JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 13, Number 3, 2006

Pp. 819–837

Maximum Likelihood Molecular Clock Comb: Analytic Solutions[1]

BENNY CHOR,[2] AMIT KHETAN,[3] and SAGI SNIR[4]

ABSTRACT

Maximum likelihood (ML) is increasingly used as an optimality criterion for selecting evolutionary trees, 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. 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). In a previous work, we devised a closed form analytic solution for the ML molecular clock fork. In this work, we extend the state of the art in the area of analytic solutions ML trees to the family of all four taxa trees under the molecular clock assumption. The change from the fork topology to the comb incurs a major increase in the complexity of the underlying algebraic system and requires novel techniques and approaches. We combine the ultrametric properties of molecular clock trees with the Hadamard conjugation to derive a number of topology dependent identities. Employing these identities, we substantially simplify the system of polynomial equations. We finally use tools from algebraic geometry (e.g., Gröbner bases, ideal saturation, resultants) and employ symbolic algebra software to obtain analytic solutions for the comb. We show that in contrast to the fork, the comb has no closed form solutions (expressed by radicals in the input data). In general, four taxa trees can have multiple ML points. In contrast, we can now prove that under the molecular clock assumption, the comb has a *unique* (local and global) ML point. (Such uniqueness was previously shown for the fork.)

Key words: maximum likelihood, phylogenetic trees, analytic solutions, Hadamard conjugation, symbolic manipulation.

1. INTRODUCTION

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

[1]A preliminary version of the results in this paper appeared in *RECOMB 2003* (Chor *et al.*, 2003).
[2]School of Computer Science, Tel-Aviv University, Tel-Aviv 39040 Israel.
[3]Department of Mathematics, University of Massachusetts at Amherst, MA.
[4]Mathematics Department, University of California, Berkeley, CA 94720.
Research supported by ISF grant 418/00.

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 are computationally intensive, but 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 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.* (1994) and Hendy and Penny (1993), together with convexity arguments. The first extension from three to four taxa was made by Chor and Snir (2004). There are two families of rooted topologies for four taxa: topologies with two taxa in each subtree of the root, which we call *fork* topologies, and topologies where one subtree of the root has three taxa, which we call *comb* topologies. Under molecular clock, the distance from each of the four leaves to the root is the same (Fig. 1). Chor and Snir (2004) showed a closed form analytical solution for the fork topology under the molecular clock assumption. The fork topology is somewhat simpler than the fork because it has two pairs of equal length edges, while the comb has only one pair of equal length edges.

In this work, we retain the substitution model (symmetric two states) used for the fork (as well as in Yang [2000] and Chor *et al.* [2001]) and handle the more involved topology—the comb. This extension incurs a major increase in the complexity of the underlying system of polynomial equations and requires novel techniques and approaches. Our starting point, like Chor *et al.* (2000) and Chor and Snir (2004), is to formulate the ML problem as one of constrained optimization, and express it in terms of Lagrange coefficients. We 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 polynomials in nine variables.

It turns out that the seemingly subtle difference between the fork and the comb has significant consequences. For the comb, the interrelationships between the variables are more complicated, and consequently the resulting system of equations is not solvable by elementary algebraic tools alone. We employ tools from algebraic geometry (e.g., Gröbner bases, ideal saturation, resultants), together with the computer algebra package Maple. We do get analytical solutions to the comb, expressed as roots of a degree 9 univariant polynomial, whose coefficients are polynomials in the input parameters. In this case, a closed form solution (expressed as radicals in the input parameters) does not exist. However, we are able to show that only one of nine complex roots of the likelihood equations corresponds to a realistic comb tree.

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, combining the results obtained in this work with the closed form solution to the fork, we show that every molecular clock 4-tree has a unique (global and local) ML point. Without the molecular clock hypothesis, this uniqueness does not hold, as proved by Steel (1994) and Chor *et al.* (2000).

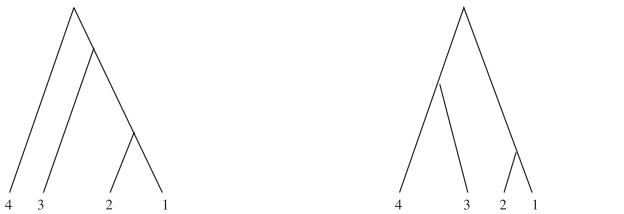

FIG. 1. The fork and comb—two rooted topologies on four taxa.

The remainder of this work is organized as follows: Section 2 introduces definitions and notations and briefly explains the Hadamard conjugation (Hendy and Penny, 1993; Hendy *et al.*, 1994) and its relation to maximum likelihood. In Section 3, we show some technical properties on general trees and particular properties of the molecular clock four taxa trees. In Section 4, we derive the analytical solution to the molecular clock comb. In Section 5, uniqueness of the solution obtained is shown, and in Section 6, we give some concluding remarks and directions for further research.

2. DEFINITIONS, NOTATIONS, AND THE HADAMARD CONJUGATION

In this section, we define the model of substitution we use, introduce useful notations, and briefly describe the Hadamard conjugation.

2.1. Definitions and notations

We start with a tree-labeling notation that will be useful for the rest of the work. For simplicity we use four taxa, but the definitions extend to any n. A *split* of the species is any partition of $\{1, 2, 3, 4\}$ into two disjoint subsets. We will identify each split by the subset which does not contain 4 (in general n), so that, for example, the split $\{\{1, 2\}, \{3, 4\}\}$ is identified by the subset $\{1, 2\}$. For brevity, to label objects subscribed by a split, we concatenate the members of the split. Each edge e of a phylogenetic tree T induces a split of the taxa, i.e., the cut induced by removing e. We denote the edge e by the cut it induces. For instance, the central edge of the tree $T = (12)(34)$ induces the split $\{\{1, 2\}, \{3, 4\}\}$, that is identified by the subset $\{1, 2\}$, and therefore this edge is denoted e_{12}. Thus, $E(T') = \{e_1, e_2, e_{12}, e_3, e_{123}\}$ (see Fig. 2).

In the Neymann (1971) two-state model, each character at a species admits one out of two states, without loss of generality $\{\mathtt{x}, \mathtt{y}\}$. Hence, a character evolving along an evolutionary tree T with n leaves induces a split pattern between the leaves admitting the state $\mathtt{x}$ and $\mathtt{y}$.

In the two-state model, the length of an edge q_e, $e \in E(T)$, in the tree T is defined as the expected number of substitutions (changes) per site along that edge. Given the edge lengths of T: $\mathbf{q} = [q_e]_{e \in E(T)}$ $(0 \leq q_e < \infty)$, the probability of generating an α-split pattern ($\alpha \subseteq \{1, \ldots, n-1\}$) is well defined. Denote this probability by $s_\alpha = Pr(\alpha\text{-split}|T, \mathbf{q})$. Using the same indexing scheme as above, we define the *expected sequence spectrum* $\mathbf{s} = [s_\alpha]_{\alpha \subseteq \{1,\ldots,n-1\}}$.

The *edges lengths* spectrum of a tree T with n leaves is the 2^{n-1} dimensional vector $\mathbf{q} = [q_\alpha]_{\alpha \subseteq \{1,\ldots,n-1\}}$, defined for any subset $\alpha \subseteq \{1, \ldots, n-1\}$ by

$$q_\alpha = \begin{cases} q_e & \text{if } e \in E(T) \text{ induces the split } \alpha, \\ -\sum_{e \in E(T)} q_e & \text{if } \alpha = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

2.2. Hadamard conjugation

The Hadamard conjugation (Hendy and Penny, 1993; Hendy *et al.*, 1994) is an invertible transformation that specifies a relation between the expected sequence spectrum $\mathbf{s}$ and the edge lengths spectrum $\mathbf{q}$

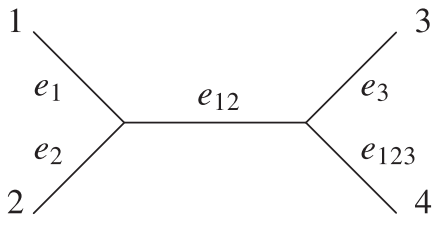

FIG. 2. The tree $T' = (12)(34)$ and its edges.

of the tree. In other words, the transformation links the probabilities of site substitutions on edges of an evolutionary tree T to the probabilities of obtaining each possible combination of characters. The Hadamard conjugation is applicable to a number of site substitution models: Neyman two-state model, Jukes–Cantor (Jukes and Cantor, 1969) model, and Kimura (1983) 2ST and 3ST models (the last two are applicable to "normal," four-state DNA). For these models, the transformation yields a powerful tool which greatly simplifies and unifies the analysis of phylogenetic data, and in particular the analytical approach to ML.

Definition 1. *A* Hadamard matrix *of order ℓ is an $\ell \times \ell$ matrix A with ± 1 entries such that $A^t A = \ell I_\ell$.*

We will use a special family of Hadamard matrices, called Sylvester matrices in MacWilliams and Sloan (1977, p. 45), defined inductively for $n \geq 0$ by $H_0 = [1]$ and $H_{n+1} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}$. For example,

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

It is convenient to index the rows and columns of H_n by lexicographically ordered subsets of $\{1, \ldots, n\}$. Denote by $h_{\alpha,\gamma}$ the (α, γ) entry of H_n, then $h_{\alpha,\gamma} = (-1)^{|\alpha \cap \gamma|}$. This implies that H_n is symmetric; namely, $H_n^t = H_n$, and thus by the definition of Hadamard matrices $H_n^{-1} = \frac{1}{2^n} H_n$.

Proposition 1 (Hendy and Penny, 1993). *Let T be a phylogenetic tree on n leaves with finite edge lengths ($q_e < \infty$ for all $e \in E(T)$). Assume that sites mutate according to a symmetric substitution model, with equal rates across sites. Let* $\mathbf{s}$ *be the expected sequence spectrum. Then*

$$\mathbf{s} = \mathbf{s}(\mathbf{q}) = H_{n-1}^{-1} \exp(H\mathbf{q}),$$

where the exponentiation function $exp(x) = e^x$ is applied element-wise to the vector $\rho = H\mathbf{q}$. That is, for $\alpha \subseteq \{1, \ldots, n-1\}$, $s_\alpha = 2^{-(n-1)} \sum_\gamma h_{\alpha,\gamma} \left(\exp\left(\sum_\delta h_{\gamma\delta} q_\delta\right)\right)$.

This transformation is called the *Hadamard conjugation*.

Definition 2. *A vector $\hat{\mathbf{s}} \in \mathcal{R}^{2^{n-1}}$ satisfying*

$$\sum_{\alpha \subseteq \{1,\ldots,n-1\}} \hat{\mathbf{s}}_\alpha = 1$$

and $H\hat{\mathbf{s}} > \mathbf{0}$ is called conservative.

For conservative data $\hat{\mathbf{s}}$, the Hadamard conjugation is invertible, yielding

$$\gamma = \gamma(\hat{\mathbf{s}}) = H_{n-1}^{-1} \ln(H\hat{\mathbf{s}})$$

where the ln function is applied element-wise to the vector $H\hat{\mathbf{s}}$. We note that γ is not necessarily the edge length spectrum of any tree. On the other hand, the expected sequence spectrum of any tree T is always conservative.

3. MAXIMUM LIKELIHOOD ON FOUR TAXA TREES

In this section, we describe how the system of equations is set up and how the molecular clock is used to simplify it. We begin with the formulation of the general maximum likelihood problem as a constrained optimization problem and the resulting system of polynomial equations. Then we use the molecular clock structure together with the Hadamard conjugation to derive a number of identities. Using the derived identities, the system is substantially simplified in both cases, to the point where analytic solutions can be derived.

3.1. General ML System

Given an input data ψ of n aligned, two-state sequences, every column in ψ induces a split. Let $\hat{s}_\alpha$ be the number of columns in ψ inducing the split α ($\alpha \subseteq \{1, \ldots, n-1\}$). The vector $\hat{s} = [\hat{s}_\alpha]_{\alpha \subseteq \{1,\ldots,n-1\}}$, indexed analogously to the expected sequence spectrum, is called the *observed sequence spectrum*. The likelihood of producing the observed spectrum $\hat{\mathbf{s}}$ given the expected spectrum $\mathbf{s}$ equals

$$L(\hat{\mathbf{s}}|\mathbf{s}) = \prod_{\alpha \subseteq \{1,\ldots,n-1\}} Pr(\alpha\text{-split} \mid \mathbf{s})^{\hat{s}_\alpha} = \prod_{\hat{s}_\alpha > 0} s_\alpha^{\hat{s}_\alpha}.$$

In the specific case of a four-taxa *unrooted* tree,

$$L(\mathbf{s}|\hat{\mathbf{s}}) = s_\emptyset^{\hat{s}_\emptyset} \cdot s_1^{\hat{s}_1} \cdot s_2^{\hat{s}_2} \cdot s_{12}^{\hat{s}_{12}} \cdot s_3^{\hat{s}_3} \cdot s_{13}^{\hat{s}_{13}} \cdot s_{23}^{\hat{s}_{23}} \cdot s_{123}^{\hat{s}_{123}}.$$

Topologically there is only one unrooted tree on four taxa. Without loss of generality, we describe the system corresponding to the tree in Fig. 2. This tree has e_{12} as its "central" edge. The expected spectrum $\mathbf{s}$ of the tree can be represented as a point in $\mathbb{R}^8$ whose edge lengths satisfy

- $q_\alpha(s) \geq 0$ for all $\alpha \in E(T)$,
- $q_\alpha(s) = 0$ for all $\alpha \notin E(T)$.

Thus, q_{13} and q_{23} must equal zero, and we can formulate the problem of maximizing the likelihood function as a constrained maximization problem: Find the maximum value of L under the two constraints $q_{13}(\mathbf{s}) = 0$ and $q_{23}(\mathbf{s}) = 0$. The approach taken by Chor *et al.* (2000) is to initially express the set of critical points using Lagrange multipliers. By Proposition 1, every q_α is a function of the expected spectrum $\mathbf{s}$, so we seek the point or points where

$$\nabla L = \lambda_1 \nabla q_{13}(\mathbf{s}) + \lambda_2 \nabla q_{23}(\mathbf{s}).$$

This gives rise to a system of ten-degree five-polynomial equations in ten variables: the eight s_α variables, and two additional "Lagrange" variables (λ_1 and λ_2). We emphasize that the eight $\hat{s}_\alpha$ are not variables—they are parameters detemined by the four *input* sequences. (For bvrevity, we will use q_α and not $q_\alpha(\mathbf{s})$ in the sequel.) The resulting system, in its full generality, is still beyond the reach of current computer algebra techniques.

We now make one additional assumption called the "molecular clock." We assign a root to the tree and require that the distances from the root to each of the four leaves is the same. This results in either the fork or the comb topologies described in the Introduction (see Fig. 1). The key to obtaining analytic solutions is to combine Hadamard conjugation with the molecular clock structure and derive a number of "clock" identities applicable to both topologies (fork and comb). Subsequently, we derive comb-specific identities. Using the derived identities, the system is substantially simplified. The molecular clock assumption yields for the comb: $q_1 = q_2$ and $q_1 + q_{12} = q_3$. We emphasize that the system of equations does not take explicitly into account *inequalities* like $q_{12} \geq 0$ or $q_{123} \geq q_{12} + q_1$. The system is hard enough to solve as it is. Of course, the final ML point (solution) should satisfy the relevant inequlities, as otherwise it would not correspond to a "real" phylogenetic tree.

3.2. Simplifying identities

The key to our simplifications is the use of length relations among the edges (the $\mathbf{q}$ variables), which follow from molecular clock, in order to derive identities on the expected spectrum variables (the $\mathbf{s}$ variables). The following relation on the expected spectrum variables is proved by Chor *et al.* (2001).

Theorem 1 (Chor *et al.*, 2001). *Let i and j be sister taxa in a phylogenetic tree T with n leaves and edge weights $\mathbf{q}$. Let $\mathbf{s}$ be the expected spectrum, such that $\mathbf{s} = H^{-1} \ln(H\mathbf{q})$; then $q_i = q_j$ implies $s_i = s_j$, and $q_i > q_j$ implies $s_i > s_j$.*

Under a molecular clock, the four-taxa molecular clock comb has a pair of sister taxa with equal edge lengths on the tree. The next theorem is a generalization of the previous one, yielding one additional identity for the molecular clock comb.

Theorem 2. *Consider a tree T on n leaves, with two sister taxa i and j such that $q_i = q_j$ (see Fig. 3). Let* $\mathbf{s}$ *be the expected spectrum, such that* $\mathbf{s} = H^{-1}\ln(H\mathbf{q})$. *Then for every* $\alpha \subseteq \{1, 2, \ldots, n-1\} \setminus \{i, j\}$, $s_{\alpha\cup\{i\}} = s_{\alpha\cup\{j\}}$.

Proof. The argument is essentially a symmetry argument, saying that i and j are interchangeable. More formally, since i and j are sister taxa in T and $q_i = q_j$, then for every other taxa k in T ($k \neq i, j$) the length of the tree paths from i to k and from j to k is the same. By the definition of the edge spectrum, this implies that for every $\beta \subseteq \{1, 2, \ldots, n-1\} \setminus \{i, j\}$, $q_{\beta\cup\{i\}} = q_{\beta\cup\{j\}}$. This means that any function of the edge spectrum, $\mathbf{q}$, is invariant under interchange of i and j. In particular, $s_{\alpha\cup\{i\}} = s_{\alpha\cup\{j\}}$. ■

The next technical claim deals with conservative points $\mathbf{s} \in \mathbb{R}^8$ (namely, $H\mathbf{s} > 0$) satisfying $s_1 = s_2$, $s_{13} = s_{23}$, and $\sum_{\alpha\subseteq\{1,2,3\}} s_\alpha = 1$. (These points need not be the expected spectrum of a tree.) This technical claim will be useful in simplifying the system of polynomial equations that we solve in Section 4.

Claim 1. *Let* $\mathbf{s} = (s_\emptyset, s_1, s_2, s_{12}, s_3, s_{13}, s_{23}, s_{123}) \in \mathbb{R}^8$ *be a conservative point satisfying* $s_1 = s_2$ *and* $s_{13} = s_{23}$, *and let* $\mathbf{q} = H^{-1}\ln H\mathbf{s}$. *Then* $\mathbf{s}$ *satisfies* $q_{13}(\mathbf{s}) = q_{23}(\mathbf{s})$.

Proof. By the Hadamard conjugation we get

$$\begin{aligned} 4(q_{23} - q_{13}) = &\ln(1 - 2s_1 - 2s_{12} - 2s_{123} - 2s_{13}) \\ &- \ln(1 - 2s_2 - 2s_{12} - 2s_{23} - 2s_{123}) \\ &- \ln(1 - 2s_1 - 2s_{12} - 2s_3 - 2s_{23}) \\ &+ \ln(1 - 2s_2 - 2s_{12} - 2s_3 - 2s_{13}). \end{aligned}$$

But by the assumption $s_1 = s_2$ and $s_{13} = s_{23}$, so $(1 - 2s_1 - 2s_{12} - 2s_{123} - 2s_{13})$ (the first term) equals $(1 - 2s_2 - 2s_{12} - 2s_{23} - 2s_{123})$ (the second term), and $(1 - 2s_1 - 2s_{12} - 2s_3 - 2s_{23})$ (third term) equals $(1 - 2s_2 - 2s_{12} - 2s_3 - 2s_{13})$ (fourth term). This implies $q_{13}(\mathbf{s}) - q_{23}(\mathbf{s}) = 0$. ■

The following two statements refer to the molecular clock comb.

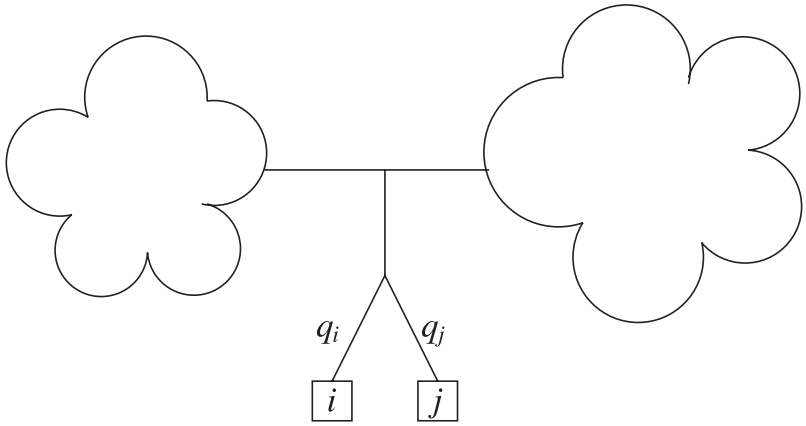

FIG. 3. A general tree with two sister taxa i and j s.t. $q_i = q_j$.

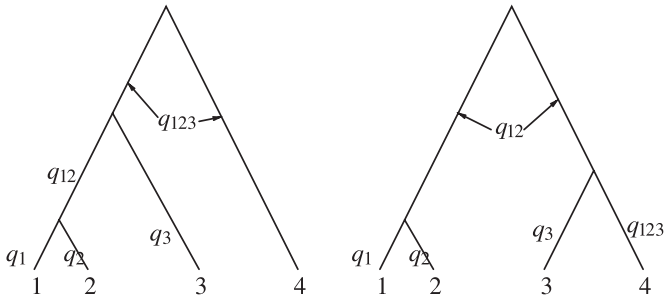

FIG. 4. Trees with edges length example.

Claim 2. *Consider the* (((12)3)4) *molecular clock comb. Then* $s_{13} = s_1 + s_{12} - s_3$.

Proof. By the Hadamard conjugation we get

$$s_1 + s_{12} - s_3 - s_{13} = -\frac{1}{4}e^{(-2q_2-2q_{12}-2q_{23}-2q_{123})} - \frac{1}{4}e^{(-2q_1-2q_{12}-2q_{13}-2q_{123})} + \frac{1}{2}e^{(-2q_3-2q_{13}-2q_{23}-2q_{123})} + \frac{1}{4}e^{(-2q_2-2q_{12}-2q_3-2q_{13})} - \frac{1}{4}e^{(-2q_1-2q_{12}-2q_3-2q_{23})}.$$

Substituting $q_{13} = q_{23} = 0$ and the (((12)3)4) molecular clock comb identities $q_1 = q_2$ and $q_3 = q_1 + q_{12}$ completes the proof. ■

Claim 3. *Let* T *be a* (((1, 2)3)4) *molecular clock comb. Then*

$$-1 + 4s_1 + 2s_3 + 2s_{123} = (-1 + 4s_1 + 4s_{12} - 2s_3 + 2s_{123})(-1 + 8s_1 + 4s_{12} - 4s_3).$$

Proof. By the Hadamard Conjugation we get

$$\begin{aligned} q_{13} = {} & -1/8\ln(1 - 2s1 - 2s12 - 2s13 - 2s123) + 1/8\ln(1 - 2s2 - 2s12 - 2s23 - 2s123) \\ & - 1/8\ln(1 - 2s1 - 2s2 - 2s13 - 2s23) - 1/8\ln(1 - 2s3 - 2s13 - 2s23 - 2s123) \\ & + 1/8\ln(1 - 2s1 - 2s12 - 2s3 - 2s23) - 1/8\ln(1 - 2s2 - 2s12 - 2s3 - 2s13) \\ & + 1/8\ln(1 - 2s1 - 2s2 - 2s3 - 2s123). \end{aligned}$$

By setting $s_2 = s_1$ as in Theorem 2 and $s_{13} = s_1 + s_{12} - s_3$ by Claim 2, we get

$$\begin{aligned} q_{13} = {} & 1/8\ln(1 - 4s1 - 4s12 + 2s3 - 2s123) - 1/8\ln(1 - 8s1 - 4s12 + 4s3) \\ & + 1/8\ln(1 - 4s1 - 2s3 - 2s123). \end{aligned}$$

By equating q_{13} to zero, taking the exponent and multiplying by the denominator, the result follows. ■

4. SOLVING THE MOLECULAR CLOCK COMB

In this section, we explain the process leading to the analytical solution of the molecular clock comb.

4.1. Overview

We begin by looking for the point or points satisfying $\nabla L = \lambda_1 \nabla q_{13}(\mathbf{s}) + \lambda_2 \nabla q_{23}(\mathbf{s})$. We apply the "comb identities" derived in Subsection 3.2. A change of the variables and of the input parameters leads to a somewhat simplified system, where the two constraints $q_{12} = 0$, $q_{13} = 0$ are already satisfied. Having no constraints to satisfy, we now look for point(s) satisfying $\nabla L = 0$. This later system has three free variables and three partial derivatives. Unfortunately, this system involves rational functions, and not just polynomials. The obvious route to tackle the problem is to equate the numerators of these rational functions to zero. There are two drawbacks to this approach: It yields high degree polynomials (total degree 5) and introduces infinitely many "spurious zeroes" of ∇L. We now apply a mixture of algebraic geometry techniques (saturation of a module, resultants) and heavy application of computer algebra software (Maple, Macaulay 2). With all these, we discover three simpler polynomials (total degree 3) in three variables with only a few spurious zeroes that we can easily eliminate.

One of these polynomials is linear in one of the variables, so we can express this variable as a polynomial in the two remaining variables (of total degree 2). We are left with two polynomials in two variables, t_1 and t_2. Viewed as polynomials in t_1, the degree of both polynomials is just 3. To find the common zeroes of these two polynomials, we compute their *resultant* with respect to t_1. After simplification, we get a single univariate polynomial in t_2, of degree 9, whose coefficients (there are 1,457 of those) are polynomials of total degree 5 in the (given) input parameters. The Galois group of this polynomial is in general the whole symmetric group S_9, so it is a "generic" degree-nine polynomial which cannot be solved by radicals. This means that we cannot get closed form solutions (expressed with radicals of the input parameters). To solve any particular instance, a univariate polynomial root finder (such as **fsolve** in Maple), is employed. All of the s_α are then readily obtainable from the solutions to this one equation. In principle, the degree-nine equation could yield up to nine solutions to the ML problem. However, in the next section, we show that the ML point in the valid region ($0 \le s_\alpha \le 1$) is unique.

4.2. Obtaining the polynomial system

Our initial likelihood function is a function of the eight s_α variables

$$L(\mathbf{s}|\hat{\mathbf{s}}) = s_{\emptyset}^{\hat{s}_{\emptyset}} \cdot s_1^{\hat{s}_1} \cdot s_2^{\hat{s}_2} \cdot s_{12}^{\hat{s}_{12}} \cdot s_3^{\hat{s}_3} \cdot s_{13}^{\hat{s}_{13}} \cdot s_{23}^{\hat{s}_{23}} \cdot s_{123}^{\hat{s}_{123}}.$$

As was proved in the previous section, an ML molecular clock solution for the (((12)3)4) comb satisfies the following constraints.

$$s_0 = 1 - s_1 - s_2 - s_3 - s_{12} - s_{13} - s_{23} - s_{123} \tag{1}$$

$$s_2 = s_1 \tag{2}$$

$$s_{13} = s_1 + s_{12} - s_3 \tag{3}$$

$$s_{23} = s_{13} \tag{4}$$

$$-1 + 4s_1 + 2s_3 + 2s_{123} = (-1 + 4s_1 + 4s_{12} - 2s_3 + 2s_{123})(-1 + 8s_1 + 4s_{12} - 4s_3) \tag{5}$$

We now employ a linear invertible change of variables.

$$t_0 = s_0 = 1 - 4s_1 + s_3 - 3s_{12} - s_{123}$$

$$t_1 = 1 - 8s_1 - 4s_{12} + 4s_3$$

$$t_2 = 1 - 4s_1 - 4s_{12} + 2s_3 - 2s_{123}$$

$$t_3 = 1 - 4s_1 - 2s_3 - 2s_{123}$$

After this change of variables, the quadratic constraint (5) can be written as $t_3 = t_1 t_2$. This change of variables is indeed invertible, and we can write the s_α variables appearing in the last four equalities in

terms of the t_β variables. The other s_α variables can be obtained by simple substitutions in (2)–(4). We can eliminate t_3 by replacing it with $t_1 t_2$:

$$s_1 = \frac{-t_1 t_2 + 1 - t_1 + t_2}{8},$$

$$s_3 = \frac{t_1}{4} + \frac{1}{4} + \frac{t_2}{2} - t_0,$$

$$s_{12} = -t_0 + \frac{1}{4} + \frac{t_1}{4} + \frac{t_2}{4} + \frac{t_1 t_2}{4},$$

$$s_{123} = -\frac{t_1 t_2}{4} - \frac{3t_2}{4} + t_0.$$

Under this change of variables, we can remove all the constraints and reduce the problem to maximizing a function of three free variables, t_0, t_1, t_2. To simplify the presentation, we rewrite the parameter space by setting

$$g_0 = c - \hat{s}_1 - \hat{s}_2 - \hat{s}_3 - \hat{s}_{12} - \hat{s}_{13} - \hat{s}_2 3 - \hat{s}_{123},$$

$$g_1 = \hat{s}_1 + \hat{s}_2,$$

$$g_{12} = \hat{s}_{12},$$

$$g_{13} = \hat{s}_{13} + \hat{s}_{23},$$

$$g_3 = \hat{s}_3,$$

$$g_{123} = \hat{s}_{123},$$

$$c = \sum \hat{s}_\alpha.$$

The resulting likelihood function is now

$$\begin{aligned} L(t) = \ln \Bigg(& t_0^{g_0} \left(-\frac{1}{8} t_1 t_2 + \frac{1}{8} - \frac{1}{8} t_1 + \frac{1}{8} t_2 \right)^{g_1} \\ & \left(-t_0 + \frac{1}{4} + \frac{1}{4} t_2 + \frac{1}{4} t_1 t_2 + \frac{1}{4} t_1 \right)^{g_{12}} \left(\frac{1}{8} t_1 t_2 + \frac{1}{8} - \frac{1}{8} t_1 - \frac{1}{8} t_2 \right)^{g_{13}} \\ & \left(\frac{1}{4} t_1 + \frac{1}{4} + \frac{1}{2} t_2 - t_0 \right)^{g_3} \left(-\frac{3}{4} t_2 + t_0 - \frac{1}{4} t_1 t_2 \right)^{g_{123}} \Bigg). \end{aligned} \tag{6}$$

To maximize $L(t)$, we look for points satisfying $\nabla L = 0$, which amounts to three equations involving rational functions:

$$\frac{\partial L}{\partial t_0} = 0, \quad \frac{\partial L}{\partial t_1} = 0, \quad \frac{\partial L}{\partial t_2} = 0.$$

The left-hand side of the resulting three equations, which we name eq_0, eq_1, and eq_2, are the following.

$$eq_0 = \frac{g_0}{t_0} - \frac{4 g_3}{t_1 + 1 + 2 t_2 - 4 t_0} - \frac{4 g_{12}}{-4 t_0 + 1 + t_2 + t_2 t_1 + t_1} - \frac{4 g_{123}}{3 t_2 - 4 t_0 + t_2 t_1}$$

$$eq_1 = \frac{g_3}{t_1 + 1 + 2 t_2 - 4 t_0} + \frac{(1 + t_2) g_{12}}{-4 t_0 + 1 + t_2 + t_2 t_1 + t_1} + \frac{g_{123} t_2}{3 t_2 - 4 t_0 + t_2 t_1} + \frac{g_{13}}{t_1 - 1} + \frac{g_1}{t_1 - 1}$$

$$eq_2 = 2 \frac{g_3}{t_1 + 1 + 2 t_2 - 4 t_0} + \frac{(1 + t_1) g_{12}}{-4 t_0 + 1 + t_2 + t_2 t_1 + t_1} + \frac{(3 + t_1) g_{123}}{3 t_2 - 4 t_0 + t_2 t_1} + \frac{g_{13}}{t_2 - 1} + \frac{g_1}{1 + t_2}$$

In order to apply the techniques of symbolic algebra, we need a system of *polynomial* equations rather than *rational* ones. The simplest way to obtain a polynomial system would be to set the three numerators equal to zero. However, this gives equations of high complexity and of degree 5 in the variables t_0, t_1, t_2. Moreover, these equations have infinitely many zeroes for any values of the parameters g_i. For example, setting $t_0 = 0$ and $t_2 = -1$ causes the numerators to vanish for any choice of t_1.

A better strategy is to use the algebraic technique of *saturation*. We are interested in simple polynomials in the variables t_0, t_1, t_2 and the parameters $g_0, g_1, g_3, g_{12}, g_{13}, g_{123}$ that are zero on any critical point of L, but are *not zero* on data points for which some denominator of eq_i is zero (the denominators of eq_i are polynomials of the t_i alone and therefore independent of the input g_i). The equations eq_i are linear in the parameters $g_0, g_1, g_3, g_{12}, g_{13}, g_{123}$; therefore, it makes sense to restrict ourselves to looking for polynomials that are also linear in these parameters.

A polynomial that is linear in the six parameters can be represented as a vector of length 6 of polynomials in t_0, t_1, t_2 only, corresponding to the coefficients of each of the parameters. The set of all such vectors is a *free module* over the polynomial ring in t_0, t_1, and t_2. The numerators $eqnum_0, eqnum_1, eqnum_2$ define a submodule. The *saturation* of this module by the product of the denominators yields those polynomials which vanish at any common zero of $eqnum_0, eqnum_1, eqnum_2$ where none of the denominators are zero. This operation can be carried out easily in the computer algebra package Macaulay 2. The commands allow one to define the appropriate submodule of a rank 6 free modules over a polynomial ring in three variables. The command "saturate" does the rest and yields three polynomials m_0, m_1, m_2 (as well as four other polynomials which we do not use). The result of our computation is the following proposition.

Proposition 2. *The ML points of the molecular clock comb are among the zeroes of the following polynomial system.*

$$\begin{aligned}
m_0 &= c(t_1^2 - 2t_2t_1 + 4t_2^2 + 8t_0 - 6t_2 - 5) \\
&\quad + g_1(-8t_0 + 4t_1 + 2t_2t_1 + 2t_2 + 8) \\
&\quad + g_3(2 + 4t_2 + 2t_1) + 2g_{12}(1 + t_1)(1 + t_2) \\
&\quad + g_{13}(-8t_0 + 2t_2t_1 + 10t_2 + 4t_1 + 8) + 2g_{123}t_2(3 + t_1) \\
m_1 &= c(-2t_2^2 - 3t_2t_1 - 3t_2 + 4t_2t_0 - 1 + 4t_1t_0 - t_1 + 2t_2^3) \\
&\quad + g_1(t_2^2 + t_2^2t_1 + 4t_0 + 2t_2t_1 - 4t_2t_0 + 2t_2 + t_1 + 1) \\
&\quad + g_3(2t_2^2 + 3t_2t_1 + t_2 + 1 + t_1) + g_{12}(1 + t_2)^2(1 + t_1) \\
&\quad + g_{13}(4t_0 + t_2^2t_1 + 5t_2^2 - 4t_2t_0 + 4t_2 + 4t_2t_1 + t_1 + 1) \\
&\quad + g_{123}t_2(3t_2 + t_2t_1 + 3t_1 + 1) \\
m_2 &= 2c(1 + t_2)(-t_2t_1 - t_2 + 2t_2t_0 - t_0 + t_1t_0) \\
&\quad + g_1(2t_2 + 4t_0 - 4t_2t_0 + 2t_2^2t_1 + 2t_2^2 + 2t_2t_1) \\
&\quad + 2g_3t_2(1 + t_1)(1 + t_2) + 2g_{12}t_2(1 + t_1)(1 + t_2) \\
&\quad + 2g_{13}(1 + t_2)(t_2 + t_2t_1 + 2t_0) + 2g_{123}t_2(1 + t_1)(1 + t_2)
\end{aligned}$$

Proof. We verify, using Maple, the following identities.

$$\begin{aligned}
m_0 &= [(t_1^2 - 2t_2t_1 + 4t_2^2 + 8t_0 - 6t_2 - 5)t_0]eq_0 \\
&\quad + [(t_1 - 1)(3 + t_1)(1 + t_1)]eq_1 + [4(t_2 - 1)(1 + t_2)t_2]eq_2 \\
m_1 &= [(-2t_2^2 - 3t_2t_1 - 3t_2 + 4t_2t_0 - 1 + 4t_1t_0 - t_1 + 2t_2^3)t_0]eq_0 \\
&\quad + [4t_0(t_1 - 1)(1 + t_1)]eq_1 + [(t_2 - 1)(1 + t_2)(t_1 + 1 + 2t_2)t_2]eq_2 \\
m_2 &= [2(1 + t_2)(-t_2t_1 - t_2 + 2t_2t_0 - t_0 + t_1t_0)t_0]eq_0 \\
&\quad + [2t_0(t_1 - 1)(1 + t_1)(1 + t_2)]eq_1 + [4(t_2 - 1)(1 + t_2)t_2t_0]eq_2
\end{aligned}$$

Therefore, the vanishing of ∇L implies the vanishing of m_0, m_1, m_2. ■

4.3. Solving the polynomial system

In this subsection, we explicitly solve for the critical points of the likelihood function L. This is done by solving the system of the polynomials $\{m_0, m_1, m_2\}$ above and removing the spurious solutions. We will call a choice of parameters *positive* if all of the g_α are positive. The g_α represent the data for our likelihood problem and so will always be nonnegative. In phylogenetically realistic data for the model, none of the g_α will be zero.

The goal will be the following theorem.

Theorem 3. *For positive data, the likelihood function L in Equation (6) has at most nine critical points. Moreover, these critical points are obtained from the zeroes of an explicit degree-nine polynomial in one variable t_2, from which we derive the other two variables t_0 and t_1.*

The proof/construction has two stages. First we solve for the critical points for almost all ("generic") choices of data g_α. We make use of the following simple fact.

Proposition 3. *Given any subset S of $\mathbb{R}^n$ defined as the zero set of a collection of polynomials (an algebraic set), either S is all of $\mathbb{R}^n$ or has strictly smaller dimension. In particular, if there is a single point p not in S, then almost all points $p \in \mathbb{R}^n$ are not in S.*

We will often use the word "generic" to refer to any data point outside some given algebraic set that is a strict subset of the whole space. So, we first prove the theorem for cases when the data is generic. In the second stage, more sophisticated algebraic geometry is invoked to argue that the proof holds for all positive data points and not just generic ones.

Consider the equations m_0, m_1, m_2. These are of low degree, as m_0 is of total degree 2, while m_1 and m_2 are of total degree 3. Furthermore, t_0 appears *linearly* in m_0, so it is possible to eliminate it.

Proposition 4. *Any solution of $m_0 = 0$ satisfies*

$$\begin{aligned}
t_0 = \frac{1}{8(g_1 + g_{13} - c)}&(6g_{123}t_2 + 2g_{123}t_2t_1 + 2g_1t_2 + 4g_1t_1 + 2g_1t_2t_1 + 8g_1 + 2g_{12} \\
&+ 2g_{12}t_2 + 2g_{12}t_1 + 2g_{12}t_2t_1 + 2g_3 + 4g_3t_2 + 2g_3t_1 + 8g_{13} + 2g_{13}t_2t_1 + 10g_{13}t_2 \\
&+ 4g_{13}t_1 + ct_1^2 - 2ct_2t_1 + 4ct_2^2 - 6ct_2 - 5c). \qquad (7)
\end{aligned}$$

Proof. This is an immediate consequence of $m_0 = 0$. ■

Substituting for t_0 in equations m_1 and m_2 leaves us with two polynomials in t_1 and t_2, which we call n_1 and n_2. The two polynomials are of degree 3 when viewed as polynomials in just t_1.

$$n_1(t_1, t_2) = a_0(t_2) + a_1(t_2)t_1 + a_2(t_2)t_1^2 + a_3(t_2)t_1^3$$

$$n_2(t_1, t_2) = b_0(t_2) + b_1(t_2)t_1 + b_2(t_2)t_1^2 + b_3(t_2)t_1^3$$

With respect to t_2, n_1 is of degree 2 and n_2 is of degree 4.

We shall see that this system has only finitely many solutions, all of which we can find. The major tool used here is the *resultant* of two polynomials. Let $f(x) = \sum_{i=0}^{d} a_i x^i$ and $g(x) = \sum_{j=0}^{d} b_j x^j$ be two polynomials in one variable x. The resultant of f and g, denoted $Res(f, g, x)$, is a polynomial in the coefficients a_i and b_j of f and g which is 0 whenever f and g have a common zero. The coefficients can themselves be unknowns, or functions of other variables, in which case the resultant replaces the two polynomials f and g by a single polynomial in one fewer variables. Moreover, the common zero of f and g can be written explicitly in terms of the partial derivatives of the resultant with respect to certain coefficients.

Theorem 4 (Gelfand *et al.*, 1994). *Let $R(a, b) = Res(f, g, x)$. If the coefficients a, b are such that f and g have a common zero, then it can be obtained by the ratio of two polynomials in the coefficients a and b:*

$$x = \frac{\partial R}{\partial a_1} \Big/ \frac{\partial R}{\partial a_0}$$

provided the denominator is nonzero, which is true if and only if the common zero is unique without multiplicity.

So we compute the resultant $R = Res(n_1, n_2, t_1)$ of n_1 and n_2 with respect to t_1, a single polynomial R in just t_2 and the parameters. Using Maple, we obtain the following.

Proposition 5. *The resultant R factors as*

$$Res(n_1, n_2, t_1) = 128c^4(g_1 + g_{13} - c)^2 t_2(-t_2^2 c + (g_1 - g_{13})t_2 + (c - g_1 - g_{13}))Elim(t_2),$$

where $Elim(t_2)$ is a polynomial of degree 9 in t_2 and degree 5 in the parameters and has 1,457 terms.

Corollary 5. *There is an explicit rational expression*

$$t_1 = \frac{p(g, t_2)}{q(g, t_2)}, \tag{8}$$

where p, q are polynomials in the parameters g_α and t_2. So, for almost all choices of parameters, we can solve for t_1 uniquely in terms of t_2; hence, no two critical points share the same t_2 variable.

Proof. Using the resultant, $R(t_2) = Res(n_1, n_2, t_1)$, Theorem 4 above yields the desired rational expression for t_1 as the quotient of two polynomials in t_2. This expression is valid as long as the denominator $q(g, t_2)$ is nonzero. The set of parameters for which both $R(t_2)$ and $q(t_2)$ have a common zero is an algebraic set. It is easy to plug in a random choice for the g_α to see that it is not the whole space of parameters. It follows from Proposition 3 that the denominator $q(g, t_2)$ is nonzero for t_2 and a zero of $R(t_2)$ and almost all choices of parameters g. ■

Obviously, we could also have constructed the resultant of n_1 and n_2 with respect to t_2, $Res(n_1, n_2, t_2)$. This would yield a polynomial in the one variable t_1 instead of t_2. This polynomial turned out to be of degree 9 too, but we found that it is of degree 7 in the parameters and has 3,891 terms. So it is somewhat more efficient to work with the elimination polynomial for t_2.

Proposition 6. *For generic choices of parameters, the t_2 coordinate of any critical point of L is among the roots of $Elim(t_2)$. It follows that for generic input parameters there are at most nine critical points of L.*

Proof. The zeros of the resultant correspond to common zeroes of n_1 and n_2, and therefore of m_0, m_1, and m_2. For a positive choice of parameters $c \neq 0$ and $g_1 + g_{13} - c \neq 0$. There are two other factors: The first is just $t_2 = 0$. The second factor is $-t_2^2 c + (g_1 - g_{13})t_2 + (c - g_1 - g_{13})$. Equating the last expression to zero, substituting into the two polynomial equations $n_1(t_1, t_2) = 0$ and $n_2(t_1, t_2) = 0$, and applying a Maple computation, we get $t_1 = -1$. Again the set of parameters where critical points have $t_2 = 0$ or $t_1 = -1$ is an algebraic set. A choice of parameters can be easily found which, along with $t_2 = 0$ or $t_1 = -1$, is not a critical point of L. This implies that this algebraic set is not the whole space, and therefore for almost all data the critical points satisfy $Elim(t_2) = 0$ which has at most nine zeroes. By Corollary 5 we can generically solve for t_1 uniquely in terms of t_2, and by Proposition 4 we can solve for t_0. Thus, there are generically at most nine critical points. ■

This leads to the following algorithm to solve for the ML point given the parameters: Use a root finder (**fsolve** in Maple) to find the nine real roots of $Elim(t_2)$. Here we must resort to numerical root finding and not a closed form solution because for a random choice of parameters the Galois group of $Elim(t_2)$ is the whole symmetric group S_9 (checked in Maple). This means that it is a "general" degree-nine polynomial. As such, it cannot be solved by radicals for this particular choice of data, and so of course it cannot be solved by radicals in general. For each of the nine t_2 roots, we do the following.

1. Solve for t_1 in terms of t_2 using the rational expression derived from the resultant in Corollary 5.
2. Solve for t_0 in terms of t_1 and t_2 using the formula in Proposition 4.
3. Back substitute to obtain the values for **s**.
4. If this is a legal solution (all s_α are in the interval (0,1)), then output it.
5. Repeat until all nine roots are examined.

In all examples we have worked out, the polynomial in question has nine real roots. The "middle" root, i.e., the fifth largest, always gives the unique critical point where the corresponding s values are all between 0 and 1. In the next section, we show that indeed there are always nine real solutions and a unique statistically valid (i.e., where all **s** values are nonnegative) ML point.

So far we have shown that there are at most nine critical points for generic data. However, we have not yet ruled out the existence of unlucky datasets for which there are more than nine critical points. Such bad datasets might yield exceptions to the uniqueness of the ML point which we wish to prove in the next section. First, we prove that there are at most nine critical points for *all* positive parameter choices and not just generic ones.

The first step is to show that there are always only *finitely* many critical points.

Proposition 7. *For any positive choices of parameters, L has finitely many critical points.*

Proof. The critical points of L are always among the roots of $\{m_0, m_1, m_2\}$. By Proposition 4, every solution has a unique value of t_0 given t_1 and t_2 (the denominator $g_1 + g_{13} - c$ in Equation (7) where $c = g_0 + g_1 + g_2 + g_{12} + g_{13} + g_{123}$ is never 0 for positive data). The values of t_2 are the roots of $Res(n_1, n_2, t_1)$. In particular, if this polynomial is not identically zero, there can be at most finitely

many values. The leading coefficient of t_2^{12} in the "whole" resultant (not just $Elim(\cdot)$) factors as

$$g_0^8(g_1 + g_13 - g_0)(-2g_0 + g_{12} + g_{123})(g_{12} + g_0 + g_{13} + g_{123} + g_1)$$

which is nonzero for positive data. Therefore the polynomial is not identically zero.

A similar argument using the resultant $Res(n_1, n_2, t_2)$ shows that there are only finitely many values for t_1 also. Therefore, for positive data, there are in total only finitely many solutions to $\{m_0, m_1, m_2\}$ and hence only finitely many critical points. ■

Next we need to understand the "components" of an algebraic set and their dimensions. Recall that an algebraic set is the set of common solutions to a collection of polynomials.

Definition 3. *An algebraic set V is said to be* irreducible *if for any two algebraic sets V_1, V_2 such that $V_1 \cup V_2 = V$, either $V_1 = V$ or $V_2 = V$.*

Every algebraic set V has a unique irreducible decomposition. Each of the irreducible sets in this decomposition are called the *components* of V. The dimension of a component is the usual topological dimension of the set. The following well-known theorem from algebraic geometry bounds the dimension of the components of an algebraic set.

Theorem 6. *Let V be an algebraic set defined by d polynomials $f_1, \ldots, f_d$ in n variables. Then every component of V has dimension at least $n - d$.*

We use this to deduce in our setting the following.

Corollary 7. *All components of the set of all critical points of L, viewed in the nine-dimensional space of the three unknowns t_0, t_1, t_2 and the six parameters, are of dimension exactly 6.*

Proof. The critical points are the algebraic set V defined by eq_0, eq_1, eq_2 in the space of the parameters and unknowns. Of course these are rational functions and not polynomials, so the algebraic set in question is the one defined by the numerators where we remove any components on which one or more of the denominators is zero. This is the geometric interpretation of the saturation discussed above. By Theorem 6, all components have dimension at least 6 which is the dimension 9 minus the number of equations, 3 (this is true both before and after the saturation). If some component had dimension greater than 6, then for a generic choice of the six parameters there would still be a remaining degree of freedom in V and hence an infinite number of critical points, a contradiction to Proposition 7. ■

We can now apply one more theorem from algebraic geometry. This is called the "specialization principle" and appears, in somewhat more sophisticated language, as Theorem 3.25 in Mumford's (1976) book on algebraic geometry.

Theorem 8. *Let V be an algebraic set in d variables and n parameters such that all components are of dimension exactly n. If for a generic choice of parameters there are k points in V ($k < \infty$), then for any choice of parameters if the number of points in V is finite then it is less than or equal to k.*

By Proposition 7, for positive data, there are always a finite number of critical points. By the above theorem, the number of critical points is at most 9 (for *any* positive data). It remains to show that every critical point is always a root of $Elim(t_2)$. We first state an auxiliary claim.

Claim 4. *Let V_1 be the the set of solutions to $\{eq_0, eq_1, eq_2\}$ and V_2 the solutions of $\{m_0, m_1, m_2\}$. Then, every component of V_1 is contained in a component of V_2.*

Proof. By Proposition 2, $V_1 \subseteq V_2$. Let C be a component of V_1 and $C_1, C_2, \ldots, C_k$ be components of V_2. Seeking contradiction, assume C is not contained in any component C_i of V_2. Then $C = \bigcup_{l=1}^{k}(C \cap C_i)$ where each $C \cap C_i$ is a *proper* subset of C. This contradicts the irreducibility of the component C. ■

Proposition 8. *The critical points of L always come from the roots of Elim*(t_2).

Proof. Let V_1 be the the set of solutions to $\{eq_0, eq_1, eq_2\}$ and V_2 the solutions of $\{m_0, m_1, m_2\}$. Recall that the resultant R has three factors, one of which is $Elim(t_2)$, and the other two vanish when $t_2 = 0$ or $t_1 = -1$. If we add the equation $t_2 = 0$ to the system m_0, m_1, m_2, we can explicitly solve to get

$$t_0 = \frac{1}{2} - \frac{g_1 + g_{12} + g_3 + g_{13}}{2c},$$

$$t_1 = 1 - \frac{2(g_1 + g_{13})}{c}.$$

Since we get a solution for almost all choices of the six parameters, the corresponding algebraic set is of dimension 6. However, as all components of V_2 are of dimension exactly 6, the above set is a component C of V_2. But we can explicitly plug in $t_2 = 0$ and the expressions for t_0 and t_1 above to see that there are choices of parameters which do not yield a critical point of L. In particular, C is not a component of V_1, and any component of V_1 contained in C must have strictly smaller dimension than 6. Since all components of V_1 are of dimension exactly 6, no component of V_1 is contained in C.

A similar argument and computation shows that $t_1 = -1$ yields a component C' of V_2 which does not contain any components of V_1. Hence, combined with Claim 4, all components of V_1 are contained in components of V_2 with $Elim(t_2) = 0$. ■

Recall from Equation (8) that $t_1 = \frac{p(g,t_2)}{q(g,t_2)}$. As we proved, only in rare cases, $p(g, t_2)$ and $q(g, t_2)$ are zero simultaneously and by Proposition 4, we can obtain t_0 uniquely from t_2 and t_1. In the rare cases when $q(g, t_2)$ is zero, t_1 can be obtained by back-substituting t_2 into n_1.

5. UNIQUE ML MOLECULAR CLOCK COMB PROOF

5.1. Overview

In this section, we prove that the ML molecular clock comb has a unique ML point in the valid region, i.e., in the region where all s_i are between 0 and 1. The proof is somewhat involved so before we move to the formal proof we give a proof sketch that might provide a broad view of the proof. We first define a linear change on our working variables t_0, t_1, t_2, converting them to u_0, u_1, u_2. As was shown above, the expected sequence variables, s_i, can be uniquely recovered from the t_i variables and hence from the u_i as well. We now look at a compact set in $\mathbb{R}^3$ corresponding to u_i and partition this compact set into 14 disjoint subsets. We prove that at 9 of the 14, the likelihood function must have a critical point. At the other sets, it is not a priori clear whether each has a critical point. By Theorem 3, L has at most nine critical points. Therefore, we can conclude that there must be *exactly* one critical point at each of the nine subsets. By the way we partitioned the compact set, only one subset corresponds to admissible values of the s_i; hence, we obtain that the ML point is unique.

The inputs $\hat{s_\alpha}$ are all nonnegative integers, since they count the number of occurrences of patterns. In the theorem, we further restrict ourselves to strictly positive data (i.e., $\hat{s}_\alpha > 0$ for all α). This requirement is reasonable in the biological context and ensures that if any s_α is zero then $L = 0$.

5.2. Proof details

Recall we want to maximize

$$L(s|\hat{s}) = {s_0}^{\hat{s}_0} {s_1}^{\hat{s}_1} {s_2}^{\hat{s}_2} {s_{12}}^{\hat{s}_{12}} {s_3}^{\hat{s}_3} {s_{13}}^{\hat{s}_{13}} {s_{23}}^{\hat{s}_{23}} {s_{123}}^{\hat{s}_{123}}.$$

Under our earlier change of variables

$$s_0 = t_0,$$

$$s_1 = s_2 = \frac{(1-t_1)(1+t_2)}{8},$$

$$s_3 = \frac{t_1}{4} + \frac{1}{4} + \frac{t_2}{2} - t_0,$$

$$s_{12} = -t_0 + \frac{1}{4} + \frac{t_1}{4} + \frac{t_2}{4} + \frac{t_1 t_2}{4},$$

$$s_{13} = s_{23} = \frac{(1-t_1)(1-t_2)}{8},$$

$$s_{123} = -\frac{t_1 t_2}{4} - \frac{3t_2}{4} + t_0.$$

Theorem 9. *If all the $\hat{s}_i$ are positive integers, then L has a unique maximum where all s_i are between 0 and 1.*

Proof. First we make another simple linear change of variables which will make some of the arguments below more transparent.

$$t_0 = u_0$$

$$t_1 = 4u_1 - 3$$

$$t_2 = 2u_2 - 1$$

Therefore, we obtain

$$s_0 = u_0,$$

$$s_1 = s_2 = u_2(1-u_1),$$

$$s_{12} = u_2(2u_1 - 1) - u_0,$$

$$s_3 = u_1 + u_2 - 1 - u_0,$$

$$s_{13} = s_{23} = (1-u_1)(1-u_2),$$

$$s_{123} = u_0 - u_1(2u_2 - 1).$$

Let K be the following compact set in $\mathbb{R}^3$:

$$K = \{(u_0, u_1, u_2) \ : -1 \le u_0 \le 1, 0 \le u_1, u_2 \le 1\}.$$

The set K above is where we are going to look for our nine critical points. Indeed, we will find nine disjoint subsets of K and prove that there is a critical point in each of these subsets. One of these subsets will correspond to all s_i in the range between 0 and 1 giving us our unique ML point.

Claim 5. *$s_1 = s_2 = u_2(1-u_1)$ and $s_{13} = s_{23} = (1-u_1)(1-u_2)$ are always between 0 and 1 in K.*

Proof. By the definition of K, u_1 and u_2 are between 0 and 1, and the claim follows. ■

This leaves s_0, s_{12}, s_3, and s_{123} as the four coordinates of s, which may be negative on K.

Definition 4. *A* sign pattern *is a sequence of four symbols, + or −, indicating whether s_0, s_{12}, s_3, and s_{123} are nonnegative or nonpositive, respectively.*

Each sign pattern corresponds to a subset of K on which that sign pattern is realized. For example, the sign pattern $++++$ corresponds to the subset of K where all s_i are nonnegative. Since the s_i sum to 1, this is the same as the subset of K for which the s_i are all between 0 and 1. A priori, there are $2^4 = 16$ possible sign patterns. These sign patterns overlap only in the boundaries where the likelihood function L equals 0.

The next two claims show that the interior of a few sign patterns is empty in K.

Claim 6. *If $s_0 < 0$, $s_{12} < 0$, and $s_3 > 0$, then $s_{123} < 0$. In particular, the interior of sign pattern $--++$ never occurs on K.*

Proof. The inequality $s_{12} < s_3$ implies $u_2(2u_1 - 1) < u_1 + u_2 - 1$. Collecting terms and factoring, we get $(1-u_1)(2u_2-1) > 0$. Since $(1-u_1) \geq 0$, for the latter condition to hold, we must have $(2u_2-1) > 0$; therefore, $s_{123} = u_0 - u_1(2u_2 - 1) < 0$ as desired. ■

Claim 7. *If $s_0 > 0$, $s_{12} > 0$, and $s_3 < 0$, then $s_{123} > 0$. Hence the sign pattern $++--$ does not occur.*

Proof. Similar to above. ■

Of the remaining 14 sign patterns, we will now show that there are 9 for which L is 0 on the boundary. The boundary of the sign patterns either comes from the boundary of K or where s_0, s_{12}, s_3, or s_{123} is 0. L is always 0 in the latter cases, so we need only look at the boundary of K.

Claim 8. *$L = 0$ on the boundary of K except possibly when $u_0 = \pm 1$ or $u_1 = 0$.*

Proof. The boundary of K corresponds to $u_2 = 1$ or 0, $u_1 = 1$ or 0, or $u_0 = \pm 1$. If $u_2 = 1$ or $u_1 = 1$, then $s_{13} = s_{23} = 0$. If $u_2 = 0$, then $s_1 = s_2 = 0$. ■

Claim 9. *When $u_0 = 1$, we are either in sign pattern $+--+$ or one of $s_i = 0$. Similarly, when $u_0 = -1$, we are either in sign pattern $-++-$ or one of $s_i = 0$.*

Proof. When $u_0 = 1$, $s_0 = 1 > 0$ and $s_{123} = 1 - u_1(2u_2 - 1) \geq 0$, while $s_{12} = u_2(2u_1 - 1) - 1 \leq 0$ and $s_3 = u_1 + u_2 - 2 \leq 0$ as desired. A similar argument shows the sign pattern is reversed when $u_0 = -1$. ■

Claim 10. *When $u_1 = 0$, either some $s_i = 0$, or both s_0 and s_{123} are negative, or both are positive but s_{12} and s_3 are both negative. In other words, we are in a sign pattern of the form $-??-$ or $+--+$.*

Proof. When $u_1 = 0$, $s_0 = s_{123} = u_0$ while $s_{12} = -u_2 - u_0$ and $s_3 = -(1 - u_2) - u_0$. Therefore, s_0 and s_{123} have the same sign; they are in fact equal, and furthermore if this sign is positive then, as $0 \leq u_2 \leq 1$, both s_3 and s_{12} must be negative as desired. ■

Corollary 10. *The nine remaining sign patterns $---+$, $-+-+$, $-+++$, $++++$, $+++-$, $++-+$, $+-++$, $+-+-$, and $+---$ give nonempty subsets of K for which L is zero on the boundary and nonzero in the interior.*

We remark that in each sign pattern, the function retains its sign (nonnegative or nonpositive). In some of these regions, the function

$$L(s|\hat{s}) = {s_0}^{\hat{s}_0} {s_1}^{\hat{s}_1} {s_2}^{\hat{s}_2} {s_{12}}^{\hat{s}_{12}} {s_3}^{\hat{s}_3} {s_{13}}^{\hat{s}_{13}} {s_{23}}^{\hat{s}_{23}} {s_{123}}^{\hat{s}_{123}}$$

may attain *negative values*, for example when a specific $\hat{s}_\alpha$ is an odd integer, and we consider the region where the corresponding s_α is negative, and all other variables are positive.

Proof. The zero on the boundary follows from the previous claims which show that nonzero boundary points can occur only on one of the other sign patterns. Zero *interior* points are impossible from the definition of L and K and the formula for s. Finally, the nonemptiness of these sets (as opposed to $--++$ or $++--$ which would be forced to have zero boundary except that they are empty) is easily checked by example:

For every sign pattern, we give an example **s** vector and show that the resulting **u** is in k.

1. • $u = [-0.4559, 0.4947, 0.0191]$
   • $[s_0, s_{12}, s_3, s_{123}] = [-0.4559, 0.4557, -0.0301, 0.0197]$
   • sign pattern $[-,+,-,+]$
2. • $u = [-0.8117, 0.8651, 0.0195]$
   • $[s_0, s_{12}, s_3, s_{123}] = [-0.8117, 0.8260, 0.6965, 0.0195]$
   • sign pattern $[-,+,+,+]$
3. • $u = [-0.1179, 0.2386, 0.2403]$
   • $[s_0, s_{12}, s_3, s_{123}] = [-0.1179, -0.0076, -0.4030, 0.0060]$
   • sign pattern $[-,-,-,+]$
4. • $u = [0.3840, 0.9090, 0.4723]$
   • $[s_0, s_{12}, s_3, s_{123}] = [0.3840, 0.0023, -0.0026, 0.4343]$
   • sign pattern $[+,+,-,+]$
5. • $u = [0.7116, 0.9310, 0.8505]$
   • $[s_0, s_{12}, s_3, s_{123}] = [0.7116, 0.0215, 0.0699, 0.0588]$
   • sign pattern $[+,+,+,+]$
6. • $u = [0.7330, 0.8544, 0.8920]$
   • $[s_0, s_{12}, s_3, s_{123}] = [0.7330, -0.1007, 0.0134, 0.0630]$
   • sign pattern $[+,-,+,+]$
7. • $u = [0.7117, 0.9091, 0.9230]$
   • $[s_0, s_{12}, s_3, s_{123}] = [0.7117, 0.0436, 0.1205, -0.0575]$
   • sign pattern $[+,+,+,-]$
8. • $u = [0.4006, 0.4712, 0.9270]$
   • $[s_0, s_{12}, s_3, s_{123}] = [0.4006, -0.4539, -0.0023, -0.0018]$
   • sign pattern $[+,-,-,-]$
9. • $u = [0.7103, 0.8228, 0.9446]$
   • $[s_0, s_{12}, s_3, s_{123}] = [0.7103, -0.1003, 0.0571, -0.0213]$
   • sign pattern $[+,-,+,-]$ ■

Corollary 11. *For any positive data, there are exactly nine critical points lying exactly one in each of the nine regions above.*

Proof. By compactness, both the minimum and maximum of L are obtained on each set. However, for positive data, L is zero on the boundary of each of these regions and nonzero on the interior. Therefore, at least one extremum must be in the interior and hence a critical point. Since there are at most nine critical points, these are the only ones, and there is exactly one in each interior. ■

In particular, the sign pattern $++++$ has a unique critical point in the interior. As we mentioned earlier, this set is exactly the set where all s_i are between 0 and 1 and the proof is complete. (Note that this critical point must be a maximum since L is positive in this set and 0 on the boundary). ■

6. CONCLUDING REMARKS

This work introduced novel applications of algebraic techniques to get analytical solutions for the four taxa ML molecular clock comb in the Neyman two-states substitution model. We believe this is a significant step in this area of research as it closes the gap formed by the analytical solution for the fork. It would be

of interest to extend these techniques to the four taxa trees *without molecular clock*, and possibly to four state characters (the models of Jukes–Cantor, Kimura, etc.). One major promise of this line of research is to rigorously prove global properties of ML trees (for example, the uniqueness of the ML point under the molecular clock model). Numeric approaches can only suggest conjectured properties. A specific open problem is the number of local ML points for nonmolecular clock, four taxa trees, where currently the only known cases are 1, 2, or infinitely many local ML points. Even though we resort to a numeric tool (root finder) at the final step of the solution, the situation is very different than the numeric methods used in ML in general. The uniqueness of local maxima for the molecular clock comb implies that hill climbing methods need not employ multiple starting points here.

ACKNOWLEDGMENTS

We would like to thank Mike Hendy and Mike Steel for very helpful discussions and for comments on an earlier version of this work. We also thank Dan Geiger and Shmuel Onn for fruitful discussions. Finally, we would like to thank the Singular team from the Centre for Computer Algebra, University of Kaiserslautern, for their assistance in the prelimary computational steps of this work.

REFERENCES

Chor, B., Hendy, M., Holland, B., and Penny, D. 2000. Multiple maxima of likelihood in phylogenetic trees: An analytic approach. *MBE* 17(10), 1529–1541. Earlier version appeared in *RECOMB 2000*.

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

Chor, B., Khetan, A., and Snir, S. 2003. Maximum likelihood on four taxa phylogenetic trees: Analytic solutions. *Proc. 7th Ann. Int. Conf. on Computational Molecular Biology (RECOMB)*, 76–83.

Chor, B., and Snir, S. 2004. Molecular clock fork phylogenies: Closed form analytic maximum likelihood solutions. *Syst. Biol.* 56(6), 963–967.

Gelfand, I., Kapranov, M., and Zelevinsky, A. 1994. *Discriminants, Resultants and Multidimensional Determinants*, Birkhauser, Boston.

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

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

Jukes, T.H., and Cantor, C.R. 1969. Evolution of protein molecules, *in* Munro, H.N., ed., *Mammalian Protein Metabolism*, 21–132, Academic Press, New York.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*, Cambridge University Press, London.

Mumford, D. 1976. *Algebraic Geometry I: Complex Projective Varieties*, Springer, Berlin.

Neymann, J. 1971. Molecular studies of evolution: A source of novel statistical problems, *in* Gupta, S., and Jackel, Y., eds., *Statistical Decision Theory and Related Topics*, 1–27, Academic Press, New York.

Steel, M. 1994. The maximum likelihood point for a phylogenetic tree is not unique. *Syst. Biol.* 43(4), 560–564.

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

Address correspondence to:
Sagi Snir
Mathematics Dept.
University of California
Berkeley, CA 94720

E-mail: ssagi@math.berkeley.edu