



ELSEVIER

Computational Statistics & Data Analysis 38 (2000) 273–283

COMPUTATIONAL
STATISTICS
& DATA ANALYSIS

www.elsevier.com/locate/cstda

Choosing the tree which actually best explains the data: another look at the bootstrap in phylogenetic reconstruction

V. Berry*, O. Gascuel, G. Caraux

*Département d'Informatique Fondamentale et Appliquée, LIRMM, 161 rue Ada,
34392 Montpellier cedex 5, France*

Abstract

We consider the problem of phylogenetic reconstruction, which consists in estimating the evolutionary history of a set of species. This unknown history is modelled as a tree and estimated from nucleotide sequences taken from the species' genome. The first goal of the estimation is to produce a tree which is structurally as close as possible to the true tree. However, most phylogenetic tree-building methods rely on optimization criteria which lead to inferring fully resolved trees, i.e. models of maximal complexity. Thus, such trees usually contain some wrong edges, too specific to the data, i.e., resulting from an overfitting effect.

We first introduce a structural goodness-of-fit criterion based on *quartets* of species. Then we describe a tree-building method inferring a fully resolved tree by optimizing this criterion. We present two descending approaches to remove unreliable edges from this tree. The first one relies on the bootstrap process (Efron, 1979) as introduced in the phylogenetic field by Felsenstein (1985). The second one is original in this context but analogous to usual methods in model calibration.

Simulations show the efficiency of both approaches, in that the structural distance between the true tree and the estimated tree is significantly reduced. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Phylogeny reconstruction; Tree estimation; Overfitting; Complexity/goodness-of-fit; Bootstrap resampling methods; Cross validation; Structural distance; Four-point condition; Quartets

* Corresponding author.

E-mail address: vberry@lirmm.fr (V. Berry)

1. Introduction

The evolution of living species may be represented by a tree (called a *phylogeny* or an *evolutionary tree*) whose leaves correspond to contemporary species and internal nodes to ancestral species. The aim of phylogenetic analysis (Swofford and Olsen, 1990) is to recover this tree. Most of the time, nucleotide sequences taken from the species' genome are used for this purpose. These sequences are homologous, which means that they originate from a same ancestral sequence that evolved differently along the edges of the evolutionary tree, from the ancestral species at the root to the contemporary species. The numerous worldwide sequencing projects have generated huge amounts of sequences, which may be used to perform phylogenetic analysis. On the other hand, phylogenetic analysis appears as an essential tool to organize and understand this material.

More formally, let S be the set of n species under study, whose evolutionary tree t is to be estimated. This tree is usually considered as fully resolved, i.e., as binary (internal nodes have degree 3). Different lengths are attributed to the edges of t , measuring the time or amount of evolutionary events between the two extremities of each edge. The path-lengths (sum of the lengths of the edges on a path) induce a tree distance $D=(d_{ij})$ on S , usually represented as an $n \times n$ matrix. Here we suppose that t , and consequently D , is unknown. The aim is to retrieve the structure of the speciation process, i.e. the subsets into which species are split by the branching pattern of t , on the basis of molecular data. This data are organized into a matrix X consisting of n sequences of p nucleotides (coded as letters: A,C,G or T). Columns of X can be considered as a sample of homologous sites. If we knew the entire genome of the species S , then we could compute D from X , assuming that the evolution conforms to a known mathematical model, such as those of Jukes and Cantor (1969) or Kimura (1980). From the additive matrix D we could then retrieve the unique tree t with which it is associated (Sattah and Tversky, 1977; Barthélemy and Guénoche, 1991) and our goal would then be achieved. Knowing only a segment of the species' genome, we can only compute an estimation of the unknown matrix D . This estimation \hat{D} is used to estimate t by minimization of a certain goodness-of-fit criterion δ . This function measures the accuracy of \hat{t} to represent \hat{D} . Let τ denote any valued tree whose leaves are associated with the species of S . $\delta(D, \tau)$ is minimum when $\tau = t$ and \hat{t} is, therefore, defined as

$$\hat{t} = \underset{\tau}{\operatorname{argmin}} \delta(\hat{D}, \tau).$$

The tree \hat{t} will generally be fully resolved. Indeed, adding any new ramification to the tree may only improve the value of the criterion. However, is the complete resolution of \hat{t} relevant? Would not the multiplicity of edges lead to an overfitting effect? By seeking to adjust \hat{D} as closely as possible, do we not wander from D at some given step, inferring edges that result in minimizing $\delta(\hat{D}, \hat{t})$ but at the same time increasing $\delta(D, \hat{t})$?

Thus, we find ourselves faced with a very classical problem in model selection (or calibration), that of finding a good compromise between complexity and (apparent) goodness-of-fit. By introducing a decomposition of the structural estimation

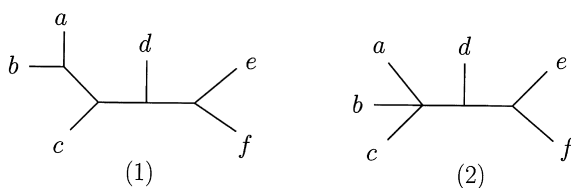


Fig. 1. Removing the bipartition $\{a,b\}|\{c,d,e,f\}$.

error in terms of bias and variance, we have shown (Berry and Gascuel, 1996) that this dilemma can also be translated in terms of bias/variance compromise, as often in an estimation context. Indeed, a new sample of sites would very probably put at least a part of \hat{t} 's structure into question. This potential variability raises doubts about the reliability of inferring fully resolved trees. For our purpose it is more interesting to search for less resolved, and thus more similar structures. In return, we have to accept a structural bias in terms of *complexity*, i.e., number of edges.

Searching for a less complex tree than \hat{t} is connected with the problem of estimating the reliability of tree edges. Among other methods (see Li and Gouy, 1990 for a survey), resampling techniques such as the bootstrap method (Efron, 1979; Efron and Tibshirani, 1993) are most commonly used.

Two strategies can be designed to find a tree of more appropriate complexity than \hat{t} . The first one, qualified as ascending, consists in successively resolving a tree of null complexity (a “star”) until no new resolution is significant. On the other hand, one may adopt a descending strategy. A tree of maximal complexity (fully resolved) is first inferred and then submitted to an elimination process successively removing its edges (Fig. 1 shows how edges are removed).

The first strategy is computationally very expensive. It requires the examination of an exponential number of alternatives (bipartitions of species) at each step. The second strategy can be implemented more easily. Many algorithms are available for constructing fully resolved trees from a distance matrix (Sattah and Tversky, 1977; Saitou and Nei, 1987; Gascuel, 1997). We are interested in this particular strategy and have investigated two approaches from this standpoint.

The first approach consists in removing from \hat{t} the edges whose reliability is lower than some a priori fixed threshold. The number of removed edges thus depends on the chosen reliability threshold. Among other methods, Felsenstein (1985) introduced a now widespread bootstrap procedure, which assigns a reliability value (a “bootstrap proportion”) to every edge of \hat{t} . We have recently shown (Berry and Gascuel, 1996) how the best reliability threshold can be chosen for this approach, according to some classical error measure.

The second approach consists in first determining the most appropriate complexity for the inferred tree, i.e., the one leading to the best complexity/goodness-of-fit compromise. The edges of the tree \hat{t} that is initially obtained are then sorted according to their importance in the minimization of $\delta(\hat{D}, \hat{t})$. The least effective edges in this sense are then removed so that the tree reaches the determined complexity.

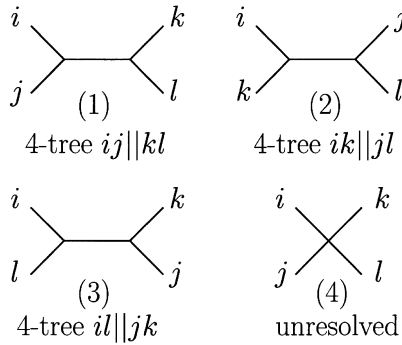


Fig. 2. Possible four-trees and irresolution for the quartet $\{a, b, c, d\}$.

In the following, we first define the distance criterion δ . Then (Section 3) we describe a specific heuristic method to infer trees from distance matrices. Next (Sections 4 and 5), we detail the two descending approaches mentioned above and the resampling techniques on which they both rely. Finally (Section 6), we report some simulation results comparing these two approaches.

2. A structural goodness-of-fit criterion

In this study our only aim is to estimate the structure of a tree, thus we use a criterion which is a direct function of the tree structure. However, the entire approach we present in this paper could be applied to many other phylogenetic criteria, e.g., distance criteria such as the least-squares or the minimum evolution principle.

The structural goodness-of-fit criterion we use is based on quartets of species, i.e., sets of four species. Any dissimilarity D is a tree distance if and only if it satisfies the *four-point condition* (Zaretskii, 1965; Buneman, 1974), also called the additivity condition. This condition requires that for any quartet of species $\{i, j, k, l\}$ the larger two of the three sums $d_{ij} + d_{kl}$, $d_{ik} + d_{jl}$, $d_{il} + d_{jk}$ be equal. If $d_{ij} + d_{kl}$ is the smallest sum, then there must be at least one edge separating i, j from k, l in the tree t representing D . This structural constraint is denoted $ij||kl$ and can be represented by a *four-tree*, as in Fig. 2(1). As can be seen from Fig. 2(1–3), four species can be separated into two pairs in three different ways. We denote these three situations respectively, $ij||kl$, $ik||jl$ and $il||jk$. When the three sums are equal, the quartet is unresolved. The latter situation is represented by a star tree on the four species (Fig. 2(4)).

In order to obtain four-trees from any distance measure \hat{D} , Sattah and Tversky (1977) and Fitch (1981) suggest using the following relaxation to the four-point condition:

$$\hat{d}_{ij} + \hat{d}_{kl} < \begin{cases} \hat{d}_{ik} + \hat{d}_{jl} \\ \hat{d}_{il} + \hat{d}_{jk} \end{cases} \rightarrow ij||kl. \tag{1}$$

The rationale for this is that such a distance pattern is best approximated by the four-tree of Fig. 2(1), i.e., when i, j are separated from k, l . This was demonstrated for several quantitative goodness-of-fit criteria (Saitou and Nei, 1987; Gascuel and Levy, 1996). If none of the three sums in (1) is smaller than the others, the star tree (Fig. 2(4)) is associated with the quartet $\{i, j, k, l\}$.

Any tree structure τ on n species can also be defined by a set of four-trees (Bandelt and Dress, 1986; Steel, 1992). These four-trees are the ones supported by its edges. More precisely, any edge e of τ separates S into two distinct parts, σ and σ' , according to the two subtrees it defines in τ . If $\sigma|_e\sigma'$ denotes the bipartition induced on S by an edge e , then

$$(\exists \sigma|_e\sigma' \in \tau \text{ s.t. } \{i, j\} \subset \sigma \text{ and } \{k, l\} \subset \sigma') \Leftrightarrow ij||kl \quad (2)$$

enables to associate a four-tree set to τ . If the paths (ij) and (kl) in τ intersect in only one node, the star is associated with the quartet $\{i, j, k, l\}$. The tree τ can be considered as structurally equivalent to the four-tree set obtained from (2).

Since any distance matrix and any tree can be associated with a four-tree set, we will use this common concept to define the goodness-of-fit criterion δ . Robinson and Foulds (1981) defined the structural distance between two trees as the cardinal of the symmetrical difference between the sets of bipartitions induced, respectively, by their edges. Similarly, we define $\delta(\hat{D}, \hat{t})$ as the cardinal of the symmetrical difference between \hat{D} and \hat{t} , expressed in terms of their associated four-tree sets, $Q_{\hat{D}}$ and $Q_{\hat{t}}$, obtained from (1) and (2), respectively:

$$\delta_1(\hat{D}, \hat{t}) = |\{q \in Q_{\hat{t}}/q \notin Q_{\hat{D}}\}|,$$

$$\delta_2(\hat{D}, \hat{t}) = |\{q \in Q_{\hat{D}}/q \notin Q_{\hat{t}}\}|,$$

$$\delta(\hat{D}, \hat{t}) = \delta_1(\hat{D}, \hat{t}) + \delta_2(\hat{D}, \hat{t}). \quad (3)$$

Thus δ measures the structural disagreement of \hat{t} and \hat{D} . In fact, $\delta_1(\hat{D}, \hat{t})$ and $\delta_2(\hat{D}, \hat{t})$, can be seen respectively as types I and II error terms. δ can be applied to any distance matrix and therefore possibly to the true distance matrix D . In this case, if $\delta(D, \hat{t})=0$ then $\hat{t} = t$.

3. Constructing trees from quartets

3.1. Obtaining a fully resolved tree

Optimizing (3) is an NP-Hard problem (Steel, 1992). As usual in phylogeny reconstruction, we rely on a heuristic method to infer the tree \hat{t} . The method is based on an agglomerative process (Fig. 3) as often in this context (Sattah and Tversky, 1977; Saitou and Nei, 1987). Starting off with a trivial tree structure, we iteratively cluster subtrees, i.e., sets of species. More precisely, the star tree (Fig. 3(1)), having only one internal node O , is initially considered. At each step two neighbor nodes i and j of O are clustered (Fig. 3(2)). That is, the species included in the subtrees of respective root i and j , are clustered as opposed to the other species of S . This

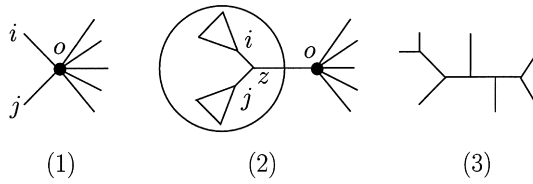


Fig. 3. Obtaining \hat{t} .

clustering results in a new node z connected to O , i and j , and in the removal of edges (Oi) and (Oj) . The edge (Oz) represents the new ramification induced in the tree. The construction process is halted when O has only three neighbors left, i.e., when no further ramification can be introduced between the species of S (Fig. 3(3)).

At each step the cluster which implies the greatest improvement to the adjustment of \hat{D} is chosen. However, for this purpose we use a slightly different criterion than the one defined before. Indeed, due to the heuristic agglomerative approach we chose, minimizing (3) at each step does not guarantee the recovery of the correct tree when processing additive data (examples are available on request). Therefore, to choose among the several possible clusters we use the following heuristic criterion (to be maximized):

$$\delta'(\hat{D}, \hat{t}) = \frac{1 + |Q_{\hat{D}}| - \delta_2(\hat{D}, \hat{t})}{\delta_1(\hat{D}, \hat{t})}. \tag{4}$$

The numerator of this expression is always positive. Whenever \hat{D} is additive and t fully resolved, at each step there is at least one cluster that does not contradict any four-tree of $Q_{\hat{D}}$. Thus $\delta_1 = 0$, implying $\delta' = \infty$. Maximizing (4) implies choosing a cluster at each step which satisfies this property. Finally, we obtain a fully resolved tree which satisfies $\delta_1 = 0$, since no four-tree from $Q_{\hat{D}}$ has been contradicted. That is, the value of the complete tree is infinite for δ' . Since this tree is fully resolved, we also have $\delta_2 = 0$, and, therefore, $\delta = 0$. In other words, the algorithm always retrieves the tree which represents an additive data matrix. Variants can also be designed for dealing with additive distance measures associated with non-fully resolved trees.

When \hat{D} is not additive, maximizing (4) leads to minimizing δ_1 and δ_2 and, therefore, to minimizing (3).

3.2. Obtaining a sequence of trees

We will see later that one of the descending approaches requires the inference of a sequence of decreasingly complex trees $[t_{\max}, \dots, t_0]$. In this expression t_{\max} is the tree \hat{t} inferred by the tree-building method described above and t_0 denotes the star tree. The other trees of the sequence are obtained by successively removing the edges of t_{\max} . Moreover, we aim at producing a tree sequence such that each tree optimizes the goodness-of-fit criterion (3), but still respects its attributed complexity. This implies that the edges of t_{\max} have to be removed in a specific order, not necessarily inverse to the order in which they were obtained since they were heuristically inferred. We rely

on a step by step elimination process. The edges of t_{\max} are reevaluated by criterion (3) and the edge whose elimination least decreases this criterion is removed. We thus obtain the next tree of the sequence. The edge elimination process is repeated until all edges of t_{\max} have been removed, i.e., t_0 has been obtained.

4. Fixed reliability threshold approach

As underlined before, there exist many methods to test the reliability of tree edges (Li and Gouy, 1990; Li and Zharkikh, 1995; Rzhetsky et al., 1995). The most widespread methods rely on resampling techniques, and more precisely on the bootstrap process (Efron, 1979; Efron and Tibshirani, 1993). Felsenstein (1985) proposed to resample the columns of the data matrix X with replacement to produce a new $n \times p$ data matrix X^* with the same number of sites. The goal of this resampling process is to simulate the sampling variance with which the observed sites are obtained. These sites represent only a small part of the species sequence and are assumed to evolve independently.

The same method as for X is used to infer a tree from X^* . We note \hat{D}^* the distance matrix between species of S computed from X^* , and \hat{t}^* the tree obtained from \hat{D}^* . According to Efron et al. (1995) we denote this process

$$X^* \rightarrow \hat{D}^* \rightarrow \hat{t}^*.$$

This whole process is independently repeated a great number of times r . Then for each edge of \hat{t} we compute the frequency (called the *bootstrap proportion*) with which it is inferred in the various trees \hat{t}^* . Each frequency f can be considered as a reliability measure for the corresponding edge. Edges inferred with low frequency will very probably be highly dependent on X and are doubtful to represent some structural aspect of the true tree t . These edges are very likely to result from the overfitting effect previously mentioned.

In introducing the bootstrap technique in the phylogenetic reconstruction field, Felsenstein aimed at evaluating the reliability of the edges of \hat{t} . Several authors discussed the use of the bootstrap proportions in a test strategy. In accordance with the findings of Hillis and Bull (1993), simulations we performed (Berry and Gascuel, 1996) have shown that 95% is a very conservative threshold, giving a method of low power. More sophisticated techniques (Efron et al., 1995; Zharkikh and Li, 1995; Larget and Simon, 1999) have recently been proposed, whose potential is to be tested in a practical context.

All these works consider the edges separately, aiming at determining their individual validity. On the other hand, we adopt here a phylogeny estimation point of view and aim at producing a tree which, *taken as a whole*, estimates the true tree t as best as possible structurally. To this purpose we use Felsenstein's procedure to decide which edges of the tree \hat{t} is to be removed in order to reduce the overfitting effect. More precisely, edges with lowest bootstrap proportions (below some fixed threshold) are removed from \hat{t} . In this estimation context, we have shown theoretically that, in expectation, 50% is the optimal threshold, if the structural distance

between the estimated and the true tree is measured according to the Robinson and Foulds (1981) structural distance measure. This is only so when assuming that f represents the probability that the corresponding edge be correct. As a matter of fact, f was shown by simulations to be close to this probability, especially when $f \approx 50\%$ (Berry and Gascuel, 1996).

Using Felsenstein's bootstrap method as described above enables us to reduce the overfitting effect and to improve the quality of the estimation of t . However, a major defect of this approach is that the threshold at which edges of \hat{t} are removed depends on simplifying hypotheses. In practice, the best value for this threshold depends on many factors, such as the tree-building method and evolutionary conditions. Also, it seems that designing a theoretical expression to fix the threshold according to these factors will be at least problematic.

5. Variable complexity level approach

The second approach we present focuses on the most appropriate complexity level α for the tree \hat{t} . To avoid the previous drawback of choosing (partly) arbitrarily a fixed threshold, we approximate α according to the reconstruction context, i.e., the particular data X and the tree-building algorithm.

For this purpose we rely on resampling techniques such as bootstrap and cross-validation to produce "independent" data pseudo-samples. From each pseudo-sample we obtain a bootstrap tree \hat{t}^* . This tree is then evaluated in the sense of \hat{D} to infer an approximation α^* of the most appropriate threshold for \hat{t} to estimate t . In other words, we draw conclusions for $t - \hat{t}$ on the basis of $\hat{t} - \hat{t}^*$, which is the basic idea of the bootstrap.

First a sequence \hat{s} of decreasingly complex trees is inferred from \hat{D} :

$$X \rightarrow \hat{D} \rightarrow \hat{s} = [\hat{t}_{\max}, \dots, \hat{t}_0]$$

with

$$\hat{t}_{\max} \succ \dots \succ \hat{t}_0,$$

where $\hat{t}_{\max} = \hat{t}$ and $t_i \succ t_j$ means that t_i contains all edges of t_j . The sequence \hat{s} is obtained as described in Section 3.2, i.e., such that its trees are of decreasing complexity and efficiency in adjusting \hat{D} in the sense of (3). This sequence of trees may be compared, to some extent, with the sequence of decision trees inferred in CART (Breiman et al., 1984), even if the purposes are different.

The original data matrix X is resampled, in the same way as previously described, to produce data pseudo-samples X^* . The corresponding distance matrix, \hat{D}^* , is computed each time and serves as a basis for obtaining a sequence \hat{s}^* of trees of decreasing complexity:

$$X^* \rightarrow \hat{D}^* \rightarrow \hat{s}^* = [\hat{t}_{\max}^*, \dots, \hat{t}_0^*],$$

$$\hat{t}_{\max}^* \succ \dots \succ \hat{t}_0^*.$$

The trees of \hat{s}^* are inferred by the same tree-building method as the one used to infer the trees of \hat{s} . Once \hat{s}^* has been obtained, the accuracy of each of its trees to

adjust the initial distance matrix \hat{D} is measured, using (3). The best complexity α^* for inferring a tree from X^* is deduced as being the one associated with the tree of \hat{s}^* which best adjusts \hat{D} , i.e., $\alpha^* = \operatorname{argmin}_{\alpha_i} (\delta(\hat{D}, \hat{t}_{\alpha_i}^*))$.

The whole process leading from X^* to α^* is repeated r' times. Remark that the same number of replications as for the fixed reliability threshold approach is not necessary. $r' = 10$ enabled to obtain sufficiently precise estimations, while $r = 1000$ is usually chosen for the first approach. An explanation is that here we aim at estimating only one value (the complexity α), whereas the previous approach has to estimate as many values as the number of edges of \hat{t} .

The complexity α , used for the tree inferred from \hat{D} , is approximated by the mean of the α^* values. Thus, the tree finally proposed to estimate t is the tree \hat{t}_α , i.e., the tree of the sequence \hat{s} which has α as complexity.

Other resampling techniques can be used with the approach described above. For example, we also investigated the cross-validation technique. Its use necessitates some adjustments to the previous scheme. Each resampling consists in splitting the original data matrix X into two half-sized matrices X_1 and X_2 . Columns of X are randomly attributed either to X_1 or to X_2 . Let \hat{D}_1 and \hat{D}_2 be the two distance matrices computed from X_1 and X_2 , respectively. \hat{D}_1 is used to produce the sequence \hat{s}^* and \hat{D}_2 to evaluate its trees with the δ criterion (3). The complexity α^* corresponds, as before, to the tree of \hat{s}^* showing the best value in this sense. The process leading from X to α^* is repeated r' times and α is deduced as described before. \hat{s} is obtained on the basis of all X and \hat{D} matrices. The tree of \hat{s} having α as complexity level is chosen as the final estimation of t .

6. Discussion

We performed computer simulations to compare the relative efficiencies of the fixed threshold and the variable complexity approaches. For this purpose we simulated the evolution of molecular sequences along the edges of a model tree, according to a classical simulation scheme (Kuhner and Felsenstein, 1994; Berry and Gascuel, 1996) described in detail in Berry (1997).

Results (Berry, 1997) showed that, in the context of the variable complexity approach, the cross validation technique is clearly less efficient than the bootstrap one. This is due to the fact that it only uses half of the data set to construct the trees, and half of the data set to evaluate them. This leads to too conservative an approach. As a result, the cross validation estimated trees were always far less resolved than the other estimates of the true tree.

Both the fixed threshold and the variable complexity approaches calibrated by a bootstrap process led to significant improvements (up to 25% in terms of structural accuracy to the correct tree) over the standard approach which proposes a fully resolved tree \hat{t} . The most important gains were obtained for data sets of poor quality. The fixed threshold approach was always slightly more efficient than the variable complexity one, but the gap between them was at most of 5%, in terms of structural accuracy. On the other hand, the variable complexity approach was much faster

to execute, since only 10 bootstrap replications were needed in order to have precise measures, while 1000 replications are usually required for the fixed threshold approach.

The variable complexity approach seems promising since it allows faster estimations than the bootstrap procedure commonly used in phylogenetic reconstruction and provides almost as accurate estimates. Moreover, it avoids the problem of arbitrary fixing a threshold to remove edges of the original estimate \hat{t} (in the fixed threshold approach, 50% is partly an empirically determined approximation of the optimal threshold).

The variable complexity approach could also be applied to other criteria such as the least-squares or the minimum evolution principles. Moreover, other parameters than the optimal complexity can be estimated by using the same approach. For example, we could aim at estimating the reliability threshold under which edges should be removed. When valued edges are of interest, the same approach could be used to determine the minimal length for edges to be kept in the tree... As can be seen, the variable complexity approach can be used in many ways that deserve further investigations.

References

- Bandelt, H.-J., Dress, A., 1986. Reconstructing the shape of a tree from observed dissimilarity data. *Adv. Appl. Math.* 7, 309–343.
- Barthélemy, J., Guénoche, A., 1991. *Trees and Proximities Representations*. Wiley, Chichester.
- Berry, V., 1997. *Méthodes et algorithmes pour reconstruire les arbres de l'Évolution*. Ph.D. Thesis, Université de Montpellier.
- Berry, V., Gascuel, O., 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. *Mol. Biol. Evol.* 13 (7), 999–1011.
- Breiman, L., Friedman, J., Olshen, A., Stone, C.J., 1984. *Classification and Regression Trees*. Belmont, Wadsworth.
- Buneman, P., 1974. A note on metric properties of trees. *J. Combin. Theory* 17(B), 48–50.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7, 1–26.
- Efron, B., Halloran, E., Holmes, S., 1995. Bootstrap confidence levels for phylogenetic trees. Technical Report Tech. Rep. 179, Stanford University.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, London.
- Felsenstein, J., 1985. Confidence limites on phylogenies: an approach using the bootstrap. *Evolution* 39 (4), 783–791.
- Fitch, W.M., 1981. A non-sequential method for constructing trees and hierarchical classifications. *J. Mol. Evol.* 18, 30–37.
- Gascuel, O., 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14, 685–695.
- Gascuel, O., Levy, D., 1996. A reduction for approximating a (non-metric) dissimilarity by a tree distance. *J. Classification* 13, 129–155.
- Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systems Biol.* 42, 192–200.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), *Mammalian Protein Metabolism*, Vol. III. Academic Press, New York, pp. 21–132 (Chapter 24).
- Kimura, M., 1980. A simple method for estimating evolutionary rates base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.

- Kuhner, M., Felsenstein, J., 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11, 459–468.
- Larget, B., Simon, D.L., 1999. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16 (6), 750–759.
- Li, W.-H., Gouy, M., 1990. Statistical tests of molecular phylogenies. In: Doolittle, R.F. (Ed.), *Molecular Evolution: Computer Analysis of Protein and Nucleic acid Sequences, Methods In Enzymology*, Vol. 183. Academic Press, New York, pp. 645–659 (Chapter 40).
- Li, W.-H., Zharkikh, A., 1995. Statistical tests of DNA phylogenies. *Systems Biol.* 44, 49–63.
- Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147.
- Rzhetsky, A., Kumar, S., Nei, M., 1995. Four-cluster analysis: a simple method to test phylogenetic hypotheses. *Mol. Biol. Evol.* 12 (1), 163–167.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstruction phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406–425.
- Sattah, S., Tversky, A., 1977. Additive similarity trees. *Psychometrika* 42, 319–345.
- Steel, M., 1992. The complexity of reconstructing trees from qualitative characters and subtrees. *J. Classification* 9, 91–116.
- Swofford, D.L., Olsen, G.J., 1990. Phylogeny reconstruction. In: Hillis, D.M., Moritz, C. (Eds.), *Molecular Systematics*. Sinauer Associates, Inc., Sunderland, MA, USA, pp. 411–501 (Chapter 11).
- Zaretskii, K., 1965. Construction of a tree on the basis of a set of distances between its leaves. *Upekki Math. Nauk.* 20, 90–92 (in Russian).
- Zharkikh, A., Li, W.-H., 1995. Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol.* 4 (1), 44–63.