Syst. Biol. 53(6):963–967, 2004

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150490522728

Molecular Clock Fork Phylogenies: Closed Form Analytic Maximum Likelihood Solutions

BENNY CHOR[1] AND SAGI SNIR[2]

[1]*School of Computer Science, Tel-Aviv University, Tel-Aviv 39040 Israel; E-mail: benny@cs.tau.ac.il(B.C.)*
[2]*Computer Science Department, Technion, Haifa 32000, Israel; E-mail: ssagi@math.berkeley.edu (s.s.)*

Abstract.—Maximum likelihood (ML) is increasingly used as an optimality criterion for selecting evolutionary trees (Felsenstein, 1981, J. Mol. Evol. 17:368–376), but finding the global optimum is a hard computational task. Because no general analytic solution is known, numeric techniques such as hill climbing or expectation maximization (EM) are used in order to find optimal parameters for a given tree. So far, analytic solutions were derived only for the simplest model—*three*-taxa, two-state characters, under a molecular clock. Quoting Ziheng Yang (2000, Proc. R. Soc. B 267:109–119), who initiated the analytic approach, "*this seems to be the simplest case, but has many of the conceptual and statistical complexities involved in phylogenetic estimation.*" In this work, we give general analytic solutions for a family of trees with *four*-taxa, two-state characters, under a molecular clock. The change from three to four taxa incurs a major increase in the complexity of the underlying algebraic system, and requires novel techniques and approaches. We start by presenting the general maximum likelihood problem on phylogenetic trees as a constrained optimization problem, and the resulting system of polynomial equations. In full generality, it is infeasible to solve this system, therefore specialized tools for the molecular clock case are developed. Four-taxa rooted trees have two topologies—the *fork* (two subtrees with two leaves each) and the *comb* (one subtree with three leaves, the other with a single leaf). We combine the ultrametric properties of molecular clock fork trees with the Hadamard conjugation (Hendy and Penny, 1993, J. Classif. 10:5–24) to derive a number of topology dependent identities. Employing these identities, we substantially simplify the system of polynomial equations for the fork. We finally employ symbolic algebra software to obtain *closed form* analytic solutions (expressed parametrically in the input data). In general, four-taxa trees can have multiple ML points (Steel, 1994, Syst. Biol. 43:560–564; Chor et al., 2000, MBE 17:1529–1541). In contrast, we can now prove that each fork topology has a *unique* (local and global) ML point. [Analytic solutions; Hadamard conjugation; maximum likelihood; molecular clock; phylogenetic reconstruction; symbolic algebra software; systems of polynomial equations.]

The study of evolution and the construction of phylogenetic (evolutionary) trees are classical subjects in biology. DNA sequences from a variety of organisms are rapidly accumulating, providing the data to a number of sequence based approaches for phylogenetic trees reconstruction. Given a set of n aligned *sequences*, the goal is to find the best explanation for the data within the model space. Among tree reconstruction approaches, maximum likelihood (Felsenstein, 1981) is increasingly used as an optimality criterion for inferring trees. In the phylogeny context, this usually means a weighted tree (the weights are parameters of the substitution model for each edge) that maximizes the likelihood of generating the observed sequences. Maximum likelihood (ML) algorithms optimize both over trees and, for a given tree, over all edge lengths. This "double optimization" makes ML algorithms computationally intensive. Still, for tractable cases ML is the method of choice. Because no general analytical solution is available, numeric techniques (such as hill climbing or expectation maximization) are used in order to find optimal likelihood values for any given tree. The first to consider *analytical solutions* for simple substitution models with a small number of taxa was Yang (2000), who worked on three taxa with two-state characters under molecular clock (Yang, 2000). Yang calls this "the simplest phylogeny estimation problem," but adds that it "has many of the conceptual and statistical complexities involved in phylogenetic estimation." The solution of Yang was generalized and its derivation was simplified by Chor et al. (2001) using the Hadamard Conjugation of Hendy et al. (Hendy and Penny, 1993; Hendy et al., 1994), together with convexity arguments. In this work we retain the symmetric two-state model of Neyman (1971), as used in (Yang, 2000; Chor et al., 2001) under molecular clock, but increase the number of taxa to four. There are two families of rooted trees on four taxa: Trees with two taxa in each subtree of the root, which we call *forks*, and trees where one subtree of the root has three taxa, which we call *combs*. Under molecular clock, the distance from each of the four leaves to the root is the same (Fig. 1). In this work we focus on the fork topology.

The change from three to four taxa incurs a major increase in the complexity of the underlying system of polynomial equations, and requires novel techniques and approaches. Our starting point, following Chor et al. (2000), is to formulate the ML problem as one of constrained optimization, and express it in terms of Lagrange coefficients (Strang, 1988: p. 415). We then use the Hadamard conjugation (Hendy and Penny, 1993; Hendy et al., 1994) to simplify the resulting system of polynomial equations. This yields a system of nine degree 5 polynomials in nine variables. This system is substantially more complex than the three-taxa system (Chor et al., 2001), and is not solvable by current techniques. (The analytical solutions in (Chor et al., 2000) were obtained for special cases where at least two out of the seven input parameters are 0.) Using the molecular clock assumption together with the specifics of the fork topology, we derive a number of identities and equations that allow a substantial simplification to the complexity of the polynomial system. By using these identities, the system of equations becomes simple enough to enable its solution—not manually, but by employing computer algebra tools (*e.g.*, Maple). This leads to the derivation of a *closed form* analytical solution, expressed

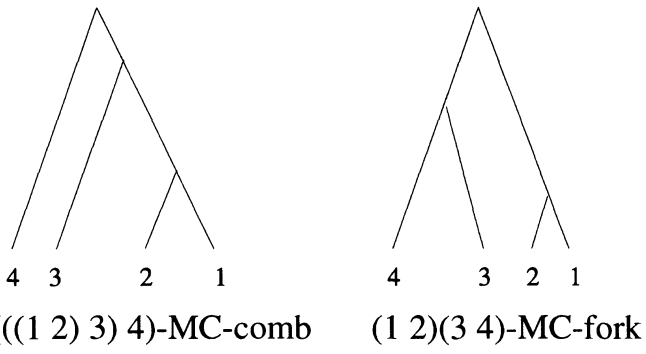

FIGURE 1. The fork and comb—two rooted trees on four taxa.

as rational functions in the input parameters. This solution is unique, implying a single local and global ML point.

Even in cases where it is feasible to derive them, analytical solutions will most probably *not* replace numeric approaches in ML based tree reconstruction packages. But the analytic solutions do reveal properties of the maximum likelihood points that are not obtainable numerically. For small trees, results along this line of research can serve as a method for checking the accuracy of the heuristic methods used in practice. This work is also useful in supertrees methods (constructing a big tree from small subtrees) that use rooted *quartets* as their building blocks (Aho et al., 1981).

The remainder of this work is organized as follows: First, we give a high level overview of our approach and methods, and briefly explain the solution. Then we give some concluding remarks and directions for further research.

METHODS—OVERVIEW

In this section we define the model of substitution that we use, introduce useful notation, and briefly describe the Hadamard conjugation and its usage for analyzing maximum likelihood. We then give the highlights (but not the details) of our analytical ML solution. A manuscript containing the full mathematical details of this work can be found in http://www.cs.tau.ac.il/~bchor/fork-full.pdf.

In the Neyman 2 states model (Neymann, 1971), each character admits one out of two states, e.g., purines and pyramidines. Without loss of generality, we denote these states by {x, y}. We use the symmetric Poisson model where for each edge e of the tree T, there is a corresponding probability p_e ($p_e \leq 1/2$) that the character states at the two incident vertices of e differ, and this probability is independent of the state at the initial vertex. The probability p_e is the probability of having an odd (1, 3, 5, ...) number of substitutions per site across the edge e. The expected number of substitutions per site across the edge e equals $q_e = -\frac{1}{2}\ln(1-2p_e)$. Measuring the tree edges by q_e ($q_e < \infty$), we get an *additive* measure on the tree (because expected values are additive). When drawing weighted trees, we usually refer to q_e as either the *length* or the *weight* of the edge e.

Such a weighted phylogenetic tree is a probabilistic model that emits any given pattern of characters at its leaves with a well defined probability. Given these "pattern generating probabilities," it is easy to compute the *likelihood* that the tree generates a given set of *observed sequences*. We now outline how this is done.

The input data to a character based tree construction algorithm (such as ML) is a set of aligned sequences. In our case, the entries of the sequences are x and y. The sequences can be represented by a matrix, ψ, where the number of rows equals the number of species (four in our case), and the number of columns equals the common length of the sequences. Interchanging any two columns of ψ does not change the likelihood. It is thus convenient to "summarize" the observed data ψ by the so called *vector of observed splits*, $\hat{\mathbf{s}}$. This vector simply counts how many sites share any specific pattern. Under a fully symmetric model, the probability of a pattern is equal to that of its complement (where all x and y are interchanged).

We make the following convention about indexing the patterns obtained in the sequences over $n = 4$ species, labeled 1, 2, 3, and 4: We identify a site pattern by the subset of species {1, 2, 3} whose character at that site is different from that of species 4. In general, for every $\alpha \subseteq \{1, \ldots, n-1\}$, an *$\alpha$-split pattern* is a pattern where all taxa in the subset α have one character (x or y), and the taxa in the complement subset have the second character (there are two such patterns). The value $\hat{s}_\alpha$ equals the number of times that α-split patterns appear in the data. If $n = 4$ then there are $2^3 = 8$ possible patterns, and the vector of observed sequence frequencies is $\hat{\mathbf{s}} = [\hat{s}_\emptyset, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_4]$. (For example, the number of occurrences of the two constant patterns

$$\begin{bmatrix}\mathbf{x}\\ \mathbf{x}\\ \mathbf{x}\\ \mathbf{x}\end{bmatrix}, \begin{bmatrix}\mathbf{y}\\ \mathbf{y}\\ \mathbf{y}\\ \mathbf{y}\end{bmatrix}$$

equals $\hat{s}_\emptyset$, whereas the number of occurrences of the two patterns separating {1, 3} from {2, 4},

$$\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\\ \mathbf{x}\\ \mathbf{y}\end{bmatrix}, \begin{bmatrix}\mathbf{y}\\ \mathbf{x}\\ \mathbf{y}\\ \mathbf{x}\end{bmatrix}$$

equals $\hat{s}_{13}$.) Table 1 depicts the vector of observed splits for three datasets of observed sequences.

Given a tree T with n leaves and edge probabilities $\mathbf{p} = [p_e]_{e \in E(T)}$ ($0 \leq p_e \leq \frac{1}{2}$), the probability of generating an α-split pattern ($\alpha \subseteq \{1, \ldots, n-1\}$) is well defined (and is equal for all sites). Denote this probability by $s_\alpha = Pr(\alpha\text{-split}|T, \mathbf{p})$. Using the same indexing scheme as above, we define the vector of *pattern generation probabilities* $\mathbf{s} = [s_\emptyset, s_1, s_2, s_{12}, s_3, s_{13}, s_{23}, s_4]$ (this vector is termed the *expected sequence spectrum* in (Hendy and Penny,

TABLE 1. Three data sets (A, B, and C) and their vector of observed splits.

	$\hat{\mathbf{s}}_A = [7, 0, 0, 1, 0, 1, 1, 0]$ Observed data (A)	$\hat{\mathbf{s}}_B = [14, 0, 0, 3, 0, 2, 1, 0]$ Observed data (B)	$\hat{\mathbf{s}}_C = [10, 2, 2, 4, 0, 1, 1, 0]$ Observed data (C)
1	xxxxyyy y x y	xxxxxxxyyyyyyy xxy xy y	xxxxxyyyyy xy yx xxyy x y
2	xxxxyyy y y x	xxxxxxxyyyyyyy xxy yx x	xxxxxyyyyy yx xy xxyy y x
3	xxxxyyy x x x	xxxxxxxyyyyyyy yyx xy x	xxxxxyyyyy yx yx yyxx x x
4	xxxxyyy x y y	xxxxxxxyyyyyyy yyx yx y	xxxxxyyyyy yx yx yyxx y y

1993)). Having this vector at hand greatly facilitates the calculation and analysis of the likelihood, because the likelihood of observing $\hat{\mathbf{s}}$, given $\mathbf{s}$, equals

$$L(\hat{\mathbf{s}} \mid \mathbf{s}) = \prod_{\alpha} Pr(\alpha\text{-split} \mid \mathbf{s})^{\hat{\mathbf{s}}_\alpha} = \prod_{\hat{\mathbf{s}}_\alpha > 0} \mathbf{S}_\alpha^{\hat{\mathbf{s}}\alpha} \qquad (1)$$

Although in principle the edge lengths $\mathbf{q} = [q_e]_{e \in E(T)}$ determine the vector of pattern generation probabilities $\mathbf{s} = [s_\emptyset, s_1, s_2, s_{12}, s_3, s_{13}, s_{23}, s_4]$, it is not obvious how to actually compute this vector, given the edge lengths. This is where the Hadamard conjugation (Hendy and Penny, 1993; Hendy et al., 1994) comes in. It is an invertible transformation that gives a concrete (and reasonably efficient) way of computing the pattern generation probabilities from the vector of edge weights, $\mathbf{q}$.

To analytically compute the maximum likelihood point (or points) for a given tree topology, we look for critical points of the likelihood function L(observation|tree parameters). Our goal is to maximize $\ln(L)$ over the eight dimensional space $\mathbf{s} = [s_\phi, s_1, s_2, s_{12}, s_3, s_{13}, s_{23}, s_4]$, bound to the eight inequalities $0 \le s_\alpha \le 1$, the equality $\sum_\alpha s_\alpha = 1$, and the constraints that s represents a point in the probability space of a tree T. That is, the corresponding edge spectrum $\mathbf{q}$ represents T. An eight-dimensional point $\mathbf{q} = [q_\emptyset, q_1, q_2, q_{12}, q_3, q_{13}, q_{23}, q_4]$ represents a four-taxa tree $T = (12)(34)$ if $q_\emptyset, q_1, q_2, q_{12}, q_3$, and q_4 are all nonnegative whereas $q_{13} = q_{23} = 0$.

In order to solve the constrained optimization problem for the tree, we first express the two constraints $q_{13} = 0$ and $q_{23} = 0$ in terms of the eight components of s, using the relation between s and q provided by the Hadamard conjugation. We now use Lagrange multipliers to find the turning points of $\ln L$, bound by the two constraints $q_{13} = 0$ and $q_{23} = 0$. This requires solving sets of equations involving partial derivatives of the likelihood, L, with respect to various pattern generation probabilities, $\partial L / \partial s_\alpha$. We get a system of nine equations in nine variables—the seven s_α and the two Lagrange multipliers μ and λ. In general, the resulting system of equations is hard to solve, even using computer algebra packages.

It is at this point that the molecular clock assumption plays a critical role. Under a molecular clock, a number of additional equalities among edge weights hold. For example, in Figure 2, $q_1 = q_2$ and $q_3 = q_4$. These relationship substantially simplify the system of equations. Not only do they enable its solution, but even let us come up with a "closed form solution," where every coordinate of

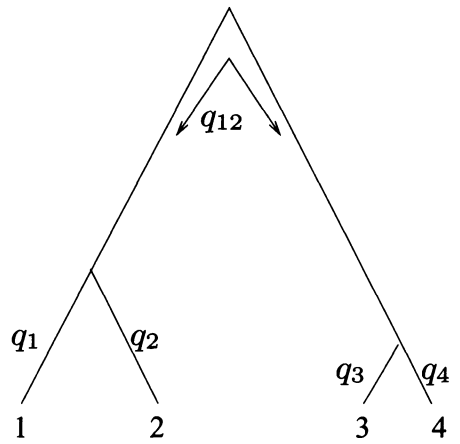

FIGURE 2. Molecular clock fork.

the ML point, $\mathbf{q}$, can be expressed as a ratio of simple, low degree polynomials in the input parameters (numbers of each observed pattern).

Without further developing the technical details, we exhibit below the analytic solutions for the ML molecular clock fork.

Theorem 1: *Let* $\hat{\mathbf{s}} = (\hat{s}_\emptyset, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_4)$ *be the vector of observed splits, and let* $\sum_\alpha \hat{s}_\alpha = c$. *Then the vector of pattern generation probabilities in the the ML* (12)(34) *molecular clock fork equals:*

$$\begin{aligned}
s_{13} = s_{23} = \big(&\hat{s}_{23}^2 + \hat{s}_1\,\hat{s}_{23} + \hat{s}_2\,\hat{s}_{23} + \hat{s}_4\,\hat{s}_{23} + 2\,\hat{s}_{13}\,\hat{s}_{23} \\
&+ \hat{s}_3\,\hat{s}_{23} + \hat{s}_{13}\,\hat{s}_3 + \hat{s}_{13}^2 + \hat{s}_{13}\,\hat{s}_4 + \hat{s}_2\,\hat{s}_3 + \hat{s}_2\,\hat{s}_4 \\
&+ \hat{s}_2\,\hat{s}_{13} + \hat{s}_1\,\hat{s}_3 + \hat{s}_1\,\hat{s}_4 + \hat{s}_1\,\hat{s}_{13}\big)2c^2
\end{aligned}$$

$$s_{12} = \frac{(c - \hat{s}_2 - \hat{s}_1 - \hat{s}_{23} - \hat{s}_{13})(c - \hat{s}_3 - \hat{s}_{23} - \hat{s}_4 - \hat{s}_{13})\hat{s}_{12}}{c^2(c - \hat{s}_1 - \hat{s}_2 - \hat{s}_3 - \hat{s}_4 - \hat{s}_{13} - \hat{s}_{23})}$$

$$\begin{aligned}
s_1 = s_2 = -\big(&\hat{s}_{23}^2 - c\hat{s}_{23} + \hat{s}_2\,\hat{s}_{23} + \hat{s}_4\,\hat{s}_{23} + \hat{s}_3\,\hat{s}_{23} + \hat{s}_1\,\hat{s}_{23} \\
&+ 2\hat{s}_{13}\,\hat{s}_{23} - c\hat{s}_{13} + \hat{s}_{13}\,\hat{s}_3 + \hat{s}_{13}^2 + \hat{s}_{13}\,\hat{s}_4 + \hat{s}_2\,\hat{s}_3 \\
&+ \hat{s}_2\,\hat{s}_4 - \hat{s}_2 c + \hat{s}_2\,\hat{s}_{13} + \hat{s}_1\,\hat{s}_3 + \hat{s}_1\,\hat{s}_4 - \hat{s}_1 c \\
&+ \hat{s}_1\,\hat{s}_{13}\big)/2c^2
\end{aligned}$$

$$\begin{aligned}
s_3 = s_4 = -\big(&\hat{s}_{13}\,\hat{s}_1 + \hat{s}_{23}\,\hat{s}_1 + \hat{s}_3\,\hat{s}_1 + \hat{s}_4\,\hat{s}_1 - \hat{s}_{23}c - \hat{s}_4 c \\
&- \hat{s}_3 c - \hat{s}_{13}c + \hat{s}_4\,\hat{s}_{23} + \hat{s}_4\,\hat{s}_{13} + \hat{s}_3\,\hat{s}_{23} + \hat{s}_3\,\hat{s}_{13} \\
&+ \hat{s}_4\,\hat{s}_2 + \hat{s}_{23}\,\hat{s}_2 + \hat{s}_{23}^2 + \hat{s}_{13}\,\hat{s}_2 + 2\,\hat{s}_{13}\,\hat{s}_{23} + \hat{s}_{13}^2 \\
&+ \hat{s}_3\,\hat{s}_2/2c^2\big)
\end{aligned}$$

For certain values, the denominator of the second equation, expressing s_{12}, may become 0, and one may wonder what happens to s_{12} in such cases. A simple arithmetic shows that the denominator equals $c^2(\hat{s}_{12} + \hat{s}_0)$. Because both $\hat{s}_{12}$ and $\hat{s}_0$ are non-negative integers, their sum is zero iff both are zero. In particular $\hat{s}_0 = 0$, meaning the input has no constant sites. If we impose the criteria of conservativeness to the data (a technical criteria, corresponding to probabilities of change smaller than 1/2 across all edges), then after a suitable manipulation, we can show that *all* values $\hat{\mathbf{s}} = (\hat{s}_\emptyset, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_4)$ are zero. In other words, this situation can only arise in the trivial and irrelevant case when there are no input data.

Our solution is expressed in terms of the *expected sequence spectrum*, or s_α variables. A more common representation of a tree is by the *edge length spectrum*, or q_α variables. The Hadamard transformation (Hendy and Penny, 1993; Hendy et al., 1994) provides an easy to compute transformation from one representation to the other. For example,

$$\begin{aligned} q_{12} = &-1/2\ \ln(1 - 2\,s_1 - 2\,s_{12} - 2\,s_4 - 2\,s_{13}) \\ &+1/8\ \ln(1 - 4\,s_1 - 4\,s_{13}) + 1/8\ \ln(1 - 4\,s_4 - 4\,s_{13}) \\ &+1/8\ \ln(1 - 4\,s_1 - 4\,s_4) \end{aligned}$$

Thus it is easy to substitute the solutions of the s_α and get the solution in terms of q_α, but writing down the result will be somewhat lengthy and a bit cumbersome.

The question of uniqueness of the ML point for phylogenetic analysis has raised considerable interest in the past (Fukami and Tateno, 1989; Tillier, 1994; Steel, 1994; Chor et al., 2000). It is now known that even four taxa ML trees exhibit datasets giving rise to multiple ML points (Steel, 1994; Chor et al., 2000). In contrast, our result implies the following:

Corollary 2: *Each molecular clock fork topology has a unique local and global maximum likelihood point.*

It is natural to ask which of the three forks yields the best ML result. The way to do it is to compute analytically the ML solution for the three forks ((12)(34)), ((13)(24)), ((14)(23)), and then substitute in the likelihood function. It would be desirable to make this choice by *direct* analytical methods, without going through the computation of the likelihood. Unfortunately, at this stage of the research we do not have the capabilities to do that.

Concluding Remarks

By applying novel algebraic techniques, we extended the state of the art in analytic solutions of ML to four taxa forks under molecular clock in the two-state model. We believe this is a significant step in this area of research. We remark that the use of Hadamard conjugation in this context seems necessary. Not only does it allow variable elimination, but the resulting system is substantially simplified. Alternatively, it is possible to formulate the (12)(34) molecular clock fork as a system with only three variables, q_1 for the two pendant edges of taxa 1 and 2, q_3 for 3 and 4, and q_{12} for the central edge (see Fig. 3). However, this formulation yields a polynomial system of total degree 12, with up to 550 monomials per equation. This should be contrasted with the system of total degree 4 and up to 63 monomials that we get for the molecular clock fork. It is this difference that enabled us to bring the problem to that degree of complexity that is solvable, when employing symbolic algebra tools.

The next step in this line of research is to consider the other molecular clock four taxa topology—the so called comb. In this case we "lose" one of the simple equalities $q_i = q_j$, and have a more complicated identity instead. This small change makes the resulting system substantially harder to solve. An analytic solution for this more involved case was obtained recently (Chor et al., 2003). The solution, however, is not given by a closed form, but rather as the root of a degree 9 univariate polynomial. Other directions are to remove the molecular clock assumption, a step that would make the resulting system much harder to solve. Another related direction is to consider *four-state* character, initially for three taxa under molecular clock.

Even in cases where it is feasible to derive them, analytical solutions will most probably *not* replace numeric approaches in ML based tree reconstruction packages. But the analytic solutions do reveal properties of the maximum likelihood points that are not obtainable numerically. For example, in this work we were able to

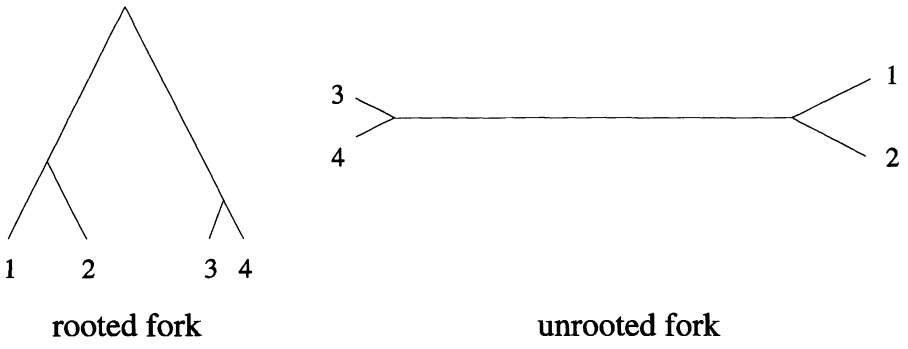

FIGURE 3. Layout of the (12)(34) molecular clock fork (left), and its unrooted version (right).

show that every molecular clock fork tree has a unique (global and local) ML point. Without the molecular clock hypothesis, this uniqueness does not hold, as proved in (Steel, 1994; Chor et al., 2000). In cases where multiple ML points may exist, it is recommended to have multiple starting points when hill climbing procedures are used. Even though this does not guarantee that the global maximum (on a given tree) is found, it improves the chances of finding it. Theoretical studies like the one here may provide a crucial tool in telling if multiple ML points (again, for a given tree) may or may not exist. Furthermore, it is worth noticing that with current knowledge we do not even know if finding the global ML point on a given tree is achievable in polynomial time.

ACKNOWLEDGMENTS

This research was supported by ISF grant 418/00. A preliminary version of these results was presented at the RECOMB03 conference in Berlin (Chor et al., 2003).

We would like to thank Mike Hendy, Mike Steel, and Ziheng Yang for very helpful discussions and for comments on earlier versions of this work. We would also like to thank the associate editor, Nick Goldman, and the two anonymous referees for helpful suggestions, and a very careful inspection and corrections throughout the refereeing process.

REFERENCES

Aho, A. V., Y. Sagiv, T. G. Szymanski, and J. D. Ullman. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. SIAM J. Comput. 10:405–421.

Chor, B., M. Hendy, B. Holland, and D. Penny. 2000. Multiple maxima of likelihood in phylogenetic trees: An analytic approach. MBE 17:1529–1541. Earlier version appeared in RECOMB 2000.

Chor, B., M. Hendy, and D. Penny. 2001. Analytic solutions for three taxon mlmc trees with variable rates across sites. In WABI 2001.

Chor, B., A. Khetan, and S. Snir. 2003. Maximum likelihood on four taxa phylogenetic trees: Analystic solutions. Pages 76–83 in Proceedig of the seventh annual international conference on computational molecular biology (RECOMB), Berlin, Germany.

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

Fukami, K., and Y. Tateno. 1989. On the uniqueness of the maximum likelihood method for estimating molecular trees: Uniqueness of the likelihood point. J. Mol. Evol. 28:460–464.

Hendy, M. D., and D. Penny. 1993. Spectral analysis of jphylogenetic data. J. Classif. 10:5–24.

Hendy, M. D., D. Penny, and M. A. Steel. 1994. Discrete fourier analysis for evolutionary trees. Proc. Natl. Acad. Sci. USA. 91:3339–3343.

Neymann, J. 1971. Molecular studies of evolution: A source of novel statistical problem. Pages 1–27 in Statistical devision theory and related topics (S. Gupta and Y. Jackel, eds.). Academic Press, New York.

Steel, M. 1994. The maximum likelihood point for a phylogenetic tree is ot unique. Syst. Biol. 43:560–564.

Strang, G. 1988. Linear algebra and its applications. Thomson Learning, USA.

Tillier, E. R. M. 1994. Maximum likelihood with multiparameter models of substitution. J. Mol. Evol. 39:409–417.

Yang, Z. 2000. Complexity of the simplest phylogenetic estimation problem. Proc. R. Soc. Lond. B 267:109–119.

First submitted 19 November 2003; reviews returned 8 February 2004; final acceptance 15 July 2004

Associate Editor: Nick Goldman