) := \left|\{e = \{u, v\} \in E : \bar{f}(u) \neq \bar{f}(v)\}\right|$ be the *changing number* of $\bar{f}$. Given a character $f : X \to R$, the *parsimony score of f* on T, denoted $ps(f, T)$, is the smallest changing number of any extension of f, i.e.:

  $$ps(f, T) := \min_{\bar{f}: V \to R, \bar{f}|_X = f} \{ch(\bar{f}, T)\}.$$

  An extension $\bar{f}$ of f for which $ch(\bar{f}, T) = ps(f, T)$ is said to be a *minimal extension*.

  Let $\mathcal{C} = (f_1, \dots, f_k)$ be a sequence of characters on X. The *parsimony score of* $\mathcal{C}$ on T, denoted $ps(\mathcal{C}, T)$, is defined by $ps(\mathcal{C}, T) := \sum_{i=1}^{k} ps(f_i, T)$.

2. Comparing given trees

Let $X_k(T)$ be the parsimony score of k random two-state characters on $T \in UB(n)$. We will see shortly (Proposition 1) that the distribution of $X_k(T)$ does not depend on the shape of T; it just depends on n. Notice that $X_k(T) = X_1 + X_2 + \cdots + X_k$, where X_i (for $i = 1, \dots, k$) form a sequence of independent and identically distributed random variables (with common distribution $X_1(T)$). If $\mathbb{P}(X_k(T) = l)$ denotes the probability that T has parsimony score l then, from Steel (1993), we have, for each $l \in [1, \lfloor n/2 \rfloor]$:

$$\mathbb{P}(X_1(T) = l) = \frac{2n - 3l}{l} \cdot \binom{n - l - 1}{l - 1} \cdot 2^{l-n}, \tag{1}$$

with $\mathbb{P}(X_1(T) = 0) = 2^{1-n}$ and $\mathbb{P}(X_1(T) = l) = 0$ for $l > \lfloor n/2 \rfloor$. Furthermore, $\mathbb{E}[X_1(T)] = \frac{3n - 2 - \left(-\frac{1}{2}\right)^{n-1}}{9} \sim \frac{n}{3}$ is the expected parsimony score of T, and $\mathbb{E}[X_k(T)] = k \cdot \mathbb{E}[X_1(T)]$. An immediate consequence of (1) is the following.

Proposition 1. *For every* $k \geqslant 1$ and $n \geqslant 2$, *the distribution of the parsimony score of k independent random binary characters (i.e.* $X_k(T)$*) is the same for all* $T \in UB(n)$.

2.1. Comparing two trees by their parsimony score

We begin this section by describing a tree rearrangement operation on binary phylogenetic trees (Semple and Steel, 2003, Chapter 2.6), namely tree bisection and reconnection (TBR). Let T be a binary phylogenetic X-tree and let $e = \{u, v\}$ be an edge of T. A TBR operation is described as follows. Let T' be the binary tree obtained from T by deleting e, adding an edge between a vertex that subdivides an edge of one component of $T \setminus e$ and a vertex that subdivides an edge of the other component of $T \setminus e$, and then suppressing any resulting degree-two vertices. In the case that a component of $T \setminus e$ consists of a single vertex, then the added edge is attached to this vertex. T' is said to be obtained from T by a single TBR operation.

Proposition 2. *Let* $T, T' \in UB(n)$.

- (i) *If T and T′ are one TBR apart, then* $\mathbb{P}(X_k(T) < X_k(T')) = \mathbb{P}(X_k(T') < X_k(T))$ *holds for all* $k \geqslant 1$.
- (ii) *If T and T′ are more than one TBR apart, then the equality* $\mathbb{P}(X_k(T) < X_k(T')) = \mathbb{P}(X_k(T') < X_k(T))$ *can fail, even for* $k = 1$ *and* $n = 6$.

Proof.

(i) From Bryant (2004, Lemma 5.1), if T and T' are one TBR apart then for any character f, $\left|ps(f, T) - ps(f, T')\right| \leqslant 1$. In particular,

$$\left|X_1(T) - X_1(T')\right| \leqslant 1. \tag{2}$$

For $k \geqslant 1$, let $\Delta_k = X_k(T) - X_k(T')$. Then if $T, T' \in UB(n)$ are one TBR apart, then $\Delta_1 = X_1(T) - X_1(T')$ is either 0, 1 or -1, by (2). Moreover, $\mathbb{P}(\Delta_1 = m) = \mathbb{P}(\Delta_1 = -m)$ for all $m \in \{0, 1 - 1\}$, since $\mathbb{E}[\Delta_1] = 0$, by Proposition 1. Furthermore, $\Delta_k = D_1 + \cdots + D_k$, where $D_1, \dots, D_k$ are independent and identically distributed as Δ_1, so we have:

$$\begin{aligned}
\mathbb{P}(\Delta_k = m) &= \sum_{\substack{m_1, \dots, m_k \in \{-1,0,1\}: \\ m_1 + \cdots + m_k = m}} \mathbb{P}(D_1 = m_1 \wedge D_2 = m_2 \wedge \cdots \wedge D_k = m_k) \\
&= \sum_{\substack{m_1, \dots, m_k \in \{-1,0,1\}: \\ m_1 + \cdots + m_k = m}} \prod_{j=1}^{k} \mathbb{P}(D_j = m_j) = \sum_{\substack{m_1, \dots, m_k \in \{-1,0,1\}: \\ m_1 + \cdots + m_k = m}} \prod_{j=1}^{k} \mathbb{P}(D_j = -m_j) \\
&= \sum_{\substack{m'_1, \dots, m'_k \in \{-1,0,1\}: \\ m'_1 + \cdots + m'_k = -m}} \mathbb{P}(D_1 = m'_1 \wedge D_2 = m'_2 \wedge \cdots \wedge D_k = m'_k) = \mathbb{P}(\Delta_k = -m).
\end{aligned}$$

This provides the equality $\mathbb{P}(X_k(T) < X_k(T')) = \mathbb{P}(X_k(T') < X_k(T))$ for all $k \geqslant 1$.

(ii) We prove this by exhibiting one counterexample, namely the trees shown in Fig. 1. Let $\Delta_k = X_k(T) - X_k(T')$. The equality $\mathbb{P}(X_k(T) < X_k(T')) = \mathbb{P}(X_k(T') < X_k(T))$ is equivalent to $\mathbb{P}(\Delta_k < 0) = \mathbb{P}(\Delta_k > 0)$.

By calculating the parsimony score for the 32 different two-state characters (without loss of generality we set $f(1) := 0$) we can assign the values that Δ_1 can take and the probability of those values. $\Delta_1 = -2$ occurs precisely when $X_1(T) = 1$ and $X_1(T') = 3$ with probability $p = \frac{1}{32}$. $\Delta_1 = -1$ occurs precisely when $X_1(T) = 1$ and $X_1(T') = 2$ or $X_1(T) = 2$ and $X_1(T') = 3$ with probability $q = \frac{3}{32}$. $\Delta_1 = +1$ occurs precisely when $X_1(T) = 2$ and $X_1(T') = 1$ or $X_1(T) = 3$ and $X_1(T') = 2$ with probability $r = \frac{5}{32}$. Since $\Delta_1 = +2$ is not possible, $\Delta_1 = 0$ with probability $1 - (p+q+r) = \frac{23}{32}$. This leads to $\mathbb{P}(\Delta_1 < 0) = \mathbb{P}(\Delta_1 = -2) + \mathbb{P}(\Delta_1 = -1) = \frac{4}{32} < \frac{5}{32} = \mathbb{P}(\Delta_1 = +1) = \mathbb{P}(\Delta_1 > 0)$. Therefore $\mathbb{P}(X_k(T) < X_k(T')) < \mathbb{P}(X_k(T') < X_k(T))$ holds for $k = 1$ and the choice of T and T' shown in Fig. 1. In other words, the probability that the symmetric tree T is more parsimonious than the caterpillar tree T' (on a single random binary character) is higher than the probability that T' is more parsimonious than T. □

3. Maximum parsimony trees

Definition 3 (*Maximum parsimony tree*). Given a sequence $\mathcal{C} = (f_1, \ldots, f_k)$ of characters on X, a phylogenetic tree T on X that minimises $ps(\mathcal{C}, T)$ is said to be a *maximum parsimony (MP) tree* for $\mathcal{C}$. The corresponding ps-value is the *parsimony or MP score of* $\mathcal{C}$, denoted $ps(\mathcal{C})$.

Notation: Given $T \in UB(n)$, let $mp_k(T)$ denote the probability that T is an MP tree for $k \geqslant 1$ random two-state characters on $[n]$. That is

$$mp_k(T) := \mathbb{P}(X_k(T) \leqslant \min_{T' \in UB(n)} \{X_k(T')\}).$$

Notice that $mp_k(T)$ is not a probability distribution on $UB(n)$ since the positive probability of ties for the most parsimonious tree ensures that the $mp_k(T)$ values will sum to a value greater than 1.

Lemma 1. *If $T_1, T_2 \in UB(n)$ have the same shape then $mp_k(T_1) = mp_k(T_2)$.*

Proof. Let $k \geqslant 1$ and let $f_1, \ldots, f_k$ be two-state characters. Then $ps(f_1, \ldots, f_k, T) = ps(f_1^\sigma, \ldots, f_k^\sigma, T^\sigma)$, where σ is an element of the group S_n of permutations on the leaf set $[n]$ of T. Notice that the map $f = (f_1, \ldots, f_k) \mapsto f^\sigma = (f_1^\sigma, \ldots, f_k^\sigma)$ is a bijection, so the number of characters f for which T is an MP tree for f equals the number of characters f for which T^σ is an MP tree for f. □

It follows from Proposition 1 and Lemma 1 that if $n \geqslant 3$ and $k = 1$, or if $k \geqslant 1$ and $n \leqslant 5$, then $mp_k(T)$ is constant for all $T \in UB(n)$. However, this does not hold more generally, as we now state.

Theorem 1. *$mp_k(T)$ is not constant for all $T \in UB(n)$ when $n = 6$ and $k = 2$. In particular, any given caterpillar tree (like T in Fig. 1) has a higher probability of being an MP tree than a symmetric tree (like T' in Fig. 1).*

The proof of Theorem 1 requires a detailed case analysis to identify the MP tree(s) for all pairs of characters (f_1, f_2); details are provided in the online supplementary material. The result is also confirmed by simulations, which are provided in the following section.

4. Asymptotic analysis

We first show that the bias exhibited in Proposition 2(ii) disappears asymptotically but the bias apparent in Theorem 1 does not.

Proposition 3. *For all $T, T' \in UB(n)$ and all n:*

$$\lim_{k\to\infty} \mathbb{P}(X_k(T) < X_k(T')) = \frac{1}{2}.$$

Proof. Let $T, T' \in UB(n)$ and $k \geqslant 1$, and let $\Delta_k = X_k(T) - X_k(T') = D_1 + D_2 + \cdots + D_k$, where the random variable $D_i = ps(f_i, T) - ps(f_i, T')$ $(i = 1, \ldots, k)$ and the D_i are independent and identically distributed. Moreover $\mathbb{E}[D_i] = 0$ and D_i has a standard deviation σ that is strictly positive and finite. To see that $\sigma > 0$, note that $\sigma^2 \geqslant \mathbb{P}[D_i \neq 0]$ by Chebychev's inequality, and D_i is nonzero whenever f_i corresponds to a two-state character that has parsimony score 1 on one of the trees T, T' and parsimony score greater than 1 on the other tree (at least one such character must exist, since $T \neq T'$, and every tree is uniquely determined by its characters of parsimony score 1). We can now apply the standard central limit theorem to deduce that for an asymptotically standard normal variable $Z_k = \frac{\Delta_k - \mathbb{E}[\Delta_k]}{\sigma\cdot\sqrt{k}}$, we have:

$$\mathbb{P}(\Delta_k < 0) = \mathbb{P}\left(Z_k < \frac{0-0}{\sigma \cdot \sqrt{k}}\right) \overset{k\to\infty}{\longrightarrow} \frac{1}{2}. \qquad \square$$

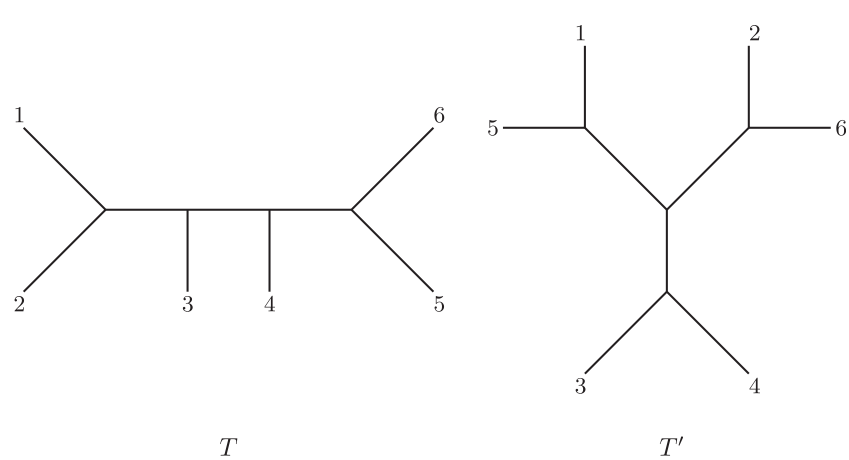

Fig. 1. Two trees $T, T' \in UB(6)$. Note that T and T' are more than one TBR apart.

Table 1
Overview of simulation results: For each alignment length, 1000 runs were evaluated.

Al. length	Av. # MP trees	# Symmetric tree was MP	# Caterpillar was MP	$\frac{\#\text{ Symmetric MP trees}}{\#\text{MP trees}}$
2	17.177	2375	14,802	0.138266
10	3.908	365	3543	0.0933982
100	1.622	119	1503	0.0733662
1000	1.166	59	1107	0.0506003
10,000	1.053	57	996	0.0541311
100,000	1.013	46	967	0.0454097

Finally, we consider the limiting behaviour of $mp_k(T)$ as $k \to \infty$, and present simulations that suggest that even for $n = 6$, this probability depends on the shape of the tree. It is easily shown that, for any $n > 1$, as $k \to \infty$, there is a unique most parsimonious tree, so $\sum_{T \in UB(n)} \lim_{k \to \infty} mp_k(T) = 1$ (see e.g. (Zhu et al., 2013) (Theorem 4(2)). In other words, $\lim_{k \to \infty} mp_k(T)$ is a probability distribution on $UB(n)$. However, the additional claim there that $mp_k(T)$ is uniform on $UB(n)$ does not hold when $n = 6$ and when either $k = 2$ (Theorem 1) or, it seems, as $k \to \infty$, as we now explain.

4.1. Simulations

We used the computer algebra system *Mathematica* to generate alignments of lengths 2, 10, 100, 1,000, 10,000 and 100,000, respectively, by sampling characters for six taxa uniformly at random out of the 32 possible binary characters (we assume without loss of generality that the state of taxon 1 is fixed, say, to state 0, whereas all other taxa can choose states 0 or 1). For each alignment, we ran an exhaustive search through the tree space of 105 unrooted binary phylogenetic trees in order to find all MP trees. For each alignment length, we did 1000 runs and we counted the average number of MP trees, as well as the number of times that each of the two tree shapes for six taxa (the caterpillar shape or the symmetric shape of T and T' in Fig. 1) were amongst the MP trees. We then calculated the ratio of the number of MP trees with a symmetric shape divided by the total number of MP trees. Note that this ratio should equal $\frac{1}{7} \approx 0.142857$ under the uniform (PDA) model, because 15 out of the 105 possible binary trees on six leaves have the symmetric shape (in general, the number of unrooted binary trees having shape τ is $n!/s(\tau)$ where $s(\tau)$ is the number of leaf permutations that fix the topology of a tree having shape τ; this gives $6!/(3!2!^3) = 15$ in our case; for further details see (Semple and Steel, 2003, Section 2.4). However, the last column of Table 1 reveals that only for the extremely short alignment of length 2 the ratio is close to this value in our simulations (and it is not exactly equal to it, by Theorem 1). Moreover, the ratio decreases away from $\frac{1}{7}$ as the alignment length increases (the small variation at alignment length 10,000 is within one standard deviation). This trend and the reported values strongly suggest that the limiting value of $mp_k(T)$ is not uniform across all trees in $UB(6)$. Note also that column 2 of Table 1 is also consistent with the finding mentioned earlier that there will be a unique MP tree with probability converging to 1 as k grows.

4.2. Concluding comments

In one sense, the two-state symmetric model is as favourable to all binary phylogenetic trees as is possible under maximum parsimony, since each tree has exactly the same probability distribution on the parsimony score of k random characters. Moreover, Proposition 3 shows that no one tree is any more likely to be an MP tree than another. It may seem somewhat surprising, therefore, that the distribution of MP trees is not uniform, even asymptotically; however this has a simple explanation. Although the characters are generated independently, and their parsimony scores is also independently distributed on any given binary tree, the MP binary tree is chosen once the k characters are given. Thus these characters are not independent random variables once we condition on a given tree being the MP tree for these characters. Moreover, once one moves away from the simple two-state model (for example, to the r-state symmetric model, or the 2-state non-symmetric model) even the uniformity of MP scores on fixed trees disappears (Steel, 1993). In summary, while maximum parsimony on random data seems, in certain senses (described above), to favour each binary tree equally, the method nevertheless exhibits a bias towards trees with certain tree shapes.

Acknowledgments

We thank the Allan Wilson Centre for help funding this work. We also thank David Bryant for pointing out that MP trees might not be uniformly distributed on sequences of random characters, and an anonymous reviewer for some helpful suggestions.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.07.010.

References

Aldous, D., 1995. Probability distributions on cladograms. In: Aldous, D., Pemantle, R.E. (Eds.), Random Discrete Structures, IMA Volumes in Mathematics and its Applications, vol. 76. Springer, pp. 1–18.
Aldous, D., 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat. Sci. 16, 23–34.
Aldous, D., Krikun, M., Popovic, L., 2011. Five statistical questions about the tree of life. Syst. Biol. 60, 318–328.
Bryant, D., 2004. The splits in the neighborhood of a tree. Ann. Comb. 8, 1–11.
Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates.
Fitch, W.M., 1971. Towards defining the course of evolution: minimum change for a species tree topology. Syst. Zool. 20, 406–416.
Foulds, L.R., Graham, R.L., 1982. The Steiner problem in phylogeny is NP-complete. Adv. Appl. Math. 3, 43–49.
Harding, E.F., 1971. The probabilities of rooted tree shapes generated by random bifurcation. Adv. Appl. Math. 3, 44–77.
Hartigan, J.A., 1973. Minimum mutation fits to a given tree. Biometrics 29, 53–65.
Hey, J., 1992. Using phylogenetic trees to study speciation and extinction. Evolution 46, 627–640.
Holton, T.A., Wilkinson, M., Pisani, D., 2014. The shape of modern tree reconstruction methods. Syst. Biol. 63 (3), 436–441.
Lambert, A., Stadler, T., 2013. Birth–death models and coalescent point processes: the shape and probability of reconstructed phylogenies. Theor. Popul. Biol. 90, 113–128.
Rosen, D.E., 1978. Vicariant patterns and historical explanation in biogeography. Syst. Biol. 27, 159–188.
Semple, C., Steel, M., 2003. Phylogenetics. Oxford University Press (Oxford Lecture Series in Mathematics and its Application).
Stadler, T., 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26, 1203–1219.
Steel, M.A., 1993. Distributions on bicoloured binary trees arising from the principle of parsimony. Discr. Appl. Math. 41 (3), 245–261.
Vinh, L.S., Fuehrer, A., von Haeseler, A., 2010. Random tree-puzzle leads to the Yule–Harding distribution. Mol. Biol. Evol. 28 (2), 873–877.
Zhu, S., Steel, M., 2013. Does random tree puzzle produce Yule–Harding trees in the many-taxon limit? Math. Biosci. 243, 109–116.