The Minimum Evolution Problem: Overview and Classification

Daniele Catanzaro
Graphes et Optimisation Mathématique (G.O.M.), Département d'Informatique, Université Libre de Bruxelles (U.L.B.), Boulevard du Triomphe, CP 210/01, Brussels B-1050, Belgium

Molecular phylogenetics studies the hierarchical evolutionary relationships among organisms by means of molecular data. These relationships are typically described by means of weighted trees, or phylogenies, whose leaves represent the observed organisms, internal vertices the intermediate ancestors, and edges the evolutionary relationships between pairs of organisms. Molecular phylogenetics provides several criteria for selecting one phylogeny from among plausible alternatives. Usually, such criteria can be expressed in terms of objective functions, and the phylogenies that optimize them are referred to as optimal. One of the most important criteria is the minimum evolution (ME) criterion, which states that the optimal phylogeny for a given set of organisms is the one whose sum of edge weights is minimal. Finding the phylogeny that satisfies the ME criterion involves solving an optimization problem, called the minimum evolution problem (MEP), which is notoriously $\mathcal{NP}$-Hard. This article offers an overview of the MEP and discusses the different versions of it that occur in the literature. © 2008 Wiley Periodicals, Inc. NETWORKS, Vol. 53(2), 112–125 2009

Keywords: network design; computational biology; phylogenetic estimation; minimum evolution

1. INTRODUCTION

One of the most important aims in systematics, the science that studies the diversity of life on Earth, is to estimate the evolutionary relationships of a given set of organisms (usually referred to as *taxa*). These relationships are described by means of a weighted tree (called a *phylogeny*, see Fig. 1) whose leaves represent taxa, internal vertices the intermediate ancestors, edges the evolutionary relationships between pairs of organisms, and edge weights the *evolutionary distances* (i.e., measures of the dissimilarity) between pairs of organisms [36].

Received March 2007; accepted February 2008
Correspondence to: D. Catanzaro; e-mail: dacatanz@ulb.ac.be
Contract grant sponsor: Belgian National Fund for Scientific Research (F.N.R.S.)
DOI 10.1002/net.20280
Published online 22 October 2008 in Wiley InterScience (www.interscience.wiley.com).

Phylogenetic estimation has been practiced since Darwin [44]. Initially, physical characters of taxa, such as morphology or physiology, were used to estimate the corresponding phylogeny [89]. Nowadays, phylogenetic estimation can also be carried out through the use of molecular data extracted from taxa, such as protein fragments, DNA and RNA sequences, or (less frequently) the whole genome [89].

Since no one can practicably observe the real evolutionary process over thousands or millions of years, there is no way to empirically validate a candidate phylogeny for a set of taxa [44]. For this reason, different researchers have proposed various criteria for selecting one phylogeny from among plausible alternatives [36]. These criteria can usually be expressed in terms of objective functions, and the phylogenies that optimize them are referred to as *optimal* [44]. Each criterion adopts a set of assumptions whose ability to describe the real evolutionary process determines the gap between the real and the *true phylogeny*, i.e., the phylogeny that one would obtain under the same set of assumptions if all the molecular data from the (set of) taxa were available [36]. If the optimal phylogeny approaches the true phylogeny as the amount of molecular data analyzed increases, then the corresponding criterion is said to be *statistically consistent* [44].

The first criterion to be proposed was the *parsimony criterion* [12]. This criterion states that under many plausible explanations of an observed phenomenon, the one requiring the fewest assumptions should be preferred [89]. Hence, under the parsimony criterion, a phylogeny is defined to be optimal (or the most parsimonious) if the sum of edge weights of each path (see Section 2) from one taxon to another in the phylogeny is minimal [50]. However, finding the most parsimonious phylogeny for a set of taxa is $\mathcal{NP}$-Hard [50]; a further drawback is that, in some circumstances (see [36], p. 113), this criterion may be statistically inconsistent [32, 81].

Two alternative families of phylogenetic estimation criteria [44] were then proposed: the likelihood criterion [33] and the distance-based criterion [36].

The likelihood criterion states that under many plausible explanations of an observed phenomenon, the one with the highest probability of occurring should be preferred [33]. Hence, under the likelihood criterion, a phylogeny is

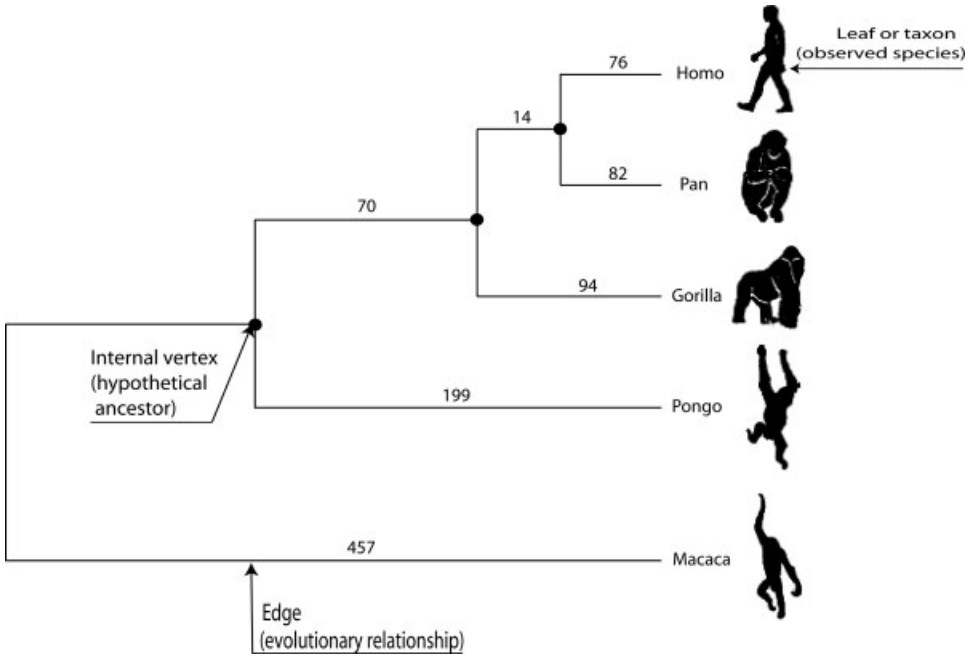

FIG. 1. An example of a phylogeny.

defined to be optimal (or the most likely) if it has the highest probability of explaining the observed taxa [33]. In contrast to the parsimony criterion, the likelihood criterion has the important benefit of being statistically consistent ([44], p. 50). Again, however, the problem of finding the most likely phylogeny is $\mathcal{NP}$-Hard [17, 82]. A recent extension of the likelihood criterion, *Bayesian inference* [27, 58, 68], uses prior and posterior probabilities to measure the quality of the phylogeny provided ([44], p. 67-69). Just as with the likelihood criterion, however, finding the optimal phylogeny using Bayesian inference is $\mathcal{NP}$-Hard [36].

The distance-based criterion aims to find a phylogeny that best fits a given matrix of evolutionary distances among pairwise molecular data [36]. Different definitions of "fitting" give rise to different distance-based criteria [36]. One of the most important distance-based criteria is that of minimum evolution (ME). This criterion states that the optimal phylogeny for a set of taxa is the one whose sum of edge weights, estimated from the corresponding evolutionary distances, is minimal [43, 85]. Phylogenies satisfying the ME criterion are determined by solving a minimum evolution problem (MEP); versions of this problem depend on how the edge weight estimation is performed. The ME criterion has the benefit of generally being statistically consistent. It also requires little computational effort to estimate the edge weights of a phylogeny compared to other statistically consistent criteria such as the likelihood criterion [33] or the Bayesian inference criterion [27, 58, 68]. On the other hand, finding the phylogeny that satisfies the ME criterion is generally $\mathcal{NP}$-Hard [22].

Here, we provide a review of the available literature on the MEP. In Section 2 we state the ME criterion and the corresponding MEP in its most general form. In Section 3 we propose a possible taxonomy of the different versions of the MEP; specifically, we classify these versions according to the type of objective function, constraints, and methods used to estimate edge weights. We then present a synopsis of the MEP in Section 4. In Section 5 we review the most widely used approaches to solving the various versions of the MEP; and finally, in Section 6, we discuss the statistical consistency of the phylogenies they provide.

2. THE MINIMUM EVOLUTION PROBLEM

In this section, we formally state the minimum evolution criterion and the corresponding minimum evolution problem in its most general form. To this end, we first introduce some preliminary definitions that will prove useful throughout the article.

Denote Γ as the set of n organisms (taxa) to be analyzed, and consider an unweighted graph $G = (V, \mathcal{E})$ (namely, a *phylogenetic graph*), where $V = V_e \cup V_i$ is the set of vertices. V_e is the set of n *leaves* representing the n taxa in Γ, and V_i the set of $(n-2)$ *internal vertices* representing the common ancestors. By analogy, $\mathcal{E} = \mathcal{E}_e \cup \mathcal{E}_i$ is the set of $\frac{3}{2}(n-1)(n-2)$ edges, where $\mathcal{E}_e$ is the set of *external edges*, i.e., the set of edges with one extreme being a leaf, and $\mathcal{E}_i$ is the set of *internal edges*, i.e., the set of edges with both extremes being internal vertices. Then a *phylogeny* of the set Γ is any spanning tree T of G such that each internal vertex has degree three, and each leaf has degree one.

It is worth noting that there is no biological reason for imposing the degree constraint on the internal vertices of a phylogeny. Nevertheless, the constraint is usually imposed as it simplifies the formalization of the MEP [94]. Moreover, the constraint is not an oversimplification because any m-ary tree can be transformed into a phylogeny by adding "dummy" vertices and edges (e.g., see Fig. 2 and [94]).

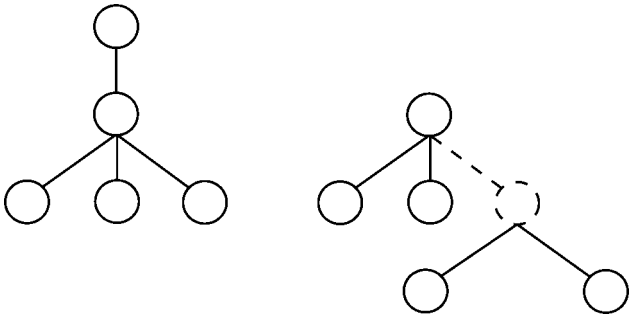

FIG. 2. The 4-ary tree (on the left) can be transformed into a phylogeny by adding a dummy vertex and edge (dashed, on the right).

We denote $\mathcal{E}(T)$ as the set of edges of a phylogeny T, $\mathcal{T}$ as the set of all the possible $(2n-5)!!$ phylogenies of Γ (where $n!!$ is the double factorial of n) [36], and we assume that a weight function $f : \mathcal{E}(T) \to \Re$ is given. We denote $\mathbf{w}$ as the $(2n-3)$-vector of edge weights associated to T, and let $L(T)$ be the length of T, i.e., the sum of the associated edge weights. We define a phylogeny T with weights $\mathbf{w}$ as a *tree metric* if all the entries of $\mathbf{w}$ are nonnegative [99].

We assume that a $n \times n$ distance matrix $\mathbf{D} = \{d_{ij}\}$ of *evolutionary distances* [11] between each pair of taxa i and j in Γ is given *a priori*. Such evolutionary distances measure the dissimilarity between pairwise molecular data, and are usually computed on the basis of a given Markov substitution model of molecular evolution (e.g., those described in [11, 36, 52, 60, 62, 67, 83, 98]) or, more rarely, by means of metric models (e.g., those described in [5, 61]). We say that a distance matrix $\mathbf{D}$ is a *dissimilarity matrix* if for each pair of distinct taxa i and j, $d_{ij} > 0$, $d_{ij} = d_{ji}$, and $d_{ii} = 0$ [99]. In addition, we say that a dissimilarity matrix $\mathbf{D}$ is *metric* if the triangle inequality also holds [34]:

$$d_{ij} \leq d_{ik} + d_{kj} \quad \forall i, j, k \in \Gamma. \tag{1}$$

Metric distance matrices are more likely to be generated when, for example, covarion models [37, 39, 57, 70] of molecular evolution (see [25]), or the models described in [5, 61] are used.

We say that a dissimilarity matrix $\mathbf{D}$ is *additive* if there exists a tree metric phylogeny such that the sum of edge weights along the path between leaves i and j is equal to d_{ij}, for all $i, j \in \Gamma$ [99] or, equivalently (see [10, 26]), if the *four-point condition* ([89], p. 146) holds, i.e., for any four taxa i, j, k, and z

$$d_{zi} + d_{kj} \leq d_{zj} + d_{ik} = d_{kz} + d_{ij}. \tag{2}$$

In fact, by referring, for example, to the phylogeny shown in Figure 3, if a phylogeny is a tree metric then the following inequality holds:

$$\begin{aligned} d_{AB} + d_{CD} &= e_A + e_B + e_C + e_D \\ &\leq d_{AD} + d_{BC} = e_A + e_B + e_C + e_D + 2e_1 \end{aligned}$$

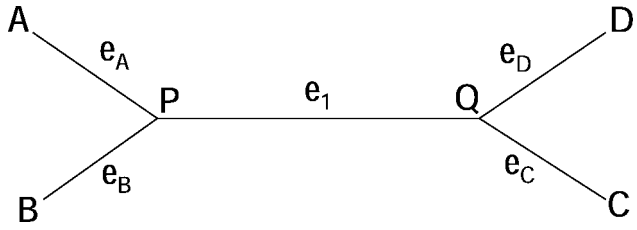

FIG. 3. An example of a phylogeny with four leaves.

(because e_1 by hypothesis is a nonnegative quantity). On the other hand, since $d_{AD} = e_A + e_1 + e_D$, $d_{BC} = e_B + e_1 + e_C$, $d_{AC} = e_A + e_1 + e_C$, and $d_{BD} = e_B + e_1 + e_D$ then the relation $d_{AD} + d_{BC} = d_{AC} + d_{BD}$ holds, and the four-point condition follows.

We say that an additive distance matrix $\mathbf{D}$ is *ultrametric* if the following inequality holds for any triplet of taxa i, j, k [29]:

$$d_{ij} \leq \max\{d_{ik}, d_{kj}\}. \tag{3}$$

Additive distance matrices, respectively ultrametric distance matrices, are generated when, for example, the Markov substitution model of molecular evolution described in [51], respectively the model described in [29], is used. As observed by Farach et al. [29], ultrametric distance matrices are highly desirable in biology because evolutionary distances, as measured in time, satisfy (3).

In accordance with the literature [36], we represent a phylogeny by means of a particular network matrix called an edge–path incidence matrix of a tree (EPT) ([76], p. 550). An EPT matrix $\mathbf{X}$ of a phylogeny T is characterized by having a row for each path between two leaves and a column for each edge. The generic entry $x_{ij,e}$ is then equal to 1 if edge e belongs to the path p_{ij} from leaf i to leaf j and is 0 otherwise. As an example, Table 1 shows the EPT matrix corresponding to the phylogeny shown in Figure 3.

The one-to-one correspondence between a phylogeny T and the associated EPT matrix $\mathbf{X}$ induces a bijection between the set of all the possible phylogenies $\mathcal{T}$ and the set of all the possible associated EPT matrices $\mathcal{X}$. Hence, given a distance matrix $\mathbf{D}$, the problem of determining a phylogeny that satisfies the ME criterion can be formalized, in its most general form, as follows:

TABLE 1. The EPT matrix corresponding to the phylogeny shown in Figure 3.

	e_A	e_B	e_C	e_D	e_1
AB	1	1	0	0	0
AC	1	0	1	0	1
AD	1	0	0	1	1
BC	0	1	1	0	1
BD	0	1	0	1	1
CD	0	0	1	1	0

TABLE 2. References classified by perspective.

Perspective	References
Version	[4, 5, 8, 9, 13, 14, 24, 25, 29, 38, 42, 44–46, 53] [55, 61, 65, 72, 74, 79, 84, 85, 90, 94, 96, 97, 99] [1, 7, 15, 16, 18–21, 25, 28, 29, 31, 41–43]
Approach to solution	[44, 48, 59, 66, 71, 80, 84, 85, 87, 88, 91–93, 95] [99, 101, 102]
Statistical consistency	[2, 23, 25, 45, 47, 54, 56, 61, 69, 75, 85, 100]

Minimum Evolution Problem

$$\min_{(\mathbf{X},\mathbf{w})} L(\mathbf{X},\mathbf{w})$$
$$s.t. \quad f(\mathbf{D},\mathbf{X},\mathbf{w}) = 0$$
$$\mathbf{X} \in \mathcal{X}, \mathbf{w} \in \Re_{0^+}^{(2n-3)}$$

where $L(\mathbf{X},\mathbf{w})$ indicates the length of a phylogeny $\mathbf{X}$ with associated edge weights $\mathbf{w}$, and $f(\mathbf{D},\mathbf{X},\mathbf{w})$ is a function correlating the distance matrix $\mathbf{D}$ with the phylogeny $\mathbf{X}$ and edge weights $\mathbf{w}$. Thus, any version of MEP is completely characterized by defining the functions $L(\mathbf{X},\mathbf{w})$ and $f(\mathbf{D},\mathbf{X},\mathbf{w})$.

Apart from some polynomial cases [29, 99], each version of MEP has been proved to be $\mathcal{NP}$-Hard [22, 29, 63, 71, 74, 101]. In the next three sections, we propose a possible taxonomy of these versions and discuss the main approaches proposed to solve them.

3. A POSSIBLE TAXONOMY OF THE LITERATURE

We classify the literature on MEP according to three main perspectives: version, the approach used to solve it, and the statistical consistency of the phylogeny provided. Here, we briefly introduce each perspective and we list in Table 2 the corresponding references.

From the first perspective, we classify the literature based on the type of function $f(\mathbf{D},\mathbf{X},\mathbf{w})$ and the structure of objective function $L(\mathbf{X},\mathbf{w})$ used. The function $f(\mathbf{D},\mathbf{X},\mathbf{w})$ imposes conditions on the differences (incongruences) between the evolutionary distances derived using the weights $\mathbf{w}$ of phylogeny $\mathbf{X}$ and the distances defined by matrix $\mathbf{D}$. Some versions in the literature, the so-called *least-squares models*, require that the sum of the squares of these differences be minimized. Other versions, typically based on *linear programming*, require that the sum of only these differences be minimized, and, further, that entries of $\mathbf{w}$ be nonnegative and satisfy the triangle inequality (1). In turn, the least-squares models can be further differentiated based on the presence (or absence) of the *positivity constraint*, i.e., the nonnegativity of edge weights. The positivity constraint has important biological implications; we discuss these, together with the versions of MEP, in Section 4.

Second, we classify the literature based on the approach used to solve the various versions of MEP. In general, the approaches used are either exact or nonexact. The former includes algorithms based on exhaustive enumeration and on branch-and-bound. The latter can be further divided into approximation algorithms and heuristics. In turn, the approaches based on heuristics can be subdivided into constructive, clustering, and constructive/clustering. We discuss this classification in Section 5.

Finally, we classify the literature based on the statistical consistency of the phylogenies provided by different versions of MEP. In Section 6, we show that some versions of MEP may lead to statistically inconsistent results; for this reason, they are generally frowned upon by the scientific community [23].

4. VERSIONS OF THE MINIMUM EVOLUTION PROBLEM

The MEP can be divided into two subproblems: (i) determining the structure of the optimal phylogeny (i.e., the entries of matrix $\mathbf{X}$), and (ii) finding the associated optimal edge weights (i.e., the entries of $\mathbf{w}$) that best fit the distance matrix $\mathbf{D}$. Historically, the latter subproblem was among the first aspects of molecular phylogenetics to be studied, and is at the core of MEP. In fact, the edge weight estimation subproblem influences directly the choice of the type of function $f(\mathbf{D},\mathbf{X},\mathbf{w})$ and of the structure of function $L(\mathbf{X},\mathbf{w})$, and hence the version of MEP.

The literature proposes two main families of edge weight estimation models (whose references are listed in Table 3): the least-squares models and the linear programming models. The former are discussed in Section 4.1 and the latter are discussed in Section 4.2. Throughout this section, if not explicitly stated, we will always intend that phylogeny $\mathbf{X}$ is assigned.

4.1. The Least-Squares Models

The least-squares models were first introduced by Cavalli-Sforza and Edwards in 1967 [13]. The authors considered each evolutionary distance d_{ij} among pairwise molecular data as uniformly distributed independent random variables satisfying the additive property (2). In other words, Cavalli-Sforza and Edwards assumed that each entry d_{ij} could be thought of as the resulting sum of mutation events w_e accumulated on

TABLE 3. References classified by type of edge weight estimation model.

Edge weight estimation model	References
Least-Squares	
OLS	[8, 13, 42, 45, 61, 84, 85, 96, 97]
WLS	[38, 72]
GLS	[9, 14, 53]
BLS	[24, 25, 44, 79, 90]
Least-squares with positivity constraint	[4, 29, 35, 46, 55, 74]
Linear programming	[5, 99]

each edge e belonging to the path p_{ij} linking taxa i and j on $\mathbf{X}$, i.e., in matrix form:

$$\mathbf{X}\mathbf{w} = \mathbf{D}^{\triangle} \tag{4}$$

where $\mathbf{D}^{\triangle}$ is the $n(n-1)/2$ vector whose components are obtained by taking row by row the entries of the strictly upper triangular matrix of $\mathbf{D}$. In general, Equation (4) may not admit solutions. For this reason, the authors proposed the use of the ordinary least-squares (OLS) to find the entries of vector $\mathbf{w}$. Specifically, the authors suggested that the values $\rho_{ij} = \sum_{e \in p_{ij}} x_{ij,e} w_e$, called *expected distance estimates* [36], should minimize the function:

$$\sum_{i=1}^{n} \sum_{j=1:j\neq i}^{n} (d_{ij} - \rho_{ij})^2 = \sum_{i=1}^{n} \sum_{j=1:j\neq i}^{n} \left(d_{ij} - \sum_{e \in p_{ij}} x_{ij,e} w_e \right)^2 .$$

Some authors disagreed with Cavalli-Sforza and Edwards' model. Specifically, Fitch and Margoliash [38] observed that, due to the common evolutionary history of the taxa analyzed and the presence of sampling errors in molecular data, the assumption that the evolutionary distances $\{d_{ij}\}$ are uniformly distributed independent random variables cannot be considered generally true. Therefore, the authors proposed to modify Cavalli-Sforza and Edwards' model by introducing the quantities $\{\omega_{ij}\}$ representing the variances of $\{d_{ij}\}$. Fitch and Margoliash called the new model weighted least-squares (WLS) and proposed to minimize the function:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{ij} \left(d_{ij} - \sum_{e \in p_{ij}} x_{ij,e} w_e \right)^2 .$$

Fitch and Margoliash [38] proposed to set $\omega_{ij} = 1/d_{ij}^2$, whereas with analogous arguments, Beyer et al. [5] set $\omega_{ij} = 1/d_{ij}$. Under WLS, the function $f(\mathbf{D}, \mathbf{X}, \mathbf{w})$ of MEP becomes:

$$(\mathbf{X}^t \mathbf{\Omega} \mathbf{X})\mathbf{w} = \mathbf{X}^t \mathbf{\Omega} \mathbf{D}^{\triangle} \tag{5}$$

where $\mathbf{\Omega}$ is a diagonal matrix with diagonal entries equal to $\{\omega_{ij}\}$.

Subsequently, Bulmer [9], Chakraborty [14], and Hasegawa et al. [53] noted that the weights $\{\omega_{ij}\}$ account for the variance of $\{d_{ij}\}$, but not for their dependencies. Consequently, they proposed to substitute the values $\{\omega_{ij}\}$ with the covariances of $\{d_{ij}\}$, and called the new model generalized least-squares (GLS). Specifically, Chakraborty [14] modeled molecular evolution as a Poisson process in which mutations are random events occurring along each path in the phylogeny, and derived the covariances of the evolutionary distances by considering path-per-path the number of mutation events observed between pairs of molecular data. A very similar approach was used by Bulmer [9] and Hasegawa et al. [53]: the former used an approximation of the Poisson process to compute the covariances of the evolutionary distances, whereas the latter used a Markov model [52]. Under GLS, the function $f(\mathbf{D}, \mathbf{X}, \mathbf{w})$ of MEP becomes:

$$(\mathbf{X}^t \mathbf{\Psi}^{-1} \mathbf{X})\mathbf{w} = \mathbf{X}^t \mathbf{\Psi}^{-1} \mathbf{D}^{\triangle} \tag{6}$$

where $\mathbf{\Psi}$ is the covariance matrix of the evolutionary distances.

The computational complexity required to solve by means of matrix formulae the above models (respectively $O(n^4)$ for the OLS and WLS models, and $O(n^6)$ for the GLS model [8]) represented in the 1970s and 1980s a serious bottleneck for their empirical application. For this reason, several authors investigated alternative strategies to reduce the computational effort required to implement them.

Vach [96] noted that the bipartition (also called *split* [3]) induced by any edge of a phylogeny can be used to approximate the OLS model. Specifically, given a phylogeny $\mathbf{X}$ and assuming the OLS edge weight estimation model, Vach proved that: (i) the value of an edge weight w_e is a function of the average distance between the leaves belonging to a bipartition induced by edge e, and (ii) such a value does not depend on the phylogeny but only on the leaves contained in the bipartition [97].

This result was reached independently by Rzhetsky and Nei [85] who provided an $O(n^3)$ algorithm to solve the OLS model [86]. This algorithm was further improved by Gascuel [42] who decreased its order of complexity to $O(n^2)$. Finally, Bryant and Waddell in [8] proposed a unified and generalized framework to speed up the solution of the OLS, WLS, and GLS models. Specifically, the authors provided an optimal algorithm to solve the OLS model, an $O(n^3)$ algorithm to solve the WLS model, and an $O(n^4)$ algorithm to solve the GLS model.

Finally, Makarenkov and Lapointe [72] have recently introduced a particular WLS model usable in all cases in which some evolutionary distances are partially given or uncertain (cases usually met, for example, when dealing with fossil data [72]). The model assumes that properties (2–3) hold for the distance matrix $\mathbf{D}$, and assigns $\{\omega_{ij}\} \in \{0, 1/2, 1\}$ as a function of the uncertainty degree of the entries $\{d_{ij}\}$. The authors proved that solving this particular version of MEP is $\mathcal{NP}$-Hard.

4.1.1. The Positivity Constraint. The additive property of the distance matrix $\mathbf{D}$ in Cavalli-Sforza and Edwards' model guarantees that the phylogeny provided by (4), (5), and (6) is a tree metric [29], i.e., implicitly imposes the constraint $\mathbf{w} \geq 0$. Unfortunately, when the distance matrix $\mathbf{D}$ is generic (e.g., it is obtained by means of Markov models, see Section 2), all the least-squares models considered so far may lead to the occurrence of negative entries in the vector $\mathbf{w}$, i.e., to a phylogeny that is not tree metric [44, 65]. Negative edge weights are infeasible both from a conceptual point of view (a distance, being an expected number of mutation events over time, cannot be negative [61]) and from a biological point of view (evolution cannot proceed backwards [77, 94]). For the latter reason at least, nontree metric phylogenies are generally unacceptable to biologists [45].

In response, some authors investigated the consequences of adding or guaranteeing the positivity constraint in the least-squares models. Gascuel and Levy [46] observed that the presence of the positivity constraint transforms ([6], p. 187)

any least-square model into a nonnegative linear regression problem which involves projecting the distance matrix **D** onto the positive cone defined by the set of tree metrics associated to a given phylogeny **X** [45]. Therefore, the authors proposed an algorithm to generate a sequence of least-squares projections of the distance matrix **D** onto the convex set of the tree metrics until an additive distance matrix (and the corresponding phylogeny) is obtained [46]. A similar approach had previously been provided by Hubert and Arabie [55]. Both algorithms are characterized by a computational complexity of $O(n^4)$.

Farach et al. [29] proposed a number of models to perturb a distance matrix **D** to achieve additive or ultrametric matrices. Specifically, the authors proposed a first model in which, given an upper and lower bound for the evolutionary distances, an additive (ultrametric) distance matrix between these two bounds must be found. In a second model, the authors assumed a distance matrix **D** in which some evolutionary distances are partially given or uncertain, and studied the possibility of assigning these entries to obtain a new distance matrix that satisfies the additive (ultrametric) property. They proposed the $\mathcal{L}_\infty$-norm and $\mathcal{L}_1$-norm to constrain the entries of **D** to satisfy the additive (ultrametric) property. The authors proved that both models can be solved in $O(n^2 + n \log n)$ time when an ultrametric distance matrix is required under the $\mathcal{L}_\infty$-norm. By contrast, the authors proved that both models become hard when an ultrametric or an additive distance matrix is required under the $\mathcal{L}_1$-norm. However, to the best of our knowledge, nothing is known about the hardness of finding an additive distance matrix under the $\mathcal{L}_\infty$-norm in either model.

Barthélemy and Guénoche [4] and Makarenkov and Leclerc [74] proposed two iterative $O(n^4)$ and $O(n^5)$ algorithms which exploit the Lagrangian relaxations of the OLS and WLS models to find a phylogeny that satisfies the positivity constraint. Specifically, starting from a leaf, the algorithms generate iteratively a phylogeny with a growing number of leaves by solving an optimization problem which finds the best non-negative edge weights that minimize the OLS (respectively WLS) model. A previous and similar algorithm, called FITCH, was also proposed by Felsenstein [35].

A different approach from those described earlier is followed in the balanced least-squares (BLS) edge weight estimation model [24, 25]. The BLS model is based on a seminal work of Pauplin [79] in which the author modifies Equation (4) by requiring that all edges of a phylogeny **X** be weighted in the same way (see Section 4.1.2). As a result, the new model allows the positivity constraint to be satisfied if the triangle inequality holds for the distance matrix **D** [25]. Desper and Gascuel proved that the BLS model is a special form of the WLS model in which the variances of the evolutionary distances are proportional to their topological distance (i.e., the number of edges belonging to the path between the corresponding endpoint taxa), and inversely proportional to the length of the molecular data [25]. Solving a BLS model requires a computational complexity of $O(n^2 + n \log n)$.

4.1.2. The Objective Function in the Least-Squares Models. Under the least-squares edge weight estimation model, several objective functions have been proposed in the literature. Later, we just give an overview of them, postponing to Section 6 our discussion of the impact that each objective function has on the statistical consistency of MEP.

Rzhetsky and Nei showed that if the entries of the distance matrix **D** were not subjected to sampling errors and all the molecular data from the (set of) taxa were available, then the total length of the true phylogeny must be the shortest [84]. Therefore, the authors suggested the use of the following objective function:

$$L(\mathbf{X}, \mathbf{w}) = \sum_{e=1}^{2n-3} w_e. \tag{7}$$

Some authors proposed modifications to (7) to deal with a non-perfectly additive distance matrix **D**. Specifically, Kidd and Sgaramella-Zonta [61] suggested the use of the following objective function:

$$L(\mathbf{X}, \mathbf{w}) = \sum_{e=1}^{2n-3} |w_e|,$$

whereas Swofford et al. [94] and Gascuel et al. [45] proposed:

$$L(\mathbf{X}, \mathbf{w}) = \sum_{e=1}^{2_n-3} \max\,(w_{e,o}).$$

Specifically, this last function is used in a well-known phylogenetic software package called PAUP 4.0* [93]. Farach et al. [29] and Argawala et al. [1] proposed the function:

$$L(\mathbf{X}, \mathbf{w}) = \|\mathbf{w}\|_\infty = \|\mathbf{X}^\dagger \mathbf{D}^\triangle\|_\infty = \max_{e=1,\ldots,2n-3} |w_e|,$$

where $\|\cdot\|_\infty$ is the $\mathcal{L}_\infty$-vector norm and $\mathbf{X}^\dagger$ is the Moore-Penrose generalized inverse of **X**. This function is particularly related to the statistical consistency of MEP and will be discussed in Section 6.

Under the BLS model, the following objective function was proposed [44]:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2^{1-\tau_{ij}} d_{ij} \tag{8}$$

where

$$\tau_{ij} = \sum_{e=1}^{2n-3} x_{ij,e}.$$

This particular form derives from the seminal works of Korostensky and Gonnet [49, 64] and Pauplin [79]. In the following we will give an intuition behind Equation (8) by referring to the phylogeny shown in Figure 3.

Korostensky and Gonnet [49, 64] introduced the concept of a *circular order* induced by a phylogeny **X**,

i.e., any tour through $\mathbf{X}$ where each edge is traversed exactly twice, and each leaf is visited once. As an example, a possible circular order of the phylogeny shown in Figure 3 is $\{e_A, e_1, e_D, e_D, e_C, e_C, e_1, e_B, e_B, e_A\}$. Korostensky and Gonnet observed the existence of a link between the OLS model and the set of circular orders induced by a phylogeny, but did not investigate the issue any further.

Almost simultaneously, Pauplin [79] proposed a modification to the OLS estimation method. Specifically, for any pair of adjacent vertices x and y of a phylogeny, Pauplin defined the *average distance* between x and y as:

$$\hat{d}_{xy} = \frac{1}{|\Gamma(x)||\Gamma(y)|} \sum_{i \in \Gamma(x), j \in \Gamma(y)} d_{ij}$$

where $\Gamma(x)$, respectively $\Gamma(y)$, indicates the set of leaves of the subtree of $\mathbf{X}$ having x, respectively y, as its root and is such that $\Gamma(x) \cap \Gamma(y) = \emptyset$. As an example, if $x = P$ and $y = Q$ in Figure 3 then $\Gamma(x) = \{A, B\}$ and $\Gamma(y) = \{C, D\}$.

Pauplin then observed that, under the OLS estimation, each weight of an internal edge of a phylogeny (e.g., the w_{e_1} in Fig. 3) can be computed through the formula:

$$\frac{1}{2}[\lambda(\hat{d}_{AD} + \hat{d}_{BC}) + (1 - \lambda)(\hat{d}_{AC} + \hat{d}_{BD}) - (\hat{d}_{AB} + \hat{d}_{CD})]$$

where in principle A, B, C, and D can be either leaves or internal vertices, and where

$$\lambda = \frac{|\Gamma(A)||\Gamma(C)| + |\Gamma(B)||\Gamma(D)|}{|\Gamma(A) \cup \Gamma(B)||\Gamma(C) \cup \Gamma(D)|}.$$

Similarly, each weight of an external edge having z as its terminal leaf and x as its internal vertex can be computed through the formula:

$$\frac{1}{2}[\hat{d}_{zx} + \hat{d}_{xy} - \hat{d}_{zy}]$$

where y is any vertex adjacent to x and different from z. In our example, if $z = A$, $x = P$, and $y = B$, then $e_A = \frac{1}{2}[\hat{d}_{AP} + \hat{d}_{BP} - \hat{d}_{AB}]$. Then, the author proposed a new edge weight estimation model in which the value of λ in (9) is independent of the cardinalities involved, and fixed to 1/2. Pauplin proved that under this condition the length of the resulting phylogeny is (8).

The link between Korostensky and Gonnet's intuition and Pauplin's new estimation method was recently noted by Semple and Steel [90]. The authors proved that, if $C(\mathbf{X})$ is the set of circular orders associated to a given phylogeny $\mathbf{X}$, and with $C(\mathbf{X})_i$ the i-th circular order, the length of the phylogeny $\mathbf{X}$ computed through (8) is exactly the average length of the circular orders induced by $\mathbf{X}$ i.e., $\frac{1}{|C(\mathbf{X})|} \sum_{i \in C(\mathbf{X})} length(C(\mathbf{X})_i)$ [73, 90].

4.2. The Linear Programming Models

Linear Programming (LP) models are currently the only alternatives to the least-squares models. The earliest article using a LP model was proposed by Beyer et al. [5]. The authors observed that if the evolutionary distances between pairs of molecular data have to reflect the number of mutation events required over time to convert one molecular sequence into another, then the evolutionary distances must satisfy the triangle inequality (1) [31]. Moreover, since any edge weight of a phylogeny is *de facto* an evolutionary distance, also the entries of vector $\mathbf{w}$ must satisfy the triangle inequality [5]. Hence, Beyer et al. proposed a LP model in which, for a given phylogeny, $(2n - 3)$ nonnegative edge weights must satisfy $n(n-1)/2$ triangle inequalities.

An analogous model was proposed by Waterman et al. [99], where the authors studied the case in which the additive property (2) also holds for the distance matrix $\mathbf{D}$. Waterman et al. mixed the additive property of the distance matrix with the triangle inequality and modified Beyer et al.'s model by imposing that:

$$w_e \geq 0 \quad e = 1, \ldots, 2n-3$$
$$\sum_{e \in p_{ij}} w_e \geq d_{ij} \quad \forall i < j, \quad i, j \in \Gamma$$

The authors were able to prove that the assignment of edge weights by LP yields at least $(n-1)$ of the 2^n triangle inequalities as equalities, and at most $(n-1)$ edge weights equal to zero.

4.2.1. The Objective Function in the LP Models. Under the LP edge weight estimation models, Beyer et al. [5] and Waterman et al. [99] suggested the following objective functions:

$$L(\mathbf{X}, \mathbf{w}) = \sum_{e=1}^{2n-3} w_e, \tag{9}$$

$$L(\mathbf{X}, \mathbf{w}) = \sum_{i=1}^{n-1} \sum_{j>i}^{n} \frac{\sum_{e \in p_{ij}} w_e - d_{ij}}{d_{ij}}, \tag{10}$$

and

$$L(\mathbf{X}, \mathbf{w}) = \sum_{i=1}^{n-1} \sum_{j>i}^{n} \frac{\sum_{e \in p_{ij}} w_e - d_{ij}}{d_{ij}^2}. \tag{11}$$

The objective function (9) is analogous to Rzhetsky and Nei's (7) in the least-squares model [5]. On the other hand, the objective function (10) represents the overall sum of the deviations between the path that connects leaves and its corresponding evolutionary distance, whereas objective function (11) differs from (10) only in the use of a χ^2-type weighting [99]. Waterman et al. noted that both (10) and (11) can be considered as analogous LP versions of the WLS when $\omega_{ij} = 1/d_{ij}$ or $\omega_{ij} = 1/d_{ij}^2$, respectively [99].

5. APPROACHES TO SOLVING THE MEP

The MEP, as formulated in MEP, is an NP-hard problem [22] for which the literature provides both exact and nonexact

TABLE 4. References classified by approach to solution.

Approach to solution	References
Exact	
Polynomial-time algorithms	[15, 20, 88, 99]
Brute-force enumerations	[93]
Branch-and-bound algorithms	[15, 16, 71, 101, 102]
Nonexact	
Approximation algorithms	[1, 16, 29, 71, 101]
Heuristics	
Constructive	[18, 19, 25, 28, 31, 95]
Clustering	[7, 20, 41–43, 48, 87, 91]
Constructive/clustering	[21, 66, 80, 84, 85]

solution algorithms. Exact algorithms have been developed much more recently and are far less numerous than nonexact algorithms. Later, we discuss both classes of algorithms and list the respective references in Table 4.

5.1. Exact Algorithms

Waterman et al. [99] proved that MEP becomes easy when the distance matrix is additive and a tree metric phylogeny is required to satisfy (4). In this case, the authors showed that the solution is unique and, to determine it, provided an $O(n^2)$ algorithm, called sequential algorithm. Culberson [20] improved the algorithm by decreasing its order of complexity to $O(n \log n)$. A slower version of Culberson's algorithm, called ADDTREE, was proposed by Sattah and Tversky [88].

When dealing with generic distance matrices and using the OLS edge weight estimation model, the only program able to provide exact solutions to the instances of MEP is, to the best of our knowledge, PAUP 4.0* [93]. PAUP 4.0* combines an ingenuous exhaustive enumeration of the possible phylogenies $\mathbf{X}$ with an efficient computation of edge weights $\mathbf{w}$. Specifically, PAUP 4.0* implements Bryant and Waddell's algorithms (see [8] and Section 4.1) which can be used to estimate the edge weight vector $\mathbf{w}$ of a given phylogeny with an overall $O(n^2)$ computational complexity. PAUP 4.0* is able to solve instances containing no more than a dozen taxa.

Wu et al. [101] studied two families of instances of MEP, called the minimum ultrametric tree (MUT) problem and the metric minimum ultrametric tree (MMUT) problem. As their names suggest, the first set of instances is characterized by ultrametric distance matrices, and the second by metric ultrametric distance matrices which differ from the first only in the presence of the distance metric property. The authors proved that both the MUT and the MMUT problems are $\mathcal{NP}$-hard [29, 101], and provided an exact algorithm based on the implicit enumeration of all phylogenies, able to find optimal solutions for datasets containing up to 25 taxa. Wu et al.'s algorithm was then improved by Chen and Chang [15] who provided a tighter lower bound for the MUT problem, and by Yu et al. [102] who provided a parallel version of the algorithm. The latter improvement allowed the authors to find the optimal solutions for datasets containing up to 38 taxa.

Lu et al. [71] introduced a version of MEP in which the input is represented by a metric distance matrix and the solution is required to be a Steiner tree of the phylogenetic graph G. They proved that such a problem is $\mathcal{MAX}\,\mathcal{SNP}$-Hard. Chen et al. [16] even formulated a more particular case, called the bottleneck steiner tree (BST) problem, where the solution of MEP is a Steiner tree in which the greatest edge weight is minimized. The authors provided an exact algorithm of complexity $O(n \log n)$ to solve the BST problem, but unfortunately, neither Lu et al. nor Chen et al. provided any relation to previous analogous versions of MEP, or any biological interpretation of the resulting phylogeny.

Finally, to the best of our knowledge, there are no exact algorithms able to tackle instances of MEP under any other least-squares or LP edge weight estimation model.

5.2. Nonexact Algorithms

In contrast to exact algorithms, nonexact algorithms generally date back further and are more numerous. Nonexact algorithms are typically heuristics, i.e., algorithms that produce reasonably good solutions in short computing time ([78], p. 401). Although in general heuristics do not provide any formal guarantee of the quality of the solution found, for some of them (hereafter referred to as *approximation algorithms*), it is possible to prove that the solution found is optimal up to a small constant factor ([78], p. 409).

Before describing the main heuristics, we need to introduce some definitions. Define a *partial phylogeny* $\mathbf{X}_m$ of a set of taxa Γ as an m-leaf phylogeny whose leaves are taxa of a subset $\Gamma' \subset \Gamma$, with $m = |\Gamma'|$. Given a partial phylogeny $\mathbf{X}_m$ of Γ, denote $V_{\mathbf{X}_m}$ and $\mathcal{E}(\mathbf{X}_m)$ as the corresponding set of vertices and edges of $\mathbf{X}_m$, respectively. We say that we *insert a leaf* $i \notin \Gamma'$ into an edge $e = (r, s)$ of $\mathbf{X}_m$ when we generate a new partial phylogeny $\mathbf{X}'_m$ of Γ with vertex-set $V_{\mathbf{X}'_m} = V_{\mathbf{X}_m} \cup \{i, t\}$ and edge-set $\mathcal{E}_{\mathbf{X}'_m} = \mathcal{E}_{\mathbf{X}_m} \cup \{(r, t), (t, s), (t, i)\} \setminus \{(r, s)\}$. In other words, we insert a leaf i into the partial phylogeny when we divide edge $e = (r, s)$ with the new vertex t and we join the leaf i to t.

5.2.1. Approximation Algorithms. The family of approximation algorithms is relatively recent: to the best of our knowledge, the first works are dated 1999. The first approximation algorithm for MEP under a least-squares edge weight estimation model was provided by Argawala et al. [1]. The authors studied a particular version of MEP which minimizes the $\mathcal{L}_\infty$-vector norm of the differences between the entries of vector $\mathbf{w}$ and the entries of an ultrametric distance matrix $\mathbf{D}$. Argawala et al. provided for such a problem an algorithm characterized by an approximation ratio ([40], p. 128; [78], p. 409) equal to 3. Their work is at the core of Atteson's article [2] which has strong implications for the statistical consistency of a number of versions of MEP. We discuss this issue more extensively in Section 6.

More recently, Wu et al. [101] provided a $1.5(1 + \lceil \log n \rceil)$-approximation algorithm for the MUT problem (see Section 5.1), and Lu et al. [71] studied a particular case of MEP in

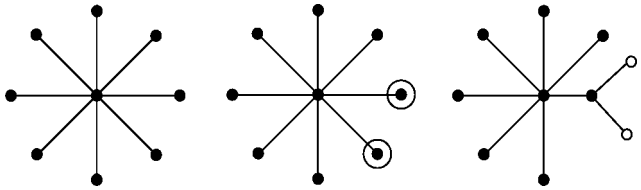

FIG. 4. Clustering heuristics: initially a graph-star is considered; subsequently two vertices (circled) are selected, marked (white vertices) and joined by an internal vertex. The algorithm is iterated on the remaining black vertices.

which edge weights are restricted to the values $\{1, 2\}$. Lu et al. proved that this particular case of MEP can be approximated with an error ratio of $8/5$, but did not provide any biological interpretation for the resulting phylogeny. Similarly, Chen et al. [16] provided a $2r$-approximation algorithm for the BST problem (see Section 5.1), where r is the best-known approximating ratio for the Steiner tree problem, but in this case also, the authors failed to provide any biological interpretation of the resulting phylogeny.

5.2.2. Heuristics. Heuristics can be classified according to the edge weight estimation model used and the strategy proposed to build a phylogeny. Specifically we can distinguish between classes of clustering, constructive, and constructive/clustering heuristics. The idea at the core of the class of clustering heuristics is the so-called *star decomposition algorithm* (see Fig. 4 and ([36], p. 48)). The algorithm starts from a graph-star containing n leaves and one internal vertex and then iteratively joins any pair of vertices with a new internal vertex until a phylogeny is obtained. By contrast, the idea at the core of the class of constructive heuristics is to start from a partial phylogeny of the set Γ and then to insert iteratively a new leaf until a phylogeny of the set Γ is obtained. Finally, the class of constructive/clustering heuristics combines the two previously cited approaches by generating typically a first phylogeny with a clustering and/or constructive heuristic and then applying a local search to it. Later, we discuss the main heuristics available in this area, postponing to Section 6 the analysis of their statistical consistency.

Heuristics Using Metric Properties. Heuristics using metric properties (triangle inequality, additive, and ultrametric properties) are typically constructive in nature. The earliest heuristic exploiting this strategy was provided by Farris in 1972 [31]. At each iteration, this heuristic computes all the possible insertions of a new leaf into a partial phylogeny and chooses the one whose length is minimal. Farris's heuristic requires metric distance matrices and exploits the triangle inequalities to compute edge weights of a given phylogeny. Tateno et al. [95] and Faith [28] extended Farris's heuristic by introducing techniques to handle ambiguities and errors in the input molecular data.

Csűrös *et al.* [18] improved the speed performances of Farris's heuristic by substituting its edge weight estimation with a new one exploiting the harmonic mean of a triplet of taxa. Csűrös [19] also showed that when the additive property holds for distance matrix **D**, his heuristic, called HGT/FP, is characterized by an order of complexity of $O(n^2)$.

Heuristics Using the OLS Model. Sneath and Sokal in 1963 [91] provided an $O(n^2)$ clustering heuristic called UPGMA to deal with instances of MEP under OLS edge weight estimation [30, 36]. UPGMA requires the distance matrix to be ultrametric to estimate the edge weights of a phylogeny. At each iteration, the algorithm joins the pair of vertices whose distance is minimal, and recomputes, by means of specific OLS formulae for ultrametric distance matrices, the distance between the new internal vertex and all other vertices not subjected to a joining process. This process is iterated until a phylogeny is obtained.

A similar clustering heuristic, called UNJ, was proposed by Gascuel [42, 43]. Unlike UPGMA, UNJ uses generic OLS formulae to estimate the edge weights of a phylogeny, so it can be applied to a distance matrix which is not necessarily ultrametric. The UNJ heuristic shares similar statistical consistency performances (see Section 6) with UPGMA but is characterized by an order of complexity of $O(n^3)$.

Desper and Gascuel [25] proposed an $O(n^2)$ constructive heuristic called greedy minimum evolution (GME). This heuristic is conceptually similar to Farris's [31] except that the type of edge weight estimation model is obtained through the OLS model. By means of particular formulae, the GME heuristic is able to increase the speed of computing the edge weight of the newly inserted leaf. Desper and Gascuel showed, through numerical experiments, that in terms of speed, GME is well suited to analyzing datasets containing thousands of taxa.

Rzhetsky and Nei [84, 85] proposed the first constructive/clustering heuristic to tackle instances of MEP under the OLS model. Starting from an initial phylogeny, the authors showed that the length of the best-so-far phylogeny could be improved by applying a local search that iteratively swaps any two leaves not linked to the same internal vertex. Rzhetsky and Nei's algorithm was subsequently improved by Kumar [66] who proposed selecting at each iteration a leaf i, and then trying all its possible $(n - 1)$ leaf-swaps on the phylogeny. Kumar showed that although his neighborhood is larger than Rzhetsky and Nei's neighborhood, as it involves the examination of a number of phylogenies that is at most an exponential function of the number of leaves n, the new algorithm provides results faster and better than that of Rzhetsky and Nei.

Heuristics Using the WLS Model. Some clustering variants of the UNJ heuristic, called BIONJ and Weighbor, were proposed by Gascuel [41] and Bruno et al. [7] respectively to provide a solution to MEP under the WLS model. These heuristics share a similar order of complexity and similar statistical consistency performances with the UNJ heuristic (see Section 6). However, they differ from UNJ

TABLE 5. Statistical consistency of different versions of MEP.

Estimation model	Objective function		
	$\sum_e w_e$	$\sum_e \lvert w_e \rvert$	$\sum_e \max(w_{e,o})$
Least-squares			
OLS	Consistent	Consistent	Consistent/unknown under p.c.
WLS	Partially consistent	Partially consistent	Partially consistent/inconsistent under p.c.
GLS	Inconsistent	Inconsistent	Inconsistent/inconsistent under p.c.
BLS	Consistent	n.a.	n.a.
Linear programming	$\sum_e w_e$	$\sum_{i,j} \frac{\sum_{e \in p_{ij}} w_e - d_{ij}}{d_{ij}}$	$\sum_{i,j} \frac{\sum_{e \in p_{ij}} w_e - d_{ij}}{d_{ij}^2}$
Beyer et al.'s model	Unknown	Unknown	Unknown
Waterman et al.'s model	Unknown	Unknown	Unknown

p.c.: positivity constraint.

mainly in that the variances and covariances of the evolutionary distances are exploited to estimate edge weights of the resulting phylogeny.

Heuristics Using the BLS Model. Saitou and Nei [87, 92] introduced an $O(n^3)$ clustering heuristic, called Neighbor-Joining (NJ), which is at the core of the previously cited UNJ, BIONJ, and Weighbor heuristics. The NJ heuristic defines *absolute distance* between any leaf i and all the others as $u_i = \sum_j d_{ij}/(n-2)$. Then, at each iteration, the NJ heuristic joins the pair of vertices (i,j) whose quantity $d_{ij} - u_i - u_j$ is the smallest and recomputes the distance between the new internal vertex and any other vertex k, not subjected to a joining process, as $(d_{ik} + d_{jk} - d_{ij})/2$. The joining step is then iterated until a phylogeny is obtained.

When Saitou and Nei introduced NJ, it was unclear which kind of edge weight estimation model was implemented by the NJ heuristic (although it was commonly accepted to be the OLS model [36]). Only recently, Desper and Gascuel [44, 48] showed that the NJ heuristic greedily optimizes MEP under the BLS model.

Desper and Gascuel [25] also proposed an $O(n^3)$ constructive heuristic called Balanced Minimum Evolution (BME). Like the GME heuristic, the BME heuristic is conceptually similar to Farris's heuristic previously cited [31], except that edge weight estimation is obtained though the BLS model. Desper and Gascuel showed that, in terms of speed, BME is outperformed by the NJ algorithm when analyzing small datasets, but is well suited to analyzing datasets containing thousands of taxa. BME returns tree-metric phylogenies when the triangle inequality holds for the distance matrix.

Da Silva et al. [21] proposed a more sophisticated constructive/clustering version of NJ. Specifically, the authors replaced the original vertex selection step of NJ with a local search. Da Silva et al. proved that the new algorithm is still comparable in terms of speed with the original version of NJ and pointed out that the new heuristic notably increases the probability of recovering optimal phylogenies of MEP under BLS edge weight estimation [21]. An earlier and very similar variant of this algorithm was proposed by Pearson et al. [80].

6. STATISTICAL CONSISTENCY OF THE MEP

Supposing that the assumptions adopted by a given criterion properly describe the real evolutionary process of a set of taxa (i.e., assuming that the real and the true phylogenies coincide), it is desirable to find a method that measures the ability (consistency) of a version of MEP to recover the true phylogeny as the amount of molecular data from the taxa increases.

In this section we discuss the statistical consistency of a number of versions of MEP. The results currently reported in the literature are summarized in Tables 5 and 6; Table 7 lists references dealing with the statistical consistency problem.

6.1. Statistical Consistency of the Edge Weight Estimation Models

Statistical consistency of the edge weight estimation models is defined as follows [45]. Assume a set Γ of n molecular data (e.g., a set of n DNA sequences) from taxa is given, and that each molecular data is characterized by a length l. Let $\mathbf{X}^*$ be the true phylogeny of Γ, $\Delta = \{\rho_{ij}\} = \{\sum_{e \in p_{ij}} x_{ij,e} w_e\}$ the

TABLE 6. Statistical consistency of a number of approaches to solving MEP.

Approach to solution	Safety radius
ADDTREE	$\epsilon/2$
Argawala's 3-approximating algorithm	$\epsilon/8$
BME	Unknown
BIONJ	$\epsilon/2$
GME	Unknown
ALSAD	$\epsilon/4$
ALSPC	Unknown
ALP	Unknown
NJ	$\epsilon/2$
Sequential algorithm	$\epsilon/2$
UNJ	$\epsilon/2$
UPGMA	ϵ

ALSAD, algorithms requiring additive distance matrices and using a least-squares model; ALSPC, algorithms using a least-squares model under the positivity constraint; ALP, algorithms using a linear programming model.

TABLE 7. References concerning statistical consistency classified by issue treated.

Issue	References
Statistical consistency of the edge weight estimation models	[23, 45, 53, 61, 85]
Bounds on the approach to the solution	[2, 25, 47, 54, 56, 69, 75, 100]

associated matrix of expected distance estimates (see Section 4.1), and $\mathbf{D}^l = \{d_{ij}^l\}$ the distance matrix, whose superscript 'l' means that its entries are estimated from pairwise molecular data having length l. We say that $\mathbf{D}^l$ is a consistent estimate of $\mathbf{\Delta}$ if the greater the amount of molecular data of taxa we have (e.g., the longer the DNA sequences), the closer $\mathbf{D}^l$ is to $\mathbf{\Delta}$: i.e.,

$$\lim_{l\to\infty} \Pr\left\{d_{ij}^l = \rho_{ij}\right\} = 1 \quad \forall i,j \in \Gamma.$$

Rzhetsky and Nei [85] proved that MEP returns statistically consistent results when the OLS model and objective function (7) are used. This result was extended by Denis and Gascuel [23] where the authors showed that the phylogenies provided by MEP remain consistent under the OLS model even when objective functions (8–8) are considered.

Denis and Gascuel also showed that MEP returns statistically consistent results when edge weights are estimated through the WLS model and when the weights ω_{ij} are set to the inverse of the product of two strictly positive constants α_i and α_j. However, in the most general case, the solutions provided by MEP under the WLS model as well as under the GLS model are inconsistent [45].

By contrast, the results provided by MEP are always consistent when edge weights are estimated through the use of the BLS model [24, 25]. Specifically, Desper and Gascuel [25] proved that the BLS model is a special form of the WLS model in which the variances ω_{ij} are directly proportional to $2^{\tau_{ij}}$, where τ_{ij} is the topological distance between taxa i and j on the optimal phylogeny, and inversely proportional to the molecular sequence length l. In summary, we can conclude that the only two known cases in which MEP is consistent under the WLS edge weights estimation are when $\omega_{ij} = 1/\alpha_i\alpha_j$ and $\omega_{ij} = 2^{\tau_{ij}} l$ respectively.

No result is known about the statistical consistency of the solutions provided by MEP when edge weights are estimated through LP. In ([36], p. 161), Felsenstein advances the hypothesis that when the divergence (and thus the evolutionary distance) among molecular data is high, this estimation method may approximate the parsimony criterion ([36], p. 1–18) which is notoriously inconsistent. However, to the best of our knowledge, no one has yet either validated or invalidated this hypothesis.

6.2. Bounds on the Statistical Consistency of Approaches to Solving the Minimum Evolution Problem

In his seminal work, Atteson [2] observed that in general a distance matrix $\mathbf{D}$ can be seen as the sum of the matrix $\mathbf{\Delta}$ of expected distance estimate associated to the true phylogeny $\mathbf{X}^*$ and an error matrix $\mathbf{E}$. Then, the author aimed to find an upper bound on the absolute value of the entries of $\mathbf{E}$ for which the phylogeny provided by a version of MEP approaches with certainty the true phylogeny. Atteson indicated with ϵ the smallest edge weight of the true phylogeny and defined *safety radius* (hereafter indicated with σ) the greatest entry of the matrix $\mathbf{E}$ (i.e., $\|\mathbf{E}\|_{\max}$) [47]. The author proved that, if the additive hypothesis holds for the distance matrix $\mathbf{D}$, an algorithm is statistically consistent if $\sigma \leq \epsilon/2$ [2]. This result was extended by Gascuel and McKenzie who proved that, when the distance matrix $\mathbf{D}$ is ultrametric, the maximum value that σ can achieve is ϵ [47].

Atteson proved that the Neighbor-Joining algorithm and ADDTREE are characterized by having $\sigma = \epsilon/2$, both for ultrametric and additive distance matrices [47]. The same result is also valid for Gascuel's BIONJ [41] and UNJ [42] algorithms, and Waterman et al.'s Sequential Algorithm [99]. Gascuel and McKenzie [47] further proved that, when the distance matrix is ultrametric, UPGMA [91] is characterized by $\sigma = \epsilon$. Moreover, the authors showed that, under ultrametric distance matrices, when an algorithm makes use of any least-squares model to estimate edge weights, the safety radius tends to zero as the number of leaves increases.

Recently, Willson [100] proved that any algorithm solving MEP under any least-squares edge weights estimation model is characterized by a safety radius at most half the safety radius of the Neighbor-Joining tree for large n. Similarly, Argawala et al.'s 3-approximating algorithm [1] is characterized by $\sigma = \epsilon/8$ ([44], p. 28).

To the best of our knowledge, no information about the safety radius of the BME and GME algorithms or the algorithms using the LP edge weight estimation model is currently known.

Different authors have studied the statistical consistency of some approaches that provide approximate solutions through evolutionary simulations [56], comparisons with (rare) known phylogenies [54], statistical evaluations [69], and congruence studies [75]. We refer the reader to Hillis's survey [54] about the statistical consistency of methods assuming the ultrametric hypothesis, and methods comparing the UPGMA with the Neighbor-Joining algorithm. Here we focus on the most recent heuristics. Desper and Gascuel [25] compared the statistical consistency of GME and BME algorithms with those of Bruno et al.'s Weighbor algorithm [7], the Neighbor-Joining algorithm [87, 92], Csűrös's HGT/FP algorithm [18, 19] and Felsenstein's FITCH algorithm [35]. The comparisons were made through simulations. Specifically, the authors used an artificial phylogeny as a reference point and computed the proportion of incorrect edges of the solutions provided by the algorithms tested. The authors showed that on small datasets (less than a few dozen taxa), FITCH performs better than other methods, but is characterized by a very slow computing time, whereas on large datasets (hundreds of taxa) Weighbor performs better. The authors also showed that on very large datasets (containing thousands of

taxa) GME and BME appear to be the best algorithms in terms of statistical consistency and computing time.

7. CONCLUSION

Although it is unlikely that evolution proceeds by following minimum paths, it is generally accepted that a minimal length phylogeny may properly fit the real phylogeny of well-conserved molecular data (i.e., data whose basic biochemical function has undergone minimal change throughout the evolution of a species) [5,36,44]. Indeed, since parallel mutations [77] are rare at the molecular level in well-conserved molecular data, in absence of convergent or reverse evolution it is reasonable to assume that a local minimum path can approximate the evolutionary process from one taxon to another [5]. That evolution follows a local rather than absolute minimum is due to: (i) the neighborhood of possible alleles that are selected at each instant of the life of a taxon is finite, and (ii) the selective forces acting on the taxon may not be constant throughout its evolution [5]. Over the long term (periods of environmental change, including the intracellular environment), the local minima will not in general combine to form an overall minimum. Therefore, a minimal length phylogeny provides a lower bound on the total number of mutation events that could have occurred along the evolution of the taxa analyzed. These are the fundamental considerations at the core of the ME criterion and of the corresponding MEP.

Here we have presented a general introduction and a review of the existing literature about the MEP. Our purpose has been to introduce a classification scheme to provide a general framework for papers appearing in this area. In particular, two main versions of the MEP have been outlined, the first concerning problems using least-squares edge weight estimation models, and the second concerning problems based on LP. This division has been further differentiated into different, approximately homogeneous sub-areas, and the basic aspects of each have been pointed out. For each, also, the most relevant issues affecting their use in tackling real-world sized problems have been outlined, as have the most interesting refinements deserving further research effort.

Acknowledgments

Thanks to Martine Labbé, Raffaele Pesenti, and the anonymous reviewers for their valuable comments on previous versions of the manuscript, and to Mike Steel for helpful and exciting discussions.

REFERENCES

[1] R. Argawala, V. Bafna, M. Farach, M. Paterson, and M. Thorup, On the approximability of numerical taxonomy (fitting distances by tree metrics), SIAM J Comput 28 (1999), 1073–1085.

[2] K. Atteson, The performance of the neighbor-joining methods of phylogenetic reconstruction, Algorithmica 25 (1999), 251–278.

[3] H. Bandelt and A. Dress, Split decomposition: A new and useful approach to phylogenetic analysis of distance data, Mol Phylogenet Evol 1 (1992), 242–252.

[4] J.P. Barthélemy and A. Guénoche, Trees and proximity representations, Wiley, New York, 1991.

[5] W.A. Beyer, M. Stein, T. Smith, and S. Ulam, A molecular sequence metric and evolutionary trees, Math Biosci 19 (1974), 9–25.

[6] A. Björck, Numerical methods for least-squares problems, SIAM, Philadelphia, PA, 1996.

[7] W.J. Bruno, M.D. Socci, and A.L. Halpern, Weighted neighbor-joining: A likelihood-based approach to distance-based phylogeny reconstruction, Mol Biol Evol 17 (2000), 189–197.

[8] D. Bryant and P. Waddell, Rapid evaluation of least-squares and minimum evolution criteria on phylogenetic trees, Mol Biol Evol 15 (1998), 1346–1359.

[9] M. Bulmer, Use of the method of generalized least-squares in reconstructing phylogenies from sequence data, Mol Biol Evol 8 (1991), 868–883.

[10] P. Buneman, "The recovery of trees from measure of dissimilarities, Archaeological and historical science," F.R. Hodson, D.G. Kendall, and P. Tautu (Editors), Edinburgh University Press, Edinburgh, UK, 1971, pp. 387–395.

[11] D. Catanzaro, R. Pesenti, and M. Milinkovitch, A non-linear optimization procedure to estimate distances and instantaneous substitution rate matrices under the GTR model, Bioinformatics 22 (2006), 708–715.

[12] L.L. Cavalli-Sforza and A.W.F. Edwards, Analysis of human evolution, Proc XI Int Congress Genet, Pergamon Press, Oxford, UK, 1963, pp. 923–933.

[13] L.L. Cavalli-Sforza and A.W.F. Edwards, Phylogenetic analysis: Models and estimation procedures, Am J Hum Genet 19 (1967), 233–257.

[14] R. Chakraborty, Estimation of time of divergence from phylogenetic studies, Can J Cytol Genet 19 (1977), 217–223.

[15] H.F. Chen and M.S. Chang, An efficient exact algorithm for the minimum ultrametric tree problem, Lecture Notes Comput Sci 3341 (2004), 282–293.

[16] Y.H. Chen, C.L.L. Lu, and C.Y. Tang, On the full and bottleneck full Steiner tree problems, Lecture Notes Comput Sci 2697 (2003), 122–129.

[17] B. Chor and T. Tuller, Maximum likelihood of evolutionary trees: Hardness and approximation, Bioinformatics 21 (2005), i97–i106.

[18] M. Csűrös, Fast recovery of evolutionary trees with thousands of nodes, J Comput Biol 9 (2002), 277–297.

[19] M. Csűrös and Y. Kao, Provably fast and accurate recovery of evolutionary trees through harmonic greedy triplets, SIAM J Comput 31 (2001), 306–322.

[20] J. Culberson and P. Rudnicki, A fast algorithm for constructing trees from distance matrices, Inform Process Lett 30 (1989), 215–220.

[21] A.E. da Silva, W.J.P. Villanueva, H. Knidel, V. Bonato, S.F. dos Reis, and F.J.V. Zuben, A multi-neighbor-joining approach for phylogenetic tree reconstruction and visualization, Genet Mol Res 4 (2005), 525–534.

[22] W.H.E. Day, Computational complexity of inferring phylogenies from dissimilarity matrices, B Math Biol 49 (1987), 461–467.

[23] F. Denis and O. Gascuel, On the consistency of the minimum evolution principle of phylogenetic inference, Discrete Appl Math 127 (2003), 66–77.

[24] R. Desper and O. Gascuel, Fast and accurate phylogeny reconstruction algorithms based on the minimum evolution principle, J Comput Biol 9 (2002), 687–705.

[25] R. Desper and O. Gascuel, Theoretical foundations of the balanced minimum evolution method of phylogenetic inference and its relationship to the weighted least-squares tree fitting, Mol Biol Evol 21 (2004), 587–598.

[26] A.J. Dobson, Unrooted trees for numerical taxonomy, J Appl Probab 11 (1974), 32–42.

[27] P. Erixon, B. Svennblad, T. Britton, and B. Oxelman, Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics, Syst Biol 52 (2003), 665–673.

[28] D.P. Faith, Distance methods and the approximation of most-parsimonious trees, Syst Zool 43 (1985), 312–325.

[29] M. Farach, S. Kannan, and T. Warnow, A robust model for finding optimal evolutionary trees, Algorithmica 13 (1995), 155–179.

[30] J.S. Farris, On the cophenetic correlation coefficient, Syst Zool 18 (1969), 279–285.

[31] J.S. Farris, Estimating phylogenetic trees from distance matrices, Am Nat 106 (1972), 645–668.

[32] J. Felsenstein, Cases in which parsimony or compatibility methods will be positively misleading, Syst Zool 27 (1978), 401–410.

[33] J. Felsenstein, Evolutionary trees from DNA sequences: A maximum likelihood approach, J Mol Evol 17 (1981), 368–376.

[34] J. Felsenstein, Distance methods for inferring phylogenies: A justification, Evolution 38 (1984), 16–24.

[35] J. Felsenstein, An alternating least-squares approach to inferring phylogenies from pairwise distances, Syst Biol 46 (1997), 101–111.

[36] J. Felsenstein, Inferring phylogenies, Sinauer Associates, Sunderland, MA, 2004.

[37] W.M. Fitch, Rate of change of concomitantly variable codons, J Mol Evol 1 (1971), 84–96.

[38] W.M. Fitch and E. Margoliash, Construction of phylogenetic trees, Sci 155 (1967), 279–284.

[39] N. Galtier, Maximum-likelihood phylogenetic analysis under covarion-like model, Mol Biol Evol 18 (2001), 866–873.

[40] M.R. Garey and D.S. Johnson, Computers and intractability: A guide to the theory of NP-completeness, Freeman, New York, 2003.

[41] O. Gascuel, BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data, Mol Biol Evol 14 (1997), 685–695.

[42] O. Gascuel, "Concerning the NJ algorithm and its unweighted version," UNJ, Mathematical hierarchies and biology, B. Mirkin, F.R. McMorris, F. Roberts, and A. Rzhetsky (Editors), American Mathematical Society, Providence, RI, 1997, pp. 149–170.

[43] O. Gascuel, On the optimization principle in phylogenetic analysis and the minimum evolution criterion, J Classif 19 (2000), 67–69.

[44] O. Gascuel, Mathematics of evolution and phylogeny, Oxford University Press, New York, 2005.

[45] O. Gascuel, D. Bryant, and F. Denis, Strengths and limitations of the minimum evolution principle, Syst Biol 50 (2001), 621–627.

[46] O. Gascuel and D. Levy, A reduction algorithm for approximating a (non-metric) dissimilarity by a tree distance, J Classif 13 (1996), 129–155.

[47] O. Gascuel and A. McKenzie, Performance analysis of hierarchical clustering algorithms, J Classif 21 (2004), 3–18.

[48] O. Gascuel and M. Steel, Neighbor-joining revealed, Mol Biol Evol 23 (2006), 1997–2000.

[49] G.H. Gonnet, C. Korostensky, and S. Benner, Evaluation measures of multiple sequence alignments, J Comput Biol 7 (2000), 261–276.

[50] R.L. Graham and L.R. Foulds, Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time, Math Biosci 60 (1982), 133–142.

[51] X. Gu and W.H. Li, A general additive distance with time-reversibility and rate variation among nucleotide sites, Proc Natl Acad Sci USA 93 (1996), 4671–4676.

[52] M. Hasegawa, H. Kishino, and T. Yano, Evolutionary trees from DNA sequences: A maximum likelihood approach, J Mol Evol 17 (1981), 368–376.

[53] M. Hasegawa, H. Kishino, and T. Yano, Dating the human-ape splitting by a molecular clock of mitochondrial DNA, J Mol Evol 22 (1985), 160–174.

[54] D.M. Hillis, Approaches for assessing phylogenetic accuracy, Syst Biol 44 (1995), 3–16.

[55] L.J. Hubert and P. Arabie, Iterative projection strategies for the least-squares fitting of tree structures to proximity data, Brit J Math Stat Psy 48 (1995), 281–317.

[56] J.P. Huelsenbeck, Performance of phylogenetic methods in simulation, Syst Biol 44 (1995), 17–48.

[57] J.P. Huelsenbeck, Testing a covariotide model of DNA substitution, Mol Biol Evol 19 (2002), 698–707.

[58] J.P. Huelsenbeck, F. Ronquist, R. Nielsen, and J.P. Bollback, Bayesian inference of phylogeny and its impact on evolutionary biology, Sci 294 (2001), 2310–2314.

[59] D. Huson, S. Nettles, and T. Warnow, Disk-covering, a fast converging method for phylogenetic tree reconstruction, J Comput Biol 6 (1999), 369–386.

[60] T.H. Jukes and C. Cantor, "Evolution of protein molecules," Mammalian protein metabolism, H.N. Munro (Editor), Academic Press, New York, 1969, pp. 21–123.

[61] K.K. Kidd and L.A. Sgaramella-Zonta, Phylogenetic analysis: Concepts and methods, Am J Hum Genet 23 (1971), 235–252.

[62] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, J Mol Evol 16 (1980), 111–120.

[63] V. King, L. Zhang, and Y. Zhou, On the complexity of distance-based evolutionary tree reconstruction, Proc Fourteenth Ann ACM-SIAM Symp Discr Algorithms, SIAM, Philadelphia, PA, 2003, pp. 444–453.

[64] C. Korostensky and G.H. Gonnet, Using traveling salesman problem algorithms for evolutionary tree construction, Bioinformatics 16 (2000), 619–627.

[65] M.K. Kuhner and J. Felsenstein, A simulation comparison of phylogeny algorithms under equal and unequal rates, Mol Biol Evol 11 (1994), 584–593.

[66] S. Kumar, A stepwise algorithm for finding minimum evolution evolutionary trees, Mol Biol Evol 13 (1996), 584–593.

[67] C. Lanave, G. Preparata, C. Saccone, and G. Serio, A new method for calculating evolutionary substitution rates, J Mol Evol 20 (1984), 86–93.

[68] B. Larget and D.L. Simon, Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees, Mol Biol Evol 16 (1999), 750–759.

[69] W.H. Li and A. Zharkikh, Statistical tests of DNA phylogenies, Syst Biol 44 (1995), 49–63.

[70] P. Lopez, D. Casane, and H. Philippe, Heterotachy: An important process of protein evolution, Mol Biol Evol 19 (2002), 1–7.

[71] C.L. Lu, C.Y. Tang, and R.C.T. Lee, The full Steiner tree problem, Theor Comput Sci 306 (2003), 55–67.

[72] V. Makarenkov and F.J. Lapointe, A weighted least-squares approach for inferring phylogenies from incomplete distance matrices, Bioinformatics 20 (2004), 2113–2121.

[73] V. Makarenkov and B. Leclerc, "Circular orders of tree metrics and their uses for the reconstruction and fitting of phylogenetic trees," Mathematical hierarchies and biology, B. Mirkin, F.R. McMorris, F. Roberts, and A. Rzhetsky (Editors), American Mathematical Society, Providence, RI, 1997, pp. 183–208.

[74] V. Makarenkov and B. Leclerc, An algorithm for the fitting of a tree metric according to a weighted least-squares criterion, J Classif 16 (1999), 3–26.

[75] M.M. Mayamoto and W.M. Fitch, Testing species phylogenies and phylogenetic methods with congruence, Syst Biol 44 (1995), 64–76.

[76] G.L. Nemhauser and L.A. Wolsey, Integer and combinatorial optimization, Wiley-Interscience, New York, 1999.

[77] R.D.M. Page and E.C. Holmes, Molecular evolution: A phylogenetic approach, Blackwell Science, Oxford, UK, 1998.

[78] C. Papadimitriou and K. Steiglitz, Combinatorial optimization: Algorithms and complexity, Dover Publications, Mineola, NY, 1998.

[79] Y. Pauplin, Direct calculation of a tree length using a distance matrix, J Mol Evol 51 (2000), 41–47.

[80] W.R. Pearson, G. Robins, and T. Zhang, Generalized neighbor-joining: More reliable phylogenetic tree reconstruction, Mol Biol Evol 16 (1999), 806–816.

[81] D. Penny, M.D. Hendy, and M.A. Steel, "Testing the theory of descendent," Phylogenetic analysis of DNA sequences, M.M. Mayamoto and J. Cracraft (Editors), Oxford University Press, New York, 1991, pp. 155–183.

[82] S. Roch, A short proof that phylogenetic tree reconstruction by maximum likelihood is hard, IEEE-ACM T Comput Biol 3 (2006), 92–94.

[83] F. Rodriguez, J.L. Oliver, A. Marin, and J.R. Medina, The general stochastic model of nucleotide substitution, J Theor Biol 142 (1990), 485–501.

[84] A. Rzhetsky and M. Nei, Statistical properties of the ordinary least-squares generalized least-squares and minimum evolution methods of phylogenetic inference, J Mol Evol 35 (1992), 367–375.

[85] A. Rzhetsky and M. Nei, Theoretical foundations of the minimum evolution method of phylogenetic inference, Mol Biol Evol 10 (1993), 1073–1095.

[86] A. Rzhetsky and M. Nei, METREE: A program package for inferring and testing minimum evolution trees, Comput Appl Biosci 10 (1994), 409–412.

[87] N. Saitou and M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees, Mol Biol Evol 4 (1987), 406–425.

[88] S. Sattah and A. Tversky, Additive similarity trees, Psychometrika 42 (1977), 319–345.

[89] C. Semple and M. Steel, Phylogenetics, Oxford University Press, New York, 2003.

[90] C. Semple and M. Steel, Cyclic permutations and evolutionary trees, Adv Appl Math 32 (2004), 669–680.

[91] P.H.A. Sneath and R.R. Sokal, Numerical taxonomy, W. K. Freeman and Company, San Francisco, CA, 1963.

[92] J.A. Studier and K.J. Keppler, A note on the neighbor-joining algorithm of Saitou and Nei, Mol Biol Evol 5 (1988), 729–731.

[93] D.L. Swofford, PAUP* version 4.0, Sinauer, Sunderland, MA, 1997.

[94] D.L. Swofford, G.J. Olsen, P.J. Waddell, and D.M. Hillis, "Phylogenetic inference," Molecular systematics, D.M. Hillis, C. Moritz, and B.K. Mable (Editors), Sinauer & Associates, Sunderland, MA, 1996, pp. 407–514.

[95] Y. Tateno, M. Nei, and F. Tajima, Accuracy of estimated phylogenetic trees from molecular data, J Mol Evol 18 (1982), 387–404.

[96] W. Vach, "Least-squares approximation of additive trees," Conceptual and numerical analysis of data, O. Opitz (Editor), Springer-Verlag, Berlin, 1989, pp. 230–238.

[97] W. Vach and P.O. Degens, Least-squares approximation of additive trees to dissimilarities: Characterization and algorithms, Comput Stat Q 3 (1991), 203–218.

[98] P.J. Waddell and M.A. Steel, General time-reversible distances with unequal rates across sites: Mixing gamma and inverse gaussian distributions with invariant sites, Mol Phylogenet Evol 8 (1997), 398–414.

[99] M.S. Waterman, T.F. Smith, M. Singh, and W.A. Beyer, Additive evolutionary trees, J Theor Biol 64 (1977), 199–213.

[100] S.J. Willson, Minimum evolution using ordinary least-squares is less robust than neighbor-joining, B Math Biol 67 (2005), 261–279.

[101] B.Y. Wu, K.M. Chao, and C.Y. Tang, Approximation and exact algorithms for constructing minimum ultrametric trees from distance matrices, J Comb Optim 3 (1999), 199–211.

[102] K.M. Yu, J. Zhou, C.Y. Lin, and C.Y. Tang, Parallel branch-and-bound algorithm for constructing evolutionary trees from distance matrix, Proc 8th Inter Conf High-Perform Comput Asia-Pacific Region, IEEE Computer Society, 2005, pp. 66–72.