Towards Optimal Distance Functions for Stochastic Substitutions Models

Ilan Gronau Shlomo Moran Irad Yavneh

March 23, 2009

Abstract

Distance based reconstruction methods of phylogenetic trees consist of two independent parts: first, inter-species distances are inferred assuming some stochastic model of sequence evolution; then the inferred distances are used to construct a tree. In this paper we concentrate on the task of inter-species distance estimation. Specifically, we characterize the family of valid distance functions for the assumed substitution model and show that deliberate selection of distance function significantly improves the accuracy of distance estimates and, consequently, also improves the accuracy of the reconstructed tree.

Our contribution consists of three parts: First, we present a general framework for constructing families of additive distance functions for stochastic evolutionary models. Then, we present a method for selecting (near) optimal distance functions, and we conclude by presenting simulation results which support our theoretical analysis.

1 Introduction.

One of the most popular approaches to phylogenetic reconstruction is the distance based approach. This approach associates *lengths* to the edges of the phylogenetic tree. The *additive distance* between two taxa is then defined as the sum of lengths of edges in the path connecting them in the tree. Traditionally, this method consists of two separate (and independent) phases: first, inter-species distances are inferred from the input DNA sequences[1]; then, the inferred distances are used to construct a tree. The inter-species distances computed in the first phase act as estimates of these additive distances. There are many methods which guarantee accurate reconstruction of the tree given the true additive distances (e.g. [27, 3, 26]). Recent research demonstrates the sensitivity of reconstruction algorithms to inaccuracies (or noise) in the input distance estimates [2, 6, 22, 10]. Some algorithms deal with this by ignoring the distance estimates which appear to be too noisy (see e.g. [6, 16, 11, 5]). In this paper we present a complementary approach which concentrates on the task of increasing the accuracy of inter-species distance estimation.

Common methods for estimation of evolutionary distances are based on some assumed model for DNA site substitution. In this view, each edge in the tree is associated with a 4×4 stochastic substitution matrix which describes the substitution process along it. Distance estimation methods typically differ by the specific site-substitution model they assume. DNA substitution models range from relatively simple models like Jukes-Cantor (JC) [17] and Kimura's 2 Parameter (K2P) [18], to more complex models like Hasegawa-Kishino-Yano (HKY) [13], Tamura-Nei (TmN) [29], and the general time reversible model (GTR) [30]. A substitution model can be identified by the set of substitution matrices it can assign to edges of the tree. The *model tree*, which is the phylogenetic tree with substitution matrices assigned to its edges, defines a joint probability distribution for sites at its leaves. The main underlying assumption behind distance estimation is that the taxon sequences are sampled according to this joint distribution. The distance between two taxa is thus estimated through the following scheme:

- The two taxon sequences are used to obtain an estimate of the marginal joint distribution of sites at the two leaves. This is usually done via maximum-likelihood, meaning that the chosen

[1] The discussion in this paper is limited to DNA sequences but can be adjusted to protein or other alphabets.

distribution is the one which maximizes the probability of observing the sequences out of all the distributions which are consistent with our assumptions on the substitution process.

- The pairwise joint distribution of step 1 is used to estimate the *rate of substitution* along the path connecting the two taxa. This is done by applying a *substitution rate (SR) function* (to be defined in Section 3) which, informally, provides an expected count for certain substitution events along the path.

This scheme is required to be *consistent*, meaning that if the pairwise joint distributions are estimated accurately in the first step, then the substitution rates computed in the second step should fit an additive metric along the tree. However, the finite length of taxon sequences implies that we are not likely to accurately recover the joint distributions. This introduces an inherent error into the estimation process. The SR function used in the second step influences the propagation of this inherent error into the distance estimates. Thus different SR functions induce different error patterns. This observation motivates the two main questions we address in this paper: **(a)** given the assumed substitution model, what are its valid SR functions? **(b)** among these SR functions, which is the one most suitable for the observed taxon-sequences?

The issue of selecting an appropriate SR function is demonstrated by the following experiments, whose results are described in Fig. 1. These experiments test various SR functions of the K2P model on a simple quartet tree. The K2P model differentiates between two types of substitutions: *transitions* (i.e., `A` $\leftrightarrow$ `G` and `C` $\leftrightarrow$ `T`), and *transversions* (i.e. {`A`,`G`} $\leftrightarrow$ {`C`,`T`}). The standard distance formula for K2P [18] estimates distances as the expected number of overall substitutions. However, there is an alternative approach which counts only transversions by reducing the substitution model to the CFN model [4, 8, 23] over two states – purines (`A`,`G`) and pyrimidines (`C`,`T`). We tested the performance of each formula in reconstructing symmetric quartets whose (single) internal edge is 5 times shorter than all four external edges. This template quartet was considered in a range of scales. 10,000 simulations were done for each quartet in a specified scale, using 500-long sequences under a transition-to-transversion ratio (ti/tv, or R) of 2. In each simulation, the topology of the quartet was resolved according to the four-point method (FPM) [6] using each of the two distance estimation formulae separately. For each formula we recorded the number of times it failed to yield the correct topology. Observing the results, the standard formula performs well in recovering quartets of small scale, but its performance dramatically deteriorates as the scale grows. The 'transversions-only' formula does not perform as well as the standard one for small scales, but it outperforms it for larger scales. In this paper we provide an analytic framework which predicts this behavior. We also use this analysis to devise an alternative distance estimation formula ('max-optimal' in the graph of Fig. 1), which clearly outperforms the other two in all scales of this symmetric quartet. The reader is referred to Section 6 for further details.

The rest of the paper is organized as follows. Section 2 provides a general presentation of substitution models. Section 3 introduces the concept of SR functions and their use in obtaining additive distances.

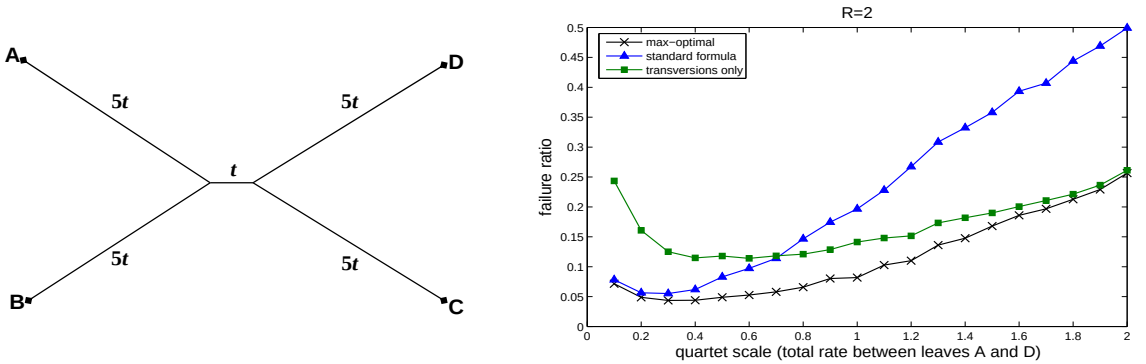

Figure 1: **Influence of distance estimation methods.** Left: the template quartet on which simulations were run. Right: ratios of reconstruction failure using the four-point method under three different distance estimation formulae.

It also provides a general recipe (based on a generalization of logdet) for constructing a large class of SR functions. While the discussion in these sections is rather general, our main interest lies in *unified* substitution models (which include the most commonly assumed substitution models). Section 4 contains the analytic framework for evaluating the propagation of error implied by the different SR functions in a given unified substitution model. This analysis is applied to the K2P model, which is essentially the simplest one for which it can yield non-trivial results (all SR functions of the JC model are shown to be equivalent). In Section 5 we present some experimental results for the K2P model which indicate a significant reduction in noise compared to its standard distance formula. Section 6 contains some concluding discussion and a list of points for further research.

2 Modelling Site Substitution in DNA Evolution.

In this section we describe the general framework used for modelling DNA sequence evolution. Our notation is similar (and often identical) to the standard terminology in this area, but is adapted to our needs. We use the convention that matrices are represented by bold capital letters ($\mathbf{P}$), while vectors and sets are denoted by capital letters (V).

A phylogenetic tree $T = (V, E)$ is defined by its vertex set V and edge set E. The leaf-set of T, denoted by $L \subset V$, corresponds to a set of extant species, whereas the internal vertices of T correspond to extinct ancestors of these extant species. We assume that there are no degree-2 vertices in the tree; it is often even assumed that all internal vertices have degree 3 (that is, the tree is *fully resolved*). Each vertex of T is associated with a fixed (but arbitrary) number of *sites*, denoted by k. Each site σ is a probability distribution on the sample space $\{\mathtt{A,G,C,T}\}^V$ (that is, a simple event in the sample space is an assignment of DNA letters to the vertices in V). For a vertex $v \in V$, σ_v denotes the marginal distribution of σ at v, and Π_v is the vector which represents this distribution, i.e., $\Pi_v(a) = \Pr(\sigma_v = a), a \in \{\mathtt{A,G,C,T}\}$. In our discussion we assume that all k sites are identical, meaning that they define k i.i.d. random variables on V. The joint distribution of $\sigma = \{\sigma_v\}_{v \in V}$ along the evolutionary tree T is described using a stochastic process, as detailed below.

2.1 Substitution models.

A *model tree* is a phylogenetic tree $T = (V, E)$ coupled with substitution parameters. Formally, it is a tuple $(T, r, \Pi_r, \{\mathbf{P}_e\}_{e \in E})$, where $r \in V$ is specified as the root of T, Π_r is the distribution-vector for σ_r, and $\{\mathbf{P}_e\}_{e \in E}$ is a set of *substitution matrices* associated with the edges of T. Each substitution matrix $\mathbf{P}_e$ is a positive row-stochastic matrix (meaning that each of its rows sums up to 1). The distribution of σ along T is described using a process of state propagation from the root toward the leaves. The root r induces directionality on T, in which each vertex $v \in V \setminus \{r\}$ has a unique *parent* (its neighbor on the path to r), and each edge is directed from a parent toward its child. The propagation process is described as follows: the state of the root is selected according to the distribution represented by Π_r; once a state $a \in \{\mathtt{A,G,C,T}\}$ is selected for vertex u, the state of a child v of u is selected according to the distribution represented by the a-th row of $\mathbf{P}_{(u,v)}$. Note that this propagation process defines the joint probability distribution of $\{\sigma_v\}_{v \in V}$, where each substitution matrix represents the conditional probability of a child-state given a parent-state, i.e., $\mathbf{P}_{(u,v)}(a, b) = \Pr[\sigma_v = b | \sigma_u = a]$.

The pairwise conditional and joint distributions associated with directed edges (via the substitution matrices) can be extended to every vertex pair $(u, v) \in V \times V$. For every such pair, the matrix $\mathbf{F}_{uv}$ describes the joint probability distribution of (σ_u, σ_v), meaning that $\mathbf{F}_{uv}(a, b) = \Pr[\sigma_u = a \wedge \sigma_v = b]$, and the matrix $\mathbf{P}_{uv}$ describes the conditional probability distribution of σ_v given σ_u, meaning that $\mathbf{P}_{uv}(a, b) = \Pr[\sigma_v = b | \sigma_u = a]$. Note that $\mathbf{P}$ matrices are always row-stochastic (their rows sum up to 1), whereas in $\mathbf{F}$ matrices the entire matrix sums to 1. Let $\mathbf{\Pi}_v = \mathbf{F}_{vv} = \mathbf{diag}(\Pi_v)$ denote the diagonal matrix whose 4 diagonal elements describe the distribution of σ_v. The $\mathbf{\Pi}$ matrices describe the basic relation between $\mathbf{P}$ matrices and $\mathbf{F}$ matrices, as follows:

$$\mathbf{P}_{vu}^T \mathbf{\Pi}_u \;=\; \mathbf{F}_{uv} \;=\; \mathbf{\Pi}_v \mathbf{P}_{uv} \;. \tag{1}$$

Different evolutionary models are specified by restrictions they apply on the substitution matrices. Hence we identify substitution models with sets of substitution matrices:

Definition 2.1. *A substitution model for* $\{\mathtt{A},\mathtt{G},\mathtt{C},\mathtt{T}\}$ *is a set* $\mathcal{M}$ *of* 4×4 *row-stochastic matrices which is closed under matrix product (i.e.* $\mathbf{P},\mathbf{Q}\in\mathcal{M}\Rightarrow\mathbf{PQ}\in\mathcal{M}$*).*

We say that a model tree belongs to a substitution model $\mathcal{M}$, if $\mathbf{P}_{uv}\in\mathcal{M}$ for every $u,v\in V(T)$. For example, the substitution model of JC is defined by the set of matrices $\mathcal{M}_{\mathsf{JC}}=\left\{(1-4p)\mathbf{I}+p\mathbf{J}:0<p<\frac{1}{4}\right\}$, where $\mathbf{I}_4$ and $\mathbf{J}_{4,4}$ are the 4×4 identity and all-one matrices (resp.). All substitution models addressed in the literature are contained in a *universal model* defined as follows:

$$\mathcal{M}_{\text{univ}}=\{\mathbf{P}:\mathbf{P}\text{ is stochastic and positive }(\mathbf{P}(a,b)>0),\text{ and }0<|\det(\mathbf{P})|<1\} \tag{2}$$

The following lemma states some basic properties of substitution matrices in $\mathcal{M}_{\text{univ}}$ which are used later on. The proof is due to Perron Frobenius theory (see e.g. [15], Chapter 8).

Lemma 2.2. *For every* $\mathbf{P}\in\mathcal{M}_{univ}$*, the following holds:*

1. 1 *is an eigenvalue of* $\mathbf{P}$ *of algebraic multiplicity one.*
2. *For every other eigenvalue* λ *of* $\mathbf{P}$*,* $0<|\lambda|<1$*.*
3. *There is a* unique *eigenvector* Π *of* $\mathbf{P}$ *s.t. for every* $a\in\{\mathtt{A},\mathtt{G},\mathtt{C},\mathtt{T}\}$*,* $\Pi(a)\geq 0$ *and* $\sum_a\Pi(a)=1$*. The eigenvalue corresponding to* Π *is* 1*.*

The vector Π guaranteed by Lemma 2.2(3) is often referred to as the *stationary vector* or the *stationary distribution* of $\mathbf{P}$.

Definition 2.3 (Reversible substitution matrix)**.** *Let* $\mathbf{P}$ *be a substitution matrix, and let* $\mathbf{\Pi}_{stat}$ *be the diagonal matrix representation of* Π_{stat} *– the (unique) stationary vector of* $\mathbf{P}$*. Then* $\mathbf{P}$ *is said to be* (time) reversible *if* $\mathbf{F}=\mathbf{\Pi}_{stat}\mathbf{P}$ *is a symmetric matrix.*

Definition 2.4 (Reversible substitution model)**.** *A substitution model* $\mathcal{M}$ *is said to be* (time) reversible *if all substitution matrices in* $\mathcal{M}$ *are time-reversible and share the same stationary vector.*

Our discussion is mostly restricted to time-reversible substitution models. We assume that when a model tree belongs to a reversible model $\mathcal{M}$, the distribution at the root Π_r is the stationary distribution of $\mathcal{M}$. Note that this guarantees the stationarity of the substitution process (i.e., $\forall u,v\in V(T),\Pi_u=\Pi_v$). An important property of reversible model trees is that $\mathbf{P}_{uv}=\mathbf{P}_{vu}$ for every vertex-pair u,v in T (through Equation §1 and Definition 2.3). This implies that the location of the root has no affect on the model tree and can be ignored. Thus a time-reversible model tree is completely defined by the pair $(T,\{\mathbf{P}_e\}_{e\in E})$. Most of our discussion and analysis focuses on reversible substitution models which have the following useful algebraic property:

Definition 2.5 (Unified Substitutions Model)**.** *A substitution model* $\mathcal{M}$ *is said to be* unified by a matrix $\mathbf{U}$ *iff every matrix* $\mathbf{P}\in\mathcal{M}$ *is triangulated by* $\mathbf{U}$*, meaning that* $\mathbf{U}^{-1}\mathbf{P}\mathbf{U}$ *is an upper-triangular matrix.* $\mathcal{M}$ *is said to be* unified *if it is unified by some matrix* $\mathbf{U}$*.* $\mathcal{M}$ *is said to be* strongly *unified if* $\mathbf{U}^{-1}\mathbf{P}\mathbf{U}$ *is a diagonal matrix for all* $\mathbf{P}\in\mathcal{M}$*.*

Most commonly assumed substitution models (like JC and K2P) are strongly unified. As we show in Section 3, the unifying property is useful in constructing and analyzing distance functions.

2.2 Continuous-time Markov processes and rate matrices

Common substitution models are often described using a continuous-time Markov process. In this section we briefly describe the substitution matrices implied by such processes. A detailed exposition of continuous-time Markov processes can be found in [25], Chapter 16, and its applications to evolutionary models are discussed, e.g., in [20].

A continuous-time Markov process is specified by a 4×4 *substitution rate matrix* (or just *rate matrix*) $\mathbf{R}$. A rate matrix is a matrix whose off-diagonal entries are all non-negative, and whose rows all sum

up to 0. A substitution matrix $\mathbf{P}$ is said to *be realized* by a continuous-time Markov process with rate $\mathbf{R}$ iff $\mathbf{P} = e^{\mathbf{R}}$. Common evolutionary models assume rate matrices which are diagonalizable, meaning that for some invertible matrix $\mathbf{U}$, $\mathbf{D} = \mathbf{U}^{-1}\mathbf{R}\mathbf{U}$ is a diagonal matrix. In such a case, $e^{\mathbf{R}} = \mathbf{U}e^{\mathbf{D}}\mathbf{U}^{-1}$, where $e^{\mathbf{D}}$ is the diagonal matrix obtained by exponentiating the diagonal elements of $\mathbf{D}$.

When possible, it is convenient to associate substitution models with the rate matrices realizing their substitution matrices, rather than with the substitution matrices themselves. This is because the composition of two substitution matrices corresponds to their product, whereas composition of two rate matrices corresponds to their sum. Of special interest are *homogeneous* substitution models, which consist of rate matrices which are all proportional to a single matrix. Formally, the homogenous substitution model of the rate matrix $\mathbf{R}$ is defined by $\mathcal{M}_{\mathbf{R}} = \{e^{t\mathbf{R}} : t > 0\}$.

The expected number of substitutions along a continuous-time Markov process acts as a natural additive distance measure. Indeed, most distance estimation formulae try to estimate a substitution count (of some or all types of substitutions) along the evolutionary path. Now, consider a *stationary* Markov process with rate matrix $\mathbf{R}$. Stationarity implies that the site distribution is the one described by the stationary vector Π_{stat} of $e^{\mathbf{R}}$ (which is the left eigenvector of $\mathbf{R}$ corresponding to eigenvalue 0). The product $\Pi_{stat}(a)\mathbf{R}(a,b)$ is the expected count for substitution $a \to b$ along this stationary Markov process . Hence, the *total substitution rate* (or number of substitutions) is given by:

$$\text{total substitution rate} \quad = \quad \sum_{a \neq b} \Pi_{stat}(a)\mathbf{R}(a,b) \quad = \quad -\sum_{a} \Pi_{stat}(a)\mathbf{R}(a,a) \tag{3}$$

Notice that if the stationary distribution is uniform ($\Pi_{stat} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$), then the total substitution rate is $\frac{1}{4}|trace(\mathbf{R})|$.

2.3 Kimura's 2 Parameter Model

We conclude this section with a short description of the K2P model [18], which we use as a case study for our analysis. A rate matrix $\mathbf{R}$ in the K2P model (see Fig. 2) is specified by two parameters: α, which specifies the rate of *transition*-type substitutions (i.e., A $\leftrightarrow$ G and C $\leftrightarrow$ T), and β, which specifies the rate of *transversion*-type substitutions (i.e. {A,G} $\leftrightarrow$ {C,T}); $\mathbf{R}$ has two distinct non-zero eigenvalues: $\lambda_1(\mathbf{R}) = -4\beta$, $\lambda_2(\mathbf{R}) = \lambda_3(\mathbf{R}) = -2\alpha - 2\beta$. The stationary distribution for this model is uniform (i.e., $\Pi_{stat} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$). The matrix $\mathbf{U}_{\mathsf{K2P}}$ in Fig. 2 diagonalizes all rate matrices, and hence also all the

$$\mathbf{R} = \begin{pmatrix} - & \alpha & \beta & \beta \\ \alpha & - & \beta & \beta \\ \beta & \beta & - & \alpha \\ \beta & \beta & \alpha & - \end{pmatrix} \qquad \Lambda(\mathbf{R}) = \begin{pmatrix} -4\beta \\ -2\alpha - 2\beta \\ -2\alpha - 2\beta \\ 0 \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} p_{\text{nil}} & p_\alpha & p_\beta & p_\beta \\ p_\alpha & p_{\text{nil}} & p_\beta & p_\beta \\ p_\beta & p_\beta & p_{\text{nil}} & p_\alpha \\ p_\beta & p_\beta & p_\alpha & p_{\text{nil}} \end{pmatrix} \qquad \Lambda(\mathbf{P}) = \begin{pmatrix} 1 - 4p_\beta \\ p_{\text{nil}} - p_\alpha \\ p_{\text{nil}} - p_\alpha \\ 1 \end{pmatrix}$$

$$\mathbf{U}_{\mathsf{K2P}} = \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{-1}{\sqrt{2}} & 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix}$$

Figure 2: **The Kimura 2-Parameter (K2P) model.** A generic rate matrix $\mathbf{R}$ of the K2P model is described at the top left. As a general convention, we order the rows and columns according to the order A,G,C,T. The matrix $\mathbf{R}$ is determined using the parameters α, β. The vector $\Lambda(\mathbf{R})$ of its eigenvalues is given next to it. The substitution matrix $\mathbf{P} = e^{\mathbf{R}}$ corresponding to $\mathbf{R}$ is described at the bottom left. It consists of three distinct values $p_{\text{nil}}, p_\alpha, p_\beta$, where $p_{\text{nil}} + p_\alpha + 2p_\beta = 1$. The vector $\Lambda(\mathbf{P})$ of its eigenvalues is given next to it. On the right, is the unifying matrix $\mathbf{U}_{\mathsf{K2P}}$ of the K2P model. The connection between α, β and $p_{\text{nil}}, p_\alpha, p_\beta$ is obtained through the equation $\Lambda(\mathbf{R}) = \ln(\Lambda(\mathbf{P}))$.

substitution matrices, in this model. Thus it *strongly unifies* this model. There are many other matrices which unify the K2P model, such as the Hadamard matrix of order 4 [14]. One notable property of $\mathbf{U}_{\mathsf{K2P}}$ from Fig. 2 is that it is *unitary*, meaning that $\mathbf{U}_{\mathsf{K2P}}^T\mathbf{U}_{\mathsf{K2P}} = \mathbf{I}_4$.

A substitution matrix $\mathbf{P}$ in the K2P model is determined by two parameters: p_α for transition-type entries, and p_β for transversion-type entries. We denote by $p_{\text{nil}} = 1 - p_\alpha - 2p_\beta$ the value of the diagonal entries of $\mathbf{P}$. The matrix $\mathbf{P}$ has two (positive) eigenvalues other than 1: $\lambda_1(\mathbf{P}) = 1 - 4p_\beta$, $\lambda_2(\mathbf{P}) = \lambda_3(\mathbf{P}) = p_{\text{nil}} - p_\alpha = 1 - 2p_\alpha - 2p_\beta$. The parameters of $\mathbf{P}$ and the parameters of $\mathbf{R}$ are tied by exponentiation of the eigenvalues, i.e., the equations $\lambda_t(\mathbf{R}) = \ln(\lambda_t(\mathbf{P}))$, $t = 1, 2$. This gives

$$p_{\text{nil}} = \frac{1}{4} + \frac{1}{4}e^{-4\beta} + \frac{1}{2}e^{-2\alpha-2\beta} \quad ; \quad p_\alpha = \frac{1}{4} + \frac{1}{4}e^{-4\beta} - \frac{1}{2}e^{-2\alpha-2\beta} \quad ; \quad p_\beta = \frac{1}{4} - \frac{1}{4}e^{-4\beta} \ .$$

The transition-to-transversion (ti/tv) ratio of a K2P rate matrix is defined by $R = \frac{\alpha}{2\beta}$. Note that a homogeneous sub-model of K2P is a substitution model that consists of K2P rate matrices which all have the same ti/tv ratio. A ratio of $R = \frac{1}{2}$ (implying that $\alpha = \beta$) results in the sub-model of JC.

3 Additive Distances and Substitution Rate functions.

In this section we explore ways of obtaining additive distances for model trees in a given substitution model. We are interested in distance metrics on the tree vertices $d : V \times V \to \mathbb{R}^+$ which obey the following basic conditions:

Locality: For every $u, v \in V$, the distance $d(u, v)$ is a function of the joint distribution matrix $\mathbf{F}_{uv}$.

Additivity: If w lies on the path connecting u and v in T, then $d(u, v) = d(u, w) + d(w, v)$.

Locality is necessary since we wish to be able to estimate the distance between two taxa i, j just by observing samples from (σ_i, σ_j). Additivity is important for reconstruction using distance-based methods (see e.g. [3, 27]). These two properties are guaranteed by using *additive functions* defined next.

Definition 3.1 (Additive functions – $\mathcal{AD}_{\mathcal{M}}$)**.** *An* additive function for a substitution model $\mathcal{M}$ *is a mapping* $\Delta : \mathcal{M} \to \mathbb{R}$ *which satisfies the following: if* $\mathbf{P}, \mathbf{Q}, \mathbf{S} \in \mathcal{M}$*, and* $\mathbf{S} = \mathbf{P}^a\mathbf{Q}^b$ *for some* $a, b \in \mathbb{R}$*, then* $\Delta(\mathbf{S}) = a\Delta(\mathbf{P}) + b\Delta(\mathbf{Q})$*.* $\mathcal{AD}_{\mathcal{M}}$ *denotes the set of all additive functions for* $\mathcal{M}$*.*

Note: When the model $\mathcal{M}$ is defined by a continuous-time Markov processes as in Section 2.2, we have that for every $\mathbf{P} \in \mathcal{M}$ and $a > 0$, $\mathbf{P}^a$ exists and is in $\mathcal{M}$. For such a model, a function Δ is additive iff it satisfies a simpler condition: for all $\mathbf{P}, \mathbf{Q} \in \mathcal{M}$, $\Delta(\mathbf{PQ}) = \Delta(\mathbf{P}) + \Delta(\mathbf{Q})$.

The additive functions we are interested in are strictly positive. We refer to such functions as *Substitution Rate* (SR) functions.

Definition 3.2 (SR functions – $\mathcal{SR}_{\mathcal{M}}$)**.** Δ *is a* substitution rate (SR) function *for a substitution model* $\mathcal{M}$ *if* $\Delta \in \mathcal{AD}_{\mathcal{M}}$ *and* $\Delta(\mathbf{P}) > 0$ *for all* $\mathbf{P} \in \mathcal{M}$*.* $\mathcal{SR}_{\mathcal{M}}$ *denotes the set of all SR functions for* $\mathcal{M}$*.*

SR functions are used to obtain additive distances using the following scheme:

Lemma 3.3. *Let* Δ *be an SR function for* $\mathcal{M}$*. Then for each model tree of* $\mathcal{M}$*, the function* $d : V \times V \to \mathbb{R}^+$ *defined by* $d(u, v) = \frac{1}{2}(\Delta(\mathbf{P}_{uv}) + \Delta(\mathbf{P}_{vu}))$ *is an additive (and local) metric on* V*. In particular, if* $\mathcal{M}$ *is time-reversible, then* $d(u, v) = \Delta(\mathbf{P}_{uv})$ *is an additive metric on* V*.*

A notable example of an SR function is the well known *logdet* function (also referred to as the *paralinear* distance measure) [28, 21, 19], given by $\Delta(\mathbf{P}) = -\ln|\det(\mathbf{P})|$. logdet is clearly additive (due to the multiplicativity of the determinant) and positive (since $0 < |det(\mathbf{P})| < 1$). Its great advantage is in being applicable to all substitution models included in $\mathcal{M}_{\text{univ}}$. This function is also implicitly used by the standard distance formulae of the JC and K2P models [17, 18]. These formulae obtain

distance estimates between two taxa i, j by applying logdet to a substitution matrix $\widehat{\mathbf{P}}_{ij}$ estimated from the taxon sequences S_i, S_j ($\widehat{\mathbf{P}}_{ij}$ is a matrix in the model which is most likely to produce S_i, S_j - more details in Section 4). We note that when restricted to JC or K2P, logdet has a natural biological interpretation: logdet($\mathbf{P}$) is proportional to the *total substitution rate* associated with $\mathbf{P}$. Recall that $\mathbf{P} = e^{\mathbf{R}}$ and so logdet($\mathbf{P}$) = logdet($e^{\mathbf{R}}$) = $-trace(\mathbf{R})$, and since the stationary distribution in JC and K2P is uniform, the total substitution rate associated with $\mathbf{R}$ is $-\frac{1}{4}trace(\mathbf{R}) = \frac{1}{4}$logdet($\mathbf{P}$).

The first step in our analysis is to figure out what are the SR functions for a given substitution model $\mathcal{M}$. Mathematically, it is more natural to characterize the set of *additive* functions $\mathcal{AD}_{\mathcal{M}}$ because it forms a linear vector space over $\mathbb{R}$: for $\Delta_1, \Delta_2 \in \mathcal{AD}_{\mathcal{M}}$ and $c_1, c_2 \in \mathbb{R}$, the function $\Delta = c_1\Delta_1 + c_2\Delta_2$ is also in $\mathcal{AD}_{\mathcal{M}}$. The dimension of this vector space determines the "richness" of $\mathcal{AD}_{\mathcal{M}}$ (and $\mathcal{SR}_{\mathcal{M}}$). In Section 3.1 below we describe a general technique for obtaining a large class of SR functions spanned by a core set of functions which are generalized versions of logdet.

3.1 Generalized logdet SR functions.

Our generalization of logdet is based on *invariant subspaces* of $\mathbb{R}^4$ (see e.g. [15], Chapter 1).

Definition 3.4 (Invariant Subspace)**.** *Let $\mathcal{H}$ be a linear subspace of $\mathbb{R}^n$, and let $\mathbf{A}$ be an $n \times n$ real matrix. Then $\mathcal{H}$ is said to be* $\mathbf{A}$-invariant *if $\mathbf{A}\mathcal{H} \subseteq \mathcal{H}$ (where $\mathbf{A}\mathcal{H} = \{\mathbf{A}V : V \in \mathcal{H}\}$). Given a collection $\mathcal{A}$ of $n \times n$ matrices, the subspace $\mathcal{H}$ is said to be $\mathcal{A}$-invariant if it is $\mathbf{A}$-invariant for every matrix $\mathbf{A} \in \mathcal{A}$.*

Given an $\mathbf{A}$-invariant subspace $\mathcal{H} \subseteq \mathbb{R}^n$ of dimension $m \leq n$, denote by $\mathbf{A}|_{\mathcal{H}}$ (the *restriction* of $\mathbf{A}$ to $\mathcal{H}$) the linear transformation induced by $\mathbf{A}$ on $\mathcal{H}$. In such a case, let $\det(\mathbf{A}|_{\mathcal{H}})$ be the determinant of the linear transformation $\mathbf{A}|_{\mathcal{H}}$ [2]. The following lemma is based on the fact that $\det(\mathbf{AB}|_{\mathcal{H}}) = \det(\mathbf{A}|_{\mathcal{H}}) \cdot \det(\mathbf{B}|_{\mathcal{H}})$.

Lemma 3.5 (Generalized logdet functions)**.** *Let $\mathcal{H}$ be an $\mathcal{M}$-invariant subspace of $\mathbb{R}^4$ for some substitution model $\mathcal{M} \subseteq \mathcal{M}_{univ}$. Then the mapping $\Delta_{\mathcal{H}} : \mathcal{M} \to \mathbb{R}^+$ defined by $\Delta_{\mathcal{H}}(\mathbf{P}) = -\ln|\det(\mathbf{P}|_{\mathcal{H}})|$ is an SR function for $\mathcal{M}$.*

The 'original' logdet function can be defined as $\Delta_{\mathbb{R}^n}$. Another example of generalized logdet functions, which is of specific interest in this work, is the family of *log-eigenvalue* functions defined below. Assume that all substitution matrices in $\mathcal{M}$ share a common eigenvector V, such that $0 < |\lambda_V(\mathbf{P})| < 1$, where $\lambda_V(\mathbf{P})$ is the eigenvalue of $\mathbf{P}$ corresponding to V (this happens, for instance, if $\mathcal{M}$ is strongly unified). Then $\mathcal{H}_V$, the 1-dimensional linear subspace spanned by V, is an $\mathcal{M}$-invariant subspace. In such a case, $\Delta_{\mathcal{H}_V}(\mathbf{P}) = -\ln|\lambda_V(\mathbf{P})|$. Log-eigenvalue SR functions are (implicitly) mentioned in several places in the literature (see e.g. [12]).

To the best of our knowledge, all published distance formulae are based on SR functions which are spanned by generalized logdet functions. The following lemma indicates that in certain strongly unified substitution models of biological interest (e.g. JC and K2P), all SR functions are indeed generalized logdet functions.

Lemma 3.6. *Let $\mathcal{M}$ be a strongly unified model, and assume that for each $\mathbf{P} \in \mathcal{M}$, all the eigenvalues of $\mathbf{P}$ are positive. Then $\mathcal{AD}_{\mathcal{M}}$ is spanned by the set of generalized logdet functions.*

Proof. Let $\mathbf{U}$ be any matrix which unifies $\mathcal{M}$. For $t = 1, \ldots, 4$, Let U^t be the t-th column of $\mathbf{U}$, and let $\mathcal{H}_t$ be the $\mathcal{M}$-invariant one-dimensional subspace spanned by U^t. Then $\Delta_{\mathcal{H}_t}$ is a generalized logdet SR function defined by $\Delta_{\mathcal{H}_t}(\mathbf{P}) = -\ln|\det(\mathbf{P}|_{\mathcal{H}_t})| = -\ln(\lambda_t(\mathbf{P}))$, where $\lambda_t(\mathbf{P})$ is the eigenvalue of $\mathbf{P}$ corresponding to the eigenvector U^t. The proof is completed by showing that for each $\Delta \in \mathcal{AD}_{\mathcal{M}}$ there are constants $c_1, \ldots, c_4$, such that for all $\mathbf{P} \in \mathcal{M}$, $\Delta(\mathbf{P}) = \sum_{t=1}^{4} c_t \ln(\lambda_t(\mathbf{P}))$.

For $\mathbf{P} \in \mathcal{M}$, let $V_{\mathbf{P}} \in \mathbb{R}^4$ be the vector $[\ln(\lambda_1(\mathbf{P})), \ldots, \ln(\lambda_4(\mathbf{P}))]$. Then since $\lambda_t(\mathbf{P}) > 0$ for all $\mathbf{P} \in \mathcal{M}$ and $t = 1..4$, the mapping $f(\mathbf{P}) = V_{\mathbf{P}}$ is 1 to 1 (implying that $V_{\mathbf{P}}$ uniquely determines $\mathbf{P}$).

[2] Recall that $\det(\mathbf{A}|_{\mathcal{H}}) = \prod_t \lambda_t^{e_t}$, where λ_t varies over the distinct eigenvalues of $\mathbf{A}|_{\mathcal{H}}$ (which are also eigenvalues of $\mathbf{A}$) and e_t is the algebraic multiplicity of λ_t.

Let $L_f = \{V_{\mathbf{P}} : \mathbf{P} \in \mathcal{M}\}$ be the image of f. For an additive function $\Delta \in \mathcal{AD}_{\mathcal{M}}$, let $T_\Delta : L_f \to \mathbb{R}$ be defined by $T_\Delta(V_{\mathbf{P}}) = \Delta(\mathbf{P})$. Since Δ is additive, we have that

$$\forall \mathbf{P}, \mathbf{Q} \in \mathcal{M}, \alpha \in \mathbb{R} : T_\Delta(V_{\mathbf{P}} + V_{\mathbf{Q}}) = T_\Delta(V_{\mathbf{P}}) + T_\Delta(V_{\mathbf{Q}}), \text{ and if } \mathbf{P}^\alpha \in \mathcal{M} \text{ then } T_\Delta(\alpha V_{\mathbf{P}}) = \alpha T_\Delta(V_{\mathbf{P}}).$$

This means that T_Δ is a linear scalar function on L_f, which can be extended to a *total* linear function T'_Δ on $\mathbb{R}^4$. For $t = 1, \ldots, 4$, let $c_t = T'_\Delta(\bar{e}_t)$, where $\bar{e}_t$ is the unit vector with 1 at the t-th entry and 0 elsewhere. Then for each vector $X = [x_1, \ldots, x_4] \in \mathbb{R}^4$, $T'_\Delta(X) = \sum_{t=1}^4 c_t x_t$. Also, for all $\mathbf{P} \in \mathcal{M}$, $V_{\mathbf{P}} = \sum_{t=1}^4 \ln(\lambda_t(\mathbf{P}))\bar{e_t}$. Hence $\Delta(\mathbf{P}) = T'_\Delta(V_{\mathbf{P}}) = \sum_{t=1}^4 c_t \ln(\lambda_t(\mathbf{P}))$ as claimed. □

3.2 Log-eigenvalue SR functions for unified substitution models

The remainder of this paper focuses on the class of SR functions spanned by log-eigenvalue functions. First, we establish that if $\mathcal{M}$ is unified, then each log-eigenvalue function is an SR functions for $\mathcal{M}$. Assume a fixed (but arbitrary) substitution model $\mathcal{M}$, unified by a 4×4 matrix $\mathbf{U}$. For every substitution matrix $\mathbf{P} \in \mathcal{M}$, Let $\mathbf{\Lambda}(\mathbf{P}) = \mathbf{U}^{-1}\mathbf{P}\mathbf{U}$ denote the upper triangular matrix associated with $\mathbf{P}$. Denote further by $\lambda_t(\mathbf{P})$ the t'th diagonal entry of $\mathbf{\Lambda}(\mathbf{P})$ ($t = 1 \ldots 4$). Note that $\lambda_t(\mathbf{P})$ is an eigenvalue of $\mathbf{P}$. Hence, since $\mathbf{P}$ is row-stochastic, then there is a $t \in \{1, 2, 3, 4\}$ s.t. $\lambda_t(\mathbf{P}) = 1$. We now show that this t is unique and common to all matrices in $\mathcal{M}$.

Lemma 3.7. *There exists $t_0 \in \{1 \ldots 4\}$ s.t. for each $\mathbf{P} \in \mathcal{M}$: $\lambda_{t_0}(\mathbf{P}) = 1$, and $0 < |\lambda_t(\mathbf{P})| < 1$ for $t \neq t_0$.*

Proof. By Lemma 2.2(1,2), we have that for each matrix $\mathbf{P} \in \mathcal{M}$ there is a unique t s.t. $\lambda_t(\mathbf{P}) = 1$, and for each $t' \neq t$, $0 < |\lambda_{t'}(\mathbf{P})| < 1$. So it remains to prove that for some fixed t_0, $\lambda_{t_0}(\mathbf{P}) = 1$ for all $\mathbf{P} \in \mathcal{M}$. Consider an arbitrary matrix $\mathbf{P} \in \mathcal{M}$. Then for some t_0, $\lambda_{t_0}(\mathbf{P}) = 1$. We show that for any other matrix $\mathbf{Q} \in \mathcal{M}$, $\lambda_{t_0}(\mathbf{Q}) = 1$ as well. Assume for contradiction that $\lambda_{t_0}(\mathbf{Q}) \neq 1$. Then for $t = 1, \ldots, 4$, it holds that $|\lambda_t(\mathbf{PQ})| = |\lambda_t(\mathbf{P})\lambda_t(\mathbf{Q})| < 1$ (here we used the fact that $\mathcal{M}$ is unified, and $\mathbf{\Lambda}(\mathbf{P}), \mathbf{\Lambda}(\mathbf{Q})$, are upper triangular). This contradicts the fact that $\mathbf{PQ} \in \mathcal{M}$ is row-stochastic and so must have 1 as an eigenvalue . □

From now on we assume for brevity that t_0 of Lemma 3.7 is 4 (if $\mathcal{M}$ is strongly unified then this can be achieved by permuting the columns of $\mathbf{U}$). The following theorem shows that log-eigenvalue functions are SR functions in $\mathcal{M}$ by showing that they are spanned by generalized logdet functions.

Theorem 3.8. *For $t \in \{1, 2, 3\}$, let $\lambda_t(\mathbf{P})$ be the t'th eigenvalue of $\mathbf{P}$ as defined above. Then the function $\Delta_t(\mathbf{P}) \stackrel{\triangle}{=} -\ln|\lambda_t(\mathbf{P})|$ is an SR function for $\mathcal{M}$.*

Proof. For $t = 1 \ldots 4$, let U^t be the t'th column of $\mathbf{U}$. Let further $\mathcal{H}_1, \mathcal{H}_2$ and $\mathcal{H}_3$ be the subspaces of $\mathbb{R}^4$ spanned by $\{U^1\}, \{U^1, U^2\}$ and $\{U^1, U^2, U^3\}$, respectively. Then, since $\mathcal{M}$ is unified by $\mathbf{U}$, we have that $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$ are all $\mathcal{M}$-invariant. The determinants of the restrictions of $\mathbf{P} \in \mathcal{M}$ to these subspaces are $\det(\mathbf{P}|_{\mathcal{H}_1}) = \lambda_1(\mathbf{P})$, $\det(\mathbf{P}|_{\mathcal{H}_2}) = \lambda_1(\mathbf{P})\lambda_2(\mathbf{P})$, and $\det(\mathbf{P}|_{\mathcal{H}_3}) = \lambda_1(\mathbf{P})\lambda_2(\mathbf{P})\lambda_3(\mathbf{P})$. Hence, by Lemma 3.5, the functions

$$-\ln|\lambda_1(\mathbf{P})| \quad , \quad -(\ln|\lambda_1(\mathbf{P})| + \ln|\lambda_2(\mathbf{P})|) \quad , \quad -(\ln|\lambda_1(\mathbf{P})| + \ln|\lambda_2(\mathbf{P})| + \ln|\lambda_3(\mathbf{P})|)$$

are all (generalized logdet) SR functions for $\mathcal{M}$. The theorem then follows by the linearity of $\mathcal{AD}_{\mathcal{M}}$, and the fact that $0 < |\lambda_t(\mathbf{P})| < 1$ for $t = 1..3$. □

The rest of the paper studies the class of SR functions spanned by the log-eigenvalue functions $\Delta_1, \Delta_2, \Delta_3$ mentioned in Theorem 3.8. Every SR function $\Delta = c_1\Delta_1 + c_2\Delta_2 + c_3\Delta_3$ in this class is defined by the three *SR coefficients* c_1, c_2, c_3. For instance, the logdet function is in this class and its SR coefficients are all 1. The "richness" of this class of SR functions is measured by its *dimension*, which is the maximal number linearly independent SR functions in the class. By definition, its dimension cannot be greater than 3, however, in some cases the functions $\Delta_1, \Delta_2, \Delta_3$ might be linearly dependent. For instance, in the JC model, for every $\mathbf{P} \in \mathsf{JC}$ we have $\lambda_1(\mathbf{P}) = \lambda_2(\mathbf{P}) = \lambda_3(\mathbf{P})$, implying that the

three log-eigenvalue functions are identical and (through Lemma 3.6) that all SR functions in $\mathcal{SR}_{\mathsf{JC}}$ are proportional to that log-eigenvalue function. In our study, proportional SR functions are considered to be *equivalent*, since they have the same relative error rate (see Section 5). Hence, there is no room for optimization in models (like JC) in which the dimension of $\mathcal{SR}_{\mathcal{M}}$ is 1.

K2P is the simplest published model which provides a non-trivial selection of SR function. Due to the structure of substitution matrices in the K2P model, this model has two linearly independent basic SR functions: $\Delta_1 \neq \Delta_2 (= \Delta_3)$. Every other SR function in $\mathcal{SR}_{\mathsf{K2P}}$ is a linear combination of these two basic functions, and every valid distance formula for K2P must use one of these SR functions. For instance, the standard distance formula for K2P [18] uses the logdet SR function which is equal to $\Delta_1 + 2\Delta_2$. On the other hand, the distance formula which considers only transversions (reducing the model to a binary one) corresponds to the SR function Δ_1.

4 Analyzing the Error in Distance Estimation.

The previous section defined the set of SR functions for a given substitution model $\mathcal{M}$, and showed that each SR function induces an additive metric D over every model tree in $\mathcal{M}$. This additive metric can be used to reconstruct the topology of the model tree from the pairwise substitution matrices $\{\mathbf{P}_{ij} : i, j \in L\}$. Unfortunately, the limited length of taxon sequences allows us only to obtain a *statistical estimate* $\widehat{\mathbf{P}}_{ij}$ for each substitution matrix $\mathbf{P}_{ij}$. Applying an SR function to the estimated substitution matrices results in a statistical estimate $\widehat{D}$ for the additive metric D induced by that SR function. The inherent stochastic error in the estimation of the substitution matrices propagates through the SR function and results in errors in the estimates of the additive distances. This deviation from the additive metric determines how accurately we will be able to reconstruct the topology of the model tree. Hence, the obvious goal is to find an SR function which minimizes the propagation of the inherent statistical error.

The first step to take in pursuing such an optimization is to provide an expression for the expected error in distance estimation associated with an SR function on any given evolutionary path. Hence, the discussion from this point on considers an arbitrary evolutionary path under some unified (and time-reversible) substitution model $\mathcal{M}$. Let $\mathbf{P} \in \mathcal{M}$ and $\mathbf{F}$ denote the substitution matrix and joint distribution matrix of this path, respectively. Given an SR function $\Delta = c_1\Delta_1 + c_2\Delta_2 + c_3\Delta_3$, we denote by $d = \Delta(\mathbf{P})$ the distance between the endpoints of the corresponding path according to Δ , and by $\widehat{d}$ the random variable corresponding to the estimated value of this distance. The model parameters $\mathbf{P}, \mathbf{F}$ determine the distribution of the inherent stochastic error, $\widehat{\mathbf{P}} - \mathbf{P}$, whereas the SR coefficients c_1, c_2, c_3 determine how this error transforms into distance estimation error, $\widehat{d} - d$. The SR function is evaluated on the path in question according to the *normalized mean square error* ($\mathsf{NMSE}[\widehat{d}]$) which is the expected relative-error-squared defined in §4 below.

$$\mathsf{NMSE}[\widehat{d}] \;=\; \mathsf{E}\left[\left(\frac{\widehat{d}-d}{d}\right)^2\right] \;. \tag{4}$$

Note that since d is constant, we have $\mathsf{NMSE}[\widehat{d}] = \frac{\mathsf{MSE}[\widehat{d}]}{d^2}$ (where $\mathsf{MSE}[\widehat{d}] \triangleq \mathsf{E}[(\widehat{d}-d)^2]$). In this section we present a general framework for approximating $\mathsf{MSE}[\widehat{d}]$ as a function of the SR coefficients and the matrix $\mathbf{P}$. This approximation is based on computing $\mathsf{MSE}(\widetilde{d})$, where $\widetilde{d}$ is a linear approximation of $\widehat{d}$ which is simpler to handle, as will be detailed soon. We show that as long as $\mathbf{P}$ does not have eigenvalues which are too close to zero, this approximation is valid. Section 4.1 presents the general scheme for computing $\mathsf{MSE}[\widetilde{d}]$, and Section 4.2 provides a specific demonstration for the K2P model.

4.1 Propagation of Error through an SR function.

Distance estimation starts with a pair of sequences S_1, S_2 corresponding to the two taxa situated at the end points of the evolutionary path. The k pairs of aligned sites form k independent samples from the joint distribution of (σ_1, σ_2) defined by the matrix $\mathbf{F}$. The k pairs are sorted according to the 16 possible

$$\left(\mathbf{F} \overset{(0)}{\Longrightarrow}\right) \widehat{\mathbf{F}} \overset{(1)}{\Longrightarrow} \widehat{\mathbf{P}} \overset{(2)}{\Longrightarrow} \widehat{\Lambda} \overset{(3)}{\Longrightarrow} \widehat{\Delta} \overset{(4)}{\Longrightarrow} \widehat{d}$$

Figure 3: **Outline of the process of distance estimation.** The matrix $\widehat{\mathbf{F}}$ consists of the statistics extracted from the taxon sequences. The matrix $\widehat{\mathbf{P}}$ is computed from $\widehat{\mathbf{F}}$ using some optimization criterion (e.g. maximum likelihood). Steps 2-4 implement an invocation of an SR function on $\widehat{\mathbf{P}}$.

types, and the relative frequencies of these types define the matrix $\widehat{\mathbf{F}}$: for every $a, b \in \{\texttt{A},\texttt{G},\texttt{C},\texttt{T}\}$, $\widehat{\mathbf{F}}(a,b)$ is the number of aligned sites of type (a,b) divided by k. Note that the 16 entries of the matrix $k \cdot \widehat{\mathbf{F}}$ are distributed multinomially (see [7]) with parameters $(k, \mathbf{F})$. This means that $\widehat{\mathbf{F}}$ is an unbiased statistical estimators of $\mathbf{F}$ (i.e., for every $a, b \in \{\texttt{A},\texttt{G},\texttt{C},\texttt{T}\}$, $\widehat{\mathbf{F}}(a,b)$ is a random variable s.t. $\mathsf{E}[\widehat{\mathbf{F}}(a,b)] = \mathbf{F}(a,b)$). Furthermore, the variances and covariances of the entries of $\widehat{\mathbf{F}}$ are given by:

$$\mathsf{VAR}[\widehat{\mathbf{F}}(a,b)] \;=\; \frac{\mathbf{F}(a,b)(1-\mathbf{F}(a,b))}{k} \quad ; \quad \mathsf{COV}[\widehat{\mathbf{F}}(a,b), \widehat{\mathbf{F}}(c,d)] \;=\; -\frac{\mathbf{F}(a,b)\mathbf{F}(c,d)}{k} \;. \tag{5}$$

Distance estimation is a (deterministic) mapping of the matrix $\widehat{\mathbf{F}}$ onto $\widehat{d}$. In order to analyze the propagation of error, we describe this mapping as a series of transformations or steps (see Fig. 3):

1. $\widehat{\mathbf{F}}$ is used to obtain $\widehat{\mathbf{P}}$, an estimation of the substitution matrix $\mathbf{P}$. For example, $\widehat{\mathbf{P}}$ can be a matrix from the unified model $\mathcal{M}$ which is most likely to produce $\widehat{\mathbf{F}}$.

2. The three non-trivial eigenvalues $\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3$ of $\widehat{\mathbf{P}}$ are determined using the unifying matrix $\mathbf{U}$ of the substitution model:[3]
$$\widehat{\lambda}_t \;=\; \sum_{s,r=1}^{4} \mathbf{U}^{-1}(t,s)\widehat{\mathbf{P}}(s,r)\mathbf{U}(r,t) \;, \quad t = 1..3 \;. \tag{6}$$

3. The three *basic distance components* are extracted from these eigenvalues by taking logarithms:
$$\widehat{\delta}_t \;\left(= \Delta_t(\widehat{\mathbf{P}})\right) \;=\; -\ln|\widehat{\lambda}_t| \;, \quad t = 1..3 \;. \tag{7}$$

4. The distance estimate $\widehat{d}$ is computed as a linear combination of the basic distance components using the SR coefficients c_1, c_2, c_3:
$$\widehat{d} \;\left(= \Delta(\widehat{\mathbf{P}})\right) \;=\; c_1\widehat{\delta}_1 + c_2\widehat{\delta}_2 + c_3\widehat{\delta}_3 \;. \tag{8}$$

Propagation of error along these steps is analyzed by associating each intermediate step with a vector of random variables: the 16-point vectors $\widehat{F}, \widehat{P}$ correspond to vectorized versions of the matrices $\widehat{\mathbf{F}}, \widehat{\mathbf{P}}$ (resp.), and the 3-point vectors $\widehat{\Lambda}, \widehat{\Delta}$ hold the values obtained in steps 2 and 3 (resp.). Each of these vectors is a statistical estimator of the appropriate model parameters: $\widehat{F}, \widehat{P}$ estimate F, P – vectorized versions of the matrices $\mathbf{F}, \mathbf{P}$; $\widehat{\Lambda} = (\lambda_1(\widehat{\mathbf{P}}), \lambda_2(\widehat{\mathbf{P}}, \lambda_3(\widehat{\mathbf{P}}))$ estimates $\Lambda = (\lambda_1(\mathbf{P}), \lambda_2(\mathbf{P}), \lambda_3(\mathbf{P}))$ and $\widehat{\Delta} = \left(-\ln|\lambda_1(\widehat{\mathbf{P}})|, -\ln|\lambda_2(\widehat{\mathbf{P}})|, -\ln|\lambda_3(\widehat{\mathbf{P}})|\right)$ estimates $\Delta = (-\ln(\lambda_1(\mathbf{P})), -\ln(\lambda_2(\mathbf{P})), -\ln(\lambda_3(\mathbf{P})))$. Each such vector $X \in \{\widehat{F}, \widehat{P}, \widehat{\Lambda}, \widehat{\Delta}\}$ is associated with an expectation vector $E(X)$ s.t. $E(i) = \mathsf{E}[x_i]$ and a covariance matrix $\mathbf{\Sigma}(X)$ s.t. $\mathbf{\Sigma}(i,j) = \mathsf{COV}[x_i, x_j]$. $\mathbf{\Sigma}(\widehat{F})$ is given by §5, and the covariance matrices of $\widehat{P}, \widehat{\Lambda}, \widehat{\Delta}$ (and $\widehat{d}$) are computed by representing each such vector as a linear transformation of its predecessor and applying the following basic lemma (whose proof is omitted):

Lemma 4.1. *Let $\widehat{X}$ be a column vector of n random variables such that $E(\widehat{X}) = X$ (i.e. $\widehat{X}$ is an unbiased statistical estimator of the vector X). Further let $\mathbf{A}$ be a constant $m \times n$ matrix and let B be a constant column-vector of length m. Then for $Y = \mathbf{A}X + B$ and $\widehat{Y} = \mathbf{A}\widehat{X} + B$ it holds that:*

[3] We note that sometimes steps 1 and 2 are combined and the eigenvalues of $\widehat{\mathbf{P}}$ are computed directly from $\widehat{\mathbf{F}}$. For instance, in the K2P model, these eigenvalues can be obtained by the diagonal elements of $\mathbf{U}^{-1}[4\widehat{\mathbf{F}}]\mathbf{U}$.

1. $E(\widehat{Y}) = Y$ *(i.e. $\widehat{Y}$ is an unbiased statistical estimator of Y).*

2. $\mathbf{\Sigma}(\widehat{Y}) = \mathbf{A} \cdot \mathbf{\Sigma}(\widehat{X}) \cdot \mathbf{A}^T$.

We are now left to show how each step of the distance estimation process can be represented by a linear transformation. We consider a common scenario where the computation of $\widehat{P}$ in step 1 is done by a linear transformation of $\widehat{F}$ (other cases will be discussed in Section 6). For instance, the maximum-likelihood estimate of $\widehat{P}$ under the K2P model is given by[4]:

$$\begin{aligned} \widehat{p_{\text{nil}}} &= \widehat{\mathbf{F}}(A,A) + \widehat{\mathbf{F}}(G,G) + \widehat{\mathbf{F}}(C,C) + \widehat{\mathbf{F}}(T,T) \ . \\ \widehat{p_\alpha} &= \widehat{\mathbf{F}}(A,G) + \widehat{\mathbf{F}}(G,A) + \widehat{\mathbf{F}}(C,T) + \widehat{\mathbf{F}}(T,C) \ . \\ \widehat{p_\beta} &= \sum_{a\in\{\text{A,G}\},b\in\{\text{C,T}\}} \frac{1}{2}(\widehat{\mathbf{F}}(a,b) + \widehat{\mathbf{F}}(b,a)) \ . \end{aligned} \tag{9}$$

Note: The formulae above provide a valid K2P substitution matrix $\widehat{\mathbf{P}}$ only when $\widehat{p_\beta} < 0.25$ and $\widehat{p_\alpha} < \widehat{p_{\text{nil}}}$. This is because the eigenvalues of a substitution matrix in the K2P model must be positive. The restriction of positive eigenvalues is common to all substitution models corresponding to a continuous-time Markov process (see Section 2.2). Therefore, in the rest of our analysis (and experiments done in Section 5) we will assume that the eigenvalues of both $\mathbf{P}$ and $\widehat{\mathbf{P}}$ are all positive. If the matrix $\widehat{\mathbf{P}}$ turns out to have a non-positive eigenvalue, the distance estimation process is failed and ignored.

Step 2 of the distance estimation process is clearly linear, as apparent in §6. Therefore, applying Lemma 4.1(1) to the first two steps gives us that $E(\widehat{P}) = P$ and $E(\widehat{\Lambda}) = \Lambda$. Furthermore, the second part of the lemma is applied twice to obtain $\mathbf{\Sigma}(\widehat{\Lambda})$. Step 3 requires a non-linear transformation (since $\ln(x)$ is a non-linear function), so we use the *delta method* [24] to replace it by a linear transformation which provides linear approximations of $\widehat{\Delta}$ and $\widehat{d}$ as follows.

Consider the two-term Taylor expansion of §7 around $\lambda_t = \mathsf{E}[\widehat{\lambda}_t]$ (assuming that $\widehat{\lambda}_t, \lambda_t > 0$):

$$\widehat{\delta}_t = -\ln(\widehat{\lambda}_t) = -\ln(\lambda_t) - \frac{(\widehat{\lambda}_t - \lambda_t)}{\lambda_t} + \frac{(\widehat{\lambda}_t - \lambda_t)^2}{2\xi^2}, \ \text{where } \min\{\lambda_t, \widehat{\lambda}_t\} \le \xi \le \max\{\lambda_t, \widehat{\lambda}_t\}. \tag{10}$$

An approximation $\widetilde{\delta}_t$ of $\widehat{\delta}_t$ is obtained by taking the first two terms in the above expression:

$$\widetilde{\delta}_t \ \triangleq \ -\ln(\lambda_t) - \frac{\widehat{\lambda}_t - \lambda_t}{\lambda_t} \ = \ 1 - \ln(\lambda_t) - \frac{\widehat{\lambda}_t}{\lambda_t}, \quad t = 1..3 \ . \tag{11}$$

Clearly, $\widetilde{\delta}_t$ is linear in $\widehat{\lambda}_t$. The approximation $\widetilde{d}$ of $\widehat{d}$ is then defined as $\widetilde{d} \triangleq c_1\widetilde{\delta}_1 + c_1\widetilde{\delta}_2 + c_1\widetilde{\delta}_3$. After this linearization, Lemma 4.1(2) can be applied twice to yield $\mathbf{\Sigma}(\widetilde{d}) = \mathsf{VAR}[\widetilde{d}]$. Furthermore, due to linearity of expectation (Lemma 4.1(1)), we get that for all $t = 1..3$, $\mathsf{E}[\widetilde{\delta}_t] = -\ln(\lambda_t) = \delta_t$, and $\mathsf{E}[\widetilde{d}] = d$, implying that $\mathsf{VAR}[\widetilde{d}] = \mathsf{MSE}[\widetilde{d}]$.

The end-to-end process of computing $\mathsf{MSE}[\widetilde{d}]$ is summarized in the following scheme:

1. Compute $\mathbf{\Sigma}(\widehat{F})$ using §5.
2. Compute the linear transformation associated with each of the 4 steps of distance estimation: for step 1 use the model-specific formula for $\widehat{P}$ (e.g. §9 for K2P); for step 2 use the unifying matrix $\mathbf{U}_{\mathcal{M}}$ and §6; for step 3 use the approximation formula of §11; and for step 4 use the SR-coefficients.
3. Apply Lemma 4.1(2) to each step, and obtain $\mathsf{VAR}[\widetilde{d}] = \mathsf{MSE}[\widetilde{d}]$.

[4]The proof that these formulae correspond to the maximum likelihood estimate (which is rather straightforward) is not presented here.

This scheme results in an approximation of $\mathsf{MSE}(\widehat{d})$ which is valid when $|\widehat{d}-\widetilde{d}|$ is much smaller than $|\widetilde{d}-d|$. Note that this condition is guaranteed if for $t=1..3$, we have $|\widehat{\delta}_t-\widetilde{\delta}_t| \ll |\widetilde{\delta}_t-\delta_t|$. Let $h_t \stackrel{\triangle}{=} \frac{|\widehat{\lambda}_t-\lambda_t|}{\lambda_t} = |\widetilde{\delta}_t-\delta_t|$. Then $|\widehat{\delta}_t-\widetilde{\delta}_t| = \frac{h_t^2\lambda_t^2}{2\xi^2} \leq \frac{1}{2}\left(\frac{h_t}{1-h_t}\right)^2$ (since $\xi \geq (1-h_t)\lambda_t$). Hence, when h_t is sufficiently small, we indeed have $|\widehat{\delta}_t-\widetilde{\delta}_t| \ll |\widetilde{\delta}_t-\delta_t|$. On the other hand, when h_t is relatively large, $|\widehat{\delta}_t-\delta_t|$ ceases to be a good approximation for $|\widetilde{\delta}_t-\delta_t|$. This typically happens when λ_t is very small. Therefore, the approximation of $\mathsf{MSE}[\widehat{d}]$ by $\mathsf{MSE}[\widetilde{d}]$ becomes less tight if for some $t=1..3$, the eigenvalue λ_t is near-zero and the SR coefficient c_t is non-negligible compared to the other two SR coefficients. This effect, dubbed *'small eigenvalue distortion'*, is observed in the simulations described in Section 5 (see Fig. 6).

4.2 Propagation of error in the K2P model.

In this section we apply the general recipe of Section 4.1 to the K2P model, simplifying the computations when possible. Recall that SR functions in the K2P model are of the form $\Delta(\mathbf{P}) = -\left(c_1\ln(\lambda_1)+c_2\ln(\lambda_2)\right)$, where λ_1,λ_2 are defined in Section 2.3. First we specify $\widehat{F}$, the vector of random variables associated with the observed joint distribution. Then, each step $s=1\ldots4$ is associated with a linear transformation defined by a matrix $\mathbf{A}_s$ and vector B_s, and Lemma 4.1(2) is applied to obtain the appropriate covariance matrix for that step.

The vector $\widehat{F}$: Assume that entries 1-4 of $\widehat{F}$ correspond to nil-substitutions (diagonal entries of $\widehat{\mathbf{F}}$), entries 5-8 correspond to transition-type substitutions, and entries 9-16 correspond to transversion-type substitutions. Then according to §5, we get:

$$\mathbf{\Sigma}(\widehat{F}) = \frac{1}{16k}\begin{pmatrix} p_{\text{nil}}(4\mathbf{I}_4 - p_{\text{nil}}\mathbf{J}_{4,4}) & -p_{\text{nil}}p_\alpha\mathbf{J}_{4,4} & -p_{\text{nil}}p_\beta\mathbf{J}_{4,8} \\ -p_{\text{nil}}p_\alpha\mathbf{J}_{4,4} & p_\alpha(4\mathbf{I}_4 - p_\alpha\mathbf{J}_{4,4}) & -p_\alpha p_\beta\mathbf{J}_{4,8} \\ -p_{\text{nil}}p_\beta\mathbf{J}_{8,4} & -p_\alpha p_\beta\mathbf{J}_{8,4} & p_\beta(4\mathbf{I}_8 - p_\beta\mathbf{J}_{8,8}) \end{pmatrix} \tag{12}$$

(where $\mathbf{I}_n$ is the $n\times n$ identity matrix and $\mathbf{J}_{n,m}$ is the $n\times m$ all-one matrix).

Step 1: The three parameters $\widehat{p_{\text{nil}}}, \widehat{p_\alpha}, \widehat{p_\beta}$ (defining the matrix $\widehat{\mathbf{P}}$) are computed as maximum-likelihood estimates of $p_{\text{nil}}, p_\alpha, p_\beta$, (which define $\mathbf{P}$) according to §9, which implies $\widehat{P} = \mathbf{A}_1\widehat{F} + B_1$, where $\widehat{P} = (\widehat{p_{\text{nil}}}\ \widehat{p_\alpha}\ \widehat{p_\beta})^T$ and

$$\mathbf{A}_1 = \begin{pmatrix} 1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2} \end{pmatrix} \quad ; \quad B_1 = \begin{pmatrix} 0\\0\\0 \end{pmatrix} \quad .$$

Plugging $\mathbf{A}_1$ and $\mathbf{\Sigma}(\widehat{F})$ in Lemma 4.1(2) results in:

$$\mathbf{\Sigma}(\widehat{P}) = \frac{1}{k}\begin{pmatrix} p_{\text{nil}}(1-p_{\text{nil}}) & -p_{\text{nil}}p_\alpha & -p_{\text{nil}}p_\beta \\ -p_{\text{nil}}p_\alpha & p_\alpha(1-p_\alpha) & -p_\alpha p_\beta \\ -p_{\text{nil}}p_\beta & -p_\alpha p_\beta & \frac{1}{2}p_\beta(1-2p_\beta) \end{pmatrix} \tag{13}$$

Step 2: The eigenvalues $\widehat{\lambda}_1, \widehat{\lambda}_2$ of the matrix $\widehat{\mathbf{P}}$ are computed. Recall that $\widehat{\lambda}_1 = 1-4\widehat{p_\beta}$, $\widehat{\lambda}_2 = \widehat{p_{\text{nil}}} - \widehat{p_\alpha}$ (see Fig. 2). Hence, $\widehat{\Lambda} = \mathbf{A}_2\widehat{P} + B_2$, where

$$\mathbf{A}_2 = \begin{pmatrix} 0&0&-4 \\ 1&-1&0 \end{pmatrix} \quad ; \quad B_2 = \begin{pmatrix} 1\\0 \end{pmatrix} \quad .$$

Plugging $\mathbf{A}_2$ and $\mathbf{\Sigma}(\widehat{P})$ in Lemma 4.1(2), and expressing through λ_1,λ_2 results in:

$$\begin{aligned} \mathbf{\Sigma}(\widehat{\Lambda}) &= \frac{1}{k}\begin{pmatrix} 8p_\beta(1-2p_\beta) & 4p_\beta(p_{\text{nil}}-p_\alpha) \\ 4p_\beta(p_{\text{nil}}-p_\alpha) & p_{\text{nil}}(1-p_{\text{nil}}) + p_\alpha(1-p_\alpha) + 2p_{\text{nil}}p_\alpha \end{pmatrix} \\ &= \frac{1}{k}\begin{pmatrix} 1-\lambda_1^2 & (1-\lambda_1)\lambda_2 \\ (1-\lambda_1)\lambda_2 & \frac{1}{2}(1+\lambda_1) - \lambda_2^2 \end{pmatrix} \end{aligned} \tag{14}$$

Step 3: The linear approximation $\widetilde{\Delta}$ of $\widehat{\Delta}$ is computed by §11, which gives us $\widetilde{\Delta} = \mathbf{A}_3\widehat{\Lambda} + B_3$, where

$$\mathbf{A}_3 = \begin{pmatrix} -\frac{1}{\lambda_1} & 0 \\ 0 & -\frac{1}{\lambda_2} \end{pmatrix} \quad ; \quad B_3 = \begin{pmatrix} 1-\ln(\lambda_1) \\ 1-\ln(\lambda_2) \end{pmatrix} \quad .$$

Plugging $\mathbf{A}_3$ and $\mathbf{\Sigma}(\widehat{\Lambda})$ in Lemma 4.1(2), and expressing through α, β results in:

$$\mathbf{\Sigma}(\widetilde{\Delta}) \;=\; \frac{1}{k}\begin{pmatrix} \frac{1-\lambda_1^2}{\lambda_1^2} & \frac{(1-\lambda_1)}{\lambda_1} \\ \frac{(1-\lambda_1)}{\lambda_1} & \frac{(1+\lambda_1)}{2\lambda_2^2} - 1 \end{pmatrix} \;=\; \frac{1}{k}\begin{pmatrix} e^{8\beta}-1 & e^{4\beta}-1 \\ e^{4\beta}-1 & \frac{1}{2}e^{4\alpha}(e^{4\beta}+1)-1 \end{pmatrix} \tag{15}$$

Step 4: Finally, consider $\widetilde{d}$ (the linear approximation of $\widehat{d}$) which is obtained by $\widetilde{d} \;=\; \mathbf{A}_4\widetilde{\Delta} + B_4$, where

$$\mathbf{A}_4 = \begin{pmatrix} c_1 & c_2 \end{pmatrix} \quad ; \quad B_4 = 0$$

Plugging $\mathbf{A}_4$ and $\mathbf{\Sigma}(\widetilde{\Delta})$ in Lemma 4.1 gives us the following expression for the variance of $\widetilde{d}$:

$$\mathsf{MSE}[\widetilde{d}] = \mathsf{VAR}[\widetilde{d}] \;=\; \frac{1}{k}\left(c_1^2(e^{8\beta}-1) + 2c_1c_2(e^{4\beta}-1) + \frac{1}{2}c_2^2(e^{4\alpha}(e^{4\beta}+1)-2)\right). \tag{16}$$

5 Finding an Optimal SR function for a single path.

In this section we use the analysis of Section 4 for the following task: given an alignment of two sequences which evolved according to the K2P model, find an SR function which (nearly) minimizes the propagation of error in estimation of the distance between the corresponding two taxa. Finding such a function seems a basic ingredient in any method for choosing an SR function which is 'best' for a given taxon set (of any size). In Section 5.1 we show how to select an SR function according to the (unknown) model parameters α, β associated with the path connecting the two taxa. In Section 5.2 we demonstrate the potential reduction in error of this approach. Then, in Section 5.3, we show how a similar approach can be used to fit the SR function to the observed taxon sequences rather than to the unknown model parameters.

5.1 Finding the SR coefficients which minimize $\mathsf{NMSE}[\widetilde{d}]$.

The formula for $\mathsf{MSE}[\widetilde{d}]$ in §16 provides a linear approximation for the actual mean square error rate, $\mathsf{MSE}[\widehat{d}]$, in the K2P model. The *normalized* mean square error $\mathsf{NMSE}[\widehat{d}] = \frac{\mathsf{MSE}[\widehat{d}]}{d^2}$, is thus approximated by:

$$\begin{aligned} \mathsf{NMSE}[\widetilde{d}] = \frac{\mathsf{MSE}[\widetilde{d}]}{d^2} \quad &= \quad \frac{c_1^2(e^{8\beta}-1) + 2c_1c_2(e^{4\beta}-1) + \frac{1}{2}c_2^2(e^{4\alpha}(e^{4\beta}+1)-2)}{k(4c_1\beta + 2c_2(\alpha+\beta))^2} \\ &\overset{(c=\frac{c_1}{c_2})}{=} \quad \frac{(e^{8\beta}-1)\;c^2 \;+\; 2(e^{4\beta}-1)\;c \;+\; \frac{1}{2}(e^{4\alpha}(e^{4\beta}+1)-2)}{k(4\beta\;c \;+\; 2(\alpha+\beta))^2} \end{aligned} \tag{17}$$

Note that $\mathsf{NMSE}[\widetilde{d}]$ depends only on the *ratio* between the SR coefficients $c = \frac{c_1}{c_2}$, which implies that proportional SR functions are equivalent under this criterion. Therefore, although $DIM(\mathcal{SR}_{\mathsf{K2P}}) = 2$, the optimization of the SR function is only one-dimensional.

Lemma 5.1. *Let $\alpha, \beta > 0$ be the K2P parameters associated with a certain evolutionary path. Then the SR-function which minimizes $\mathsf{NMSE}[\widetilde{d}]$ along this path is given by SR coefficients c_1, c_2, s.t.*

$$\frac{c_1}{c_2} \;=\; c_{OPT} \;\triangleq\; \frac{\frac{e^{4\beta}+1}{e^{4\beta}-1}(e^{4\alpha}-1)\beta \;-\; \alpha}{(e^{4\beta}-1)\beta \;+\; (e^{4\beta}+1)\alpha} \;. \tag{18}$$

Proof Outline. [5] In order to find the value of c which minimizes $\mathsf{NMSE}[\widetilde{d}]$, the partial derivative of §17 is taken with respect to c.

$$\frac{\partial(\mathsf{NMSE}[\widetilde{d}])}{\partial c} \;=\; \frac{\big((e^{4\beta}-1)^2\beta \;+\; (e^{8\beta}-1)\alpha\big)\;\; c \;-\;\; \big((e^{4\beta}+1)(e^{4\alpha}-1)\beta \;-\; (e^{4\beta}-1)\alpha\big)}{2k(2\beta\; c \;+\; \alpha+\beta)^3} \quad .$$

It is easy to see that $c_{\mathtt{OPT}}$ of §18 is the unique point where this derivative is nullified. The proof is completed by showing that (a) $c_{\mathtt{OPT}}$ is positive (implying that it corresponds to a valid SR function), and (b) $c_{\mathtt{OPT}}$ is a local minimum point of $\mathsf{NMSE}[\widetilde{d}]$. □

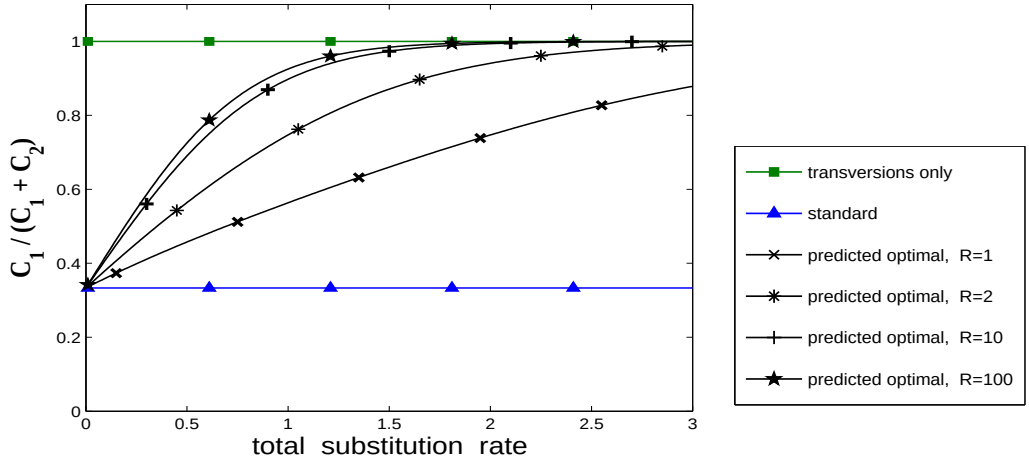

Figure 4: **Predicted optimal SR coefficients.** The graph above contains plots of $\frac{c_1}{c_1+c_2}$ for various SR-functions. The horizontal lines correspond to the standard SR function for K2P (where $c_2 = 2c_1$) and the SR function considering only transversions (where $c_2 = 0$). The other plots correspond to predicted optimal SR coefficients computed using §18 and corresponding to different ti/tv ratios R (1, 2, 10, and 100). The X axis corresponds to the total substitution rate which is $\alpha + 2\beta$.

An SR function using SR coefficients as in Lemma 5.1 is referred to as the *predicted optimal SR function* for the parameters α, β. A closer look at §18 shows that whenever $\alpha \geq \beta$, $c_{\mathtt{OPT}} \geq \frac{1}{2}$, and when $\alpha = \beta$ (implying the JC model), $c_{\mathtt{OPT}} = \frac{1}{2}$. Note that $c_{\mathtt{OPT}} = \frac{1}{2}$ corresponds to the logdet SR-function, which is used by the standard distance formula of the K2P model. It can also be shown that $c_{\mathtt{OPT}}$ converges to $\frac{1}{2}$ whenever the total substitution rate $(\alpha + 2\beta)$ converges to 0, regardless of the ti/tv ratio R (see also Fig. 4). Similarly, if $\alpha > \beta$ and the total substitution rate goes to ∞, so does $c_{\mathtt{OPT}}$. Note that $c_{\mathtt{OPT}} = \infty$ $(c_2 = 0)$ corresponds to the SR function which counts only transversions. Intuitively, when $\alpha > \beta$ and the total substitution rate is large, the transversion count is less noisy than the total substitutions count, and hence the weight c_2 given to transition events is reduced. An obvious consequence of this analysis is that the standard distance estimation formula of K2P (using the logdet SR function) is predicted to be near optimal only in the extreme cases where either the total substitution rate is very small or the ti/tv ratio R is close to 0.5 (i.e., $\alpha \cong \beta$).

5.2 Predicted and observed reduction in error.

The first thing we check about the predicted optimal SR function is the extent of reduction in error it is predicted to yield. The error rate is measured by the normalized *root* mean square error (NRMSE) which is the square root of the NMSE. $\mathsf{NRMSE}[\widetilde{d}]$, given by the square-root of §17, is used as the *predicted error rate* of an SR function, whereas $\mathsf{NRMSE}[\widehat{d}]$ denotes the actual error rate. The predicted optimal SR function, given by §18, is compared under various settings of the model parameters α, β

[5] The complete proof of Lemma 5.1 appears in Appendix A.

to the standard (or logdet) SR function (where $c_2 = 2c_1$) and the 'transversions-only' SR function (where $c_2 = 0$). The graphs in Fig. 5 describe the results of this comparison for $R = 2$ and $R = 10$. Consider the extreme cases first. When the total substitution rate is high, the standard SR function is predicted to be extremely noisy, and the optimal SR function (which converges to 'transversions only') is predicted to be much more accurate. In the other extreme, when the total substitution rate is very small, the standard SR function is predicted to be near optimal, and the 'transversions-only' SR function is predicted to be much less accurate. Comparing the two graphs in Fig. 5 demonstrates that these differences in predicted performance are enhanced when the ti/tv ratio R grows. It is also worth noting that for intermediate substitution rates, the optimal SR function is predicted to perform better than each of the other two.

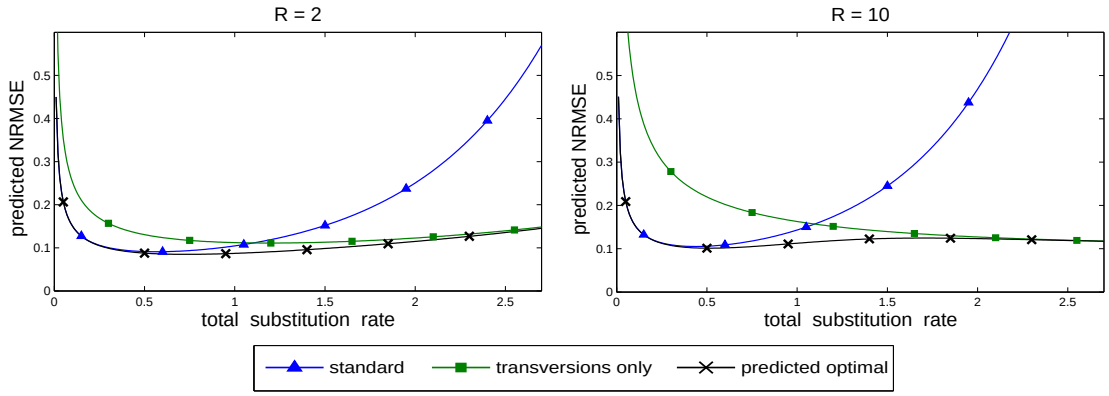

Figure 5: **Predicted error rates ($\mathsf{NRMSE}[\widetilde{d}]$).** The predicted error rates are given by the square-root of §17, where the sequence length k is taken to be 500, and the ti/tv ratio R is taken to be 2 (left graph) and 10 (right graph). The X-axis corresponds to the total substitution rate ($\alpha + 2\beta$). Each graph contains three plots: one corresponding to the standard SR function (where $c_2 = 2c_1$), one corresponding to the 'transversions only' SR function (where $c_2 = 0$), and one corresponding to the predicted optimal SR function (where $c_1/c_2 = c_{\mathtt{OPT}}$ as in §18). Note that the predicted optimal SR function changes with the values of α and β along the x-axis (unlike the other two functions which are constant).

Next, we compare the predicted error rates ($\mathsf{NRMSE}[\widetilde{d}]$) with actual error rates ($\mathsf{NRMSE}[\widehat{d}]$) using empirical observations. The empirically observed $\mathsf{NRMSE}[\widehat{d}]$ was computed using 10,000 repetitions of sequence evolution simulations for each sampled value of the model parameters α, β. Results are shown in Fig. 6 for $R = 2$, but a similar picture is observed for other values of R. This comparison indicates that $\mathsf{NRMSE}[\widetilde{d}]$ approximates $\mathsf{NRMSE}[\widehat{d}]$ very tightly in most cases. The cases where there is deviation can be explained by 'small eigenvalue distortion'. Recall that small eigenvalue distortion occurs when an SR function gives non-negligible weight to a base function $\Delta_t = -\ln(\lambda_t)$ where the eigenvalue λ_t is very small. Now, since in these simulations λ_2 is always smaller than λ_1 (implied by the fact that $\beta < \alpha$), then the base function Δ_2 is more prone to distortion than Δ_1. In fact, it is apparent that when the total rate does not exceed 2, the 'transversions-only' SR function (which is proportional to Δ_1) shows no distortion (distortion there appears only for higher rates). On the other hand, the standard SR function (which puts $\frac{2}{3}$ of its weight on Δ_2) does show significant distortion. Distortion of lower extent is also observed for the predicted optimal SR function. Although the weight it puts on Δ_2 reduces as the total substitution rate increases, it is not reduced enough to avoid distortion altogether.

There is another consequence of 'small eigenvalue distortion' visible in this graph, other than deviation of the predicted error rate from the observed one. Small eigenvalues (especially λ_2) also cause the unstable behavior of the empirically observed $\mathsf{NRMSE}[\widehat{d}]$. This behavior is caused by simulation instances which produce values of $\widehat{\lambda}_2$ that are extremely small. These outliers imply a very high estimation error ($(\widehat{d} - d)^2$), which significantly influences on the observed NRMSE, resulting in the

observed spikes in the plots. One apparent flaw of the predicted optimal SR functions is that, unlike the 'transversions-only' SR function, it is rather sensitive to these outliers (leading after some point to inferior performance). A more basic problem in using this SR function is that its SR coefficients depend on the *unknown* model parameters α, β. In Section 5.3 we present a solution to the second problem which also (somewhat surprisingly) provides a solution to the first.

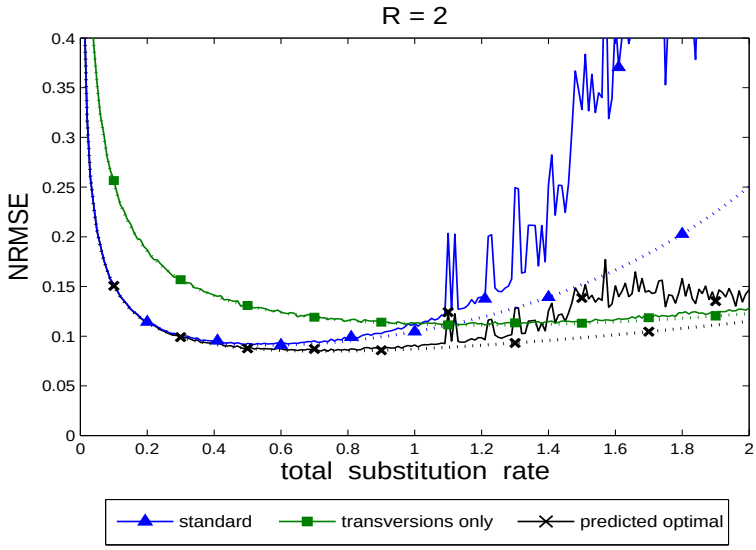

Figure 6: **Predicted vs. empirically observed NRMSE.** The predicted values of NRMSE$[\widetilde{d}]$ (plotted in dotted lines) are given by the square-root of §17. The empirically observed error rates (plotted in solid lines) are computed using 10,000 independent simulations of sequence evolution (with $R = 2$ and $k = 500$). Simulations in which one of the eigenvalues of $\widehat{\mathbf{P}}$ ($\widehat{\lambda}_1$ or $\widehat{\lambda}_2$) was non-positive were discarded and replaced by 'valid' simulations.

5.3 Fitting the SR function to the input sequences.

Recall that our goal is to fit an SR function to the *observed taxon sequences* (rather than to the unknown model parameters). A natural way to achieve this goal is to use an SR function which is optimal w.r.t. the model parameters estimated from the input sequences. In the case of K2P, this means obtaining the SR coefficients by plugging in $\widehat{\alpha}$ and $\widehat{\beta}$ (which correspond to the matrix $\widehat{\mathbf{P}}$) in §18. We refer to this as the *adaptive optimal SR function*, as it adapts itself to the observed sequences. Although this adaptive strategy seems natural, it is not a priori obvious that it yields reasonably accurate distance estimates. Fig. 7 contains results of experiments which test the accuracy of this approach. Again, 10,000 repetitions were used to compute the empirical values of the NRMSE. Three strategies of distance estimation were considered: the *adaptive strategy* suggested above; the *predicted optimal* SR function obtained by plugging in §18 the true model parameters α, β (usually unknown to the user); and the *a posteriori optimal* SR function which is the 'non-adaptive' SR function which in hindsight minimizes the observed NRMSE over the 10,000 simulations. The SR coefficients of the latter function were set by scanning the ratio $\frac{c_1}{c_1+c_2}$ in the interval $[0, 1]$ in steps of 0.01, and choosing the ratio which a posteriori leads to minimal observed NRMSE. Note that in the adaptive strategy the SR function changes throughout the 10,000 simulations according to their outcome. Hence, in this strategy (unlike the other two) the reference distance d changes from one simulation to another.

One observation which was already apparent in Fig. 6 is that the SR function predicted to be optimal (through §18) is not always optimal. Indeed, in some cases the a posteriori optimal SR function is much more accurate. However, the most important (and encouraging) observation is that the adaptive strategy, which is the only one that can be implemented in practice, outperforms even the a posteriori

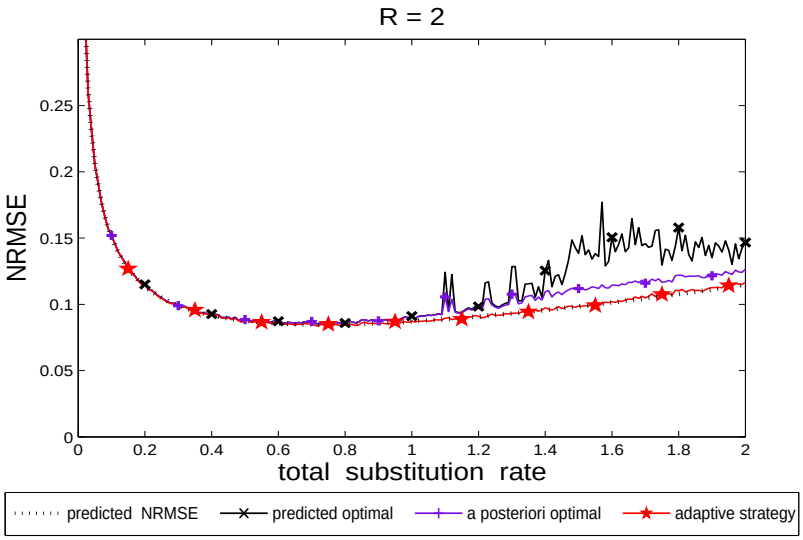

Figure 7: **Fitting SR functions to the input sequences.** The solid plots describe the empirically observed NRMSE of three distance estimation strategies: the predicted optimal SR function, obtained by plugging α, β in §18; the a posteriori optimal SR function which is the 'non-adaptive' one that in hindsight has the minimal error rate; and the adaptive strategy obtained by plugging $\widehat{\alpha}, \widehat{\beta}$ in §18. Computations were done using 10,000 independent simulations of sequence evolution with $R = 2$ and $k = 500$. The dotted plot corresponds to $\mathsf{NRMSE}[\widetilde{d}]$ of the predicted optimal SR function (as in Fig. 6).

optimal SR function. In fact, the adaptive strategy seems to behave very much as we predict the optimal SR function to behave (dotted plot in Fig. 7). This is probably due to the fact that the adaptive strategy is much less sensitive to outliers than the other strategies: when it encounters an outlier in which $\widehat{\alpha}$ is extremely large (and hence also very 'noisy'), it decreases c_2 in a way which significantly reduces the contribution of $\widehat{\alpha}$ to the estimated distance in that simulation. In this sense, the adaptive SR function is self-correcting. The comparison of the adaptive strategy to the a posteriori optimal SR function indicates that the adaptive strategy actually outperforms *any constant* SR function. This strong result was consistently observed in other settings as well.

5.3.1 A Bayesian approach.

One possible way to explain the good performance of the adaptive approach is to view it through a Bayesian setting [9]. In the Bayesian setting, the observed taxon sequences induce a probability distribution over the space of all possible substitution matrices in the substitution model $\mathcal{M}$. Hence, the estimated distance $\widehat{d}$ is considered to be constant and the reference distance d is considered to be random. For instance, assuming that, a priori, all K2P model parameters are equally likely (i.e., a *flat prior*), sequences with observed parameters $(\widehat{p_{\mathrm{nil}}}, \widehat{p_\alpha}, \widehat{p_\beta})$ imply an a posterior probability distribution for $(p_{\mathrm{nil}}, p_\alpha, 2p_\beta)$ which is the Dirichlet distribution [7] with parameters $(k\widehat{p_{\mathrm{nil}}} + 1, k\widehat{p_\alpha} + 1, 2k\widehat{p_\beta} + 1)$. Now denote $\overline{p_{\mathrm{nil}}} = \frac{k\widehat{p_{\mathrm{nil}}}+1}{k+3}$, $\overline{p_\alpha} = \frac{k\widehat{p_\alpha}+1}{k+3}$, $\overline{p_\beta} = \frac{k\widehat{p_\beta}+\frac{1}{2}}{k+3}$. Then the expectation of $(p_{\mathrm{nil}}, p_\alpha, p_\beta)$ under this posterior distribution is $(\overline{p_{\mathrm{nil}}}, \overline{p_\alpha}, \overline{p_\beta})$, and the covariance matrix of these random variable can also be easily expressed using $k, \overline{p_{\mathrm{nil}}}, \overline{p_\alpha}, \overline{p_\beta}$. Consequently, a similar process to the one detailed in Section 4.2 can be applied to approximate $\mathsf{E}[(\widehat{d} - d)^2]$ via serial application of Lemma 4.1 (where d is the random variable). The expression obtained is actually very similar to the one in §16:

$$\mathsf{MSE}[d] \;\approx\; \frac{1}{k+4}\left(c_1^2(e^{8\overline{\beta}} - 1) + 2c_1c_2(e^{4\overline{\beta}} - 1) + \frac{1}{2}c_2^2(e^{4\overline{\alpha}}(e^{4\overline{\beta}} + 1) - 2)\right). \tag{19}$$

The rate parameters $\overline{\alpha}, \overline{\beta}$ are the ones associated with the K2P substitution matrix with entries $\overline{p_{\text{nil}}}, \overline{p_\alpha}, \overline{p_\beta}$. The predicted optimal SR function in this setting is thus obtained by plugging $\overline{\alpha}, \overline{\beta}$ in §18. Now, since $\overline{\alpha}, \overline{\beta}$ are very close to $\widehat{\alpha}, \widehat{\beta}$ for $k \gg 1$, (implied by the fact that $\overline{p_{\text{nil}}}, \overline{p_\alpha}, \overline{p_\beta}$ are very close to $\widehat{p_{\text{nil}}}, \widehat{p_\alpha}, \widehat{p_\beta}$), this ('Bayesian') SR function is indeed very similar to the one obtained by our adaptive strategy. The Bayesian approach can be useful if we wish to introduce prior beliefs on the space of substitution matrices in $\mathcal{M}$ which are different from the flat (equiprobable) one. The choice of SR function is then made according to the combined information in these priors and the observed sequences.

6 Discussion.

In this paper we introduced an adaptive approach for selecting distance functions on sequence data. This was done by introducing the concept of SR functions, which define distances under stochastic substitution models. We showed how SR functions can be constructed using the generalized logdet technique, and characterized the sets of SR functions available for certain types of substitution models. Then we presented a general framework for analyzing the stochastic noise introduced by the computation of SR functions from input sequences (which evolved according to the assumed model). Finally, we used this analysis for finding near optimal distance functions for the K2P model, and introduced simulations which demonstrated the advantages of these functions over the SR functions which are commonly used for this model.

Our analysis and experiments focused on the problem of finding the best SR function for a given evolutionary path. We argue that this result provides a basic tool for significantly improving the performance of distance-based phylogenetic reconstruction methods. Preliminary evidence for this can be found in experiments we ran over quartet trees, which were already mentioned in Section 1. These experiments considered symmetric quartets in different scales (as described in Fig. 1). The main objective was to test the effectiveness of various distance formulae in reconstructing quartets, using the common four-point method (see Section1 for more details on the experimental setting). Along side the standard K2P formula and the 'transversions only' formula (which both use non-adaptive SR functions), we tested an adaptive formula which we call 'max-optimal'. This formula is based on an adaptive SR function defined as follows: (1) $\widehat{\alpha}_{ij}, \widehat{\beta}_{ij}$ are estimated for every taxon pair $\{i,j\} \subseteq \{\text{A,B,C,D}\}$; (2) the adaptive optimal coefficient-ratio c_{ij} is obtained for every pair by plugging $\widehat{\alpha}_{ij}, \widehat{\beta}_{ij}$ in §18; (3) the chosen adaptive SR function is the one using SR coefficients c_1, c_2 s.t. $\frac{c_1}{c_2} = \max\{c_{ij}\}$. The motivation behind this heuristic solution is to minimize the maximal noise within the 6 estimated distances: the SR coefficients chosen in the third step are optimal for the longest observed path in the quartet, which also induces the maximal expected noise.

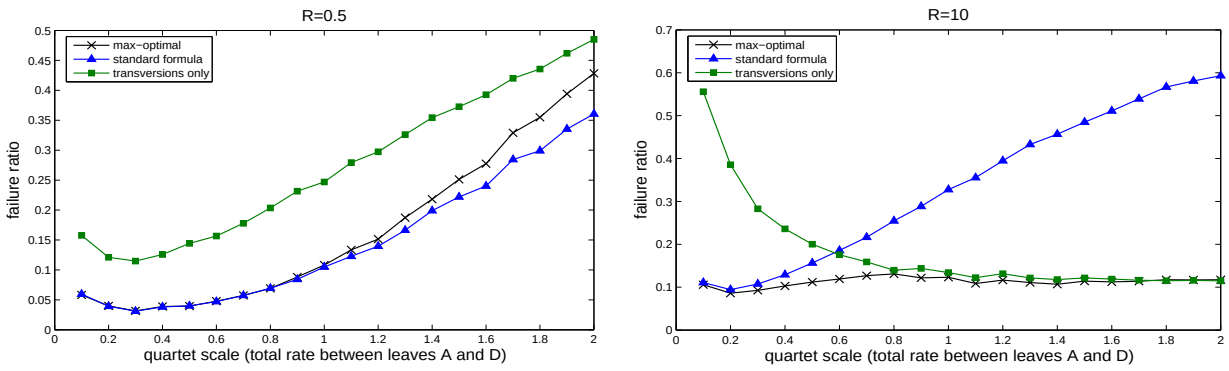

Figure 8: **Reconstructing Symmetric Quartets.** The performance of various distance formulae was tested in reconstructing symmetric quartets using FPM. These graph, which are a sequel to the one in Fig. 1, present results for simulations done using ti/tv ratios of $R = \frac{1}{2}, 10$.

Although 'max-optimal' is purely heuristic, it yields very good results in our experiments. Fig. 1 presents results of simulations where the ti/tv ratio R is 2, and Fig. 8 presents additional results obtained for $R = \frac{1}{2}, 10$. We note that similar results were obtained for other values of R and for non-symmetric quartets. As already mentioned in Section 1, the standard formula performs well for small-scale (or short) quartets, and the 'transversions only' formula preforms well on large-scale (or long) quartets. Our 'max-optimal' heuristic outperforms both of them in most cases. The only clear exception is when the ti/tv ratio is $\frac{1}{2}$ (implying th JC model). Recall that when $R = \frac{1}{2}$, the logdet SR function (used by the standard formula) yields minimal expected noise for *all evolutionary paths* (regardless of their total rate). Hence, in this case we expect the standard formula to perform optimally, regardless of the scale of the quartet. The 'max optimal' heuristic provides a coarse fit of the SR function to the scale of the quartet (or tree, in general), but it is probably not the optimal adaptive strategy. Searching for better strategies remains the main open question from this work. This and other future research topics suggested by our results are discussed below.

- Lemma 3.6 characterizes some substitution models in which all SR functions are spanned by generalized logdet functions. It is interesting to figure out whether this result can be generalized for other substitution models. This question is closely related to the following well studied problem: what are the real valued functions f for which $f(\mathbf{AB}) = f(\mathbf{A})f(\mathbf{B})$ for *all* non-singular matrices $\mathbf{A}, \mathbf{B}$ (see [1], pp. 349-353). The main difference is that in our context, the equality is required to hold only for limited subsets of matrices.

- Our framework for analyzing the stochastic error introduced by SR functions is based on the assumption that the entries of the substitution matrix $\widehat{\mathbf{P}}_{ij}$ which is most likely to produce a given sequence-pair S_i, S_j, are obtained by linear transformations of the observed joint distribution $\widehat{\mathbf{F}}_{ij}$. This assumption is not valid for some common (and unified) models, (namely ones in which the stationary distribution is non-uniform, e.g., the Tamura Nei model). In such cases a different approach may be needed, or alternatively good SR functions should be found by heuristic methods.

- Our quartet experiments demonstrate the potential of the adaptive approach in improving the accuracy of distance based methods. This suggests that finding good (near optimal) SR functions for more general substitution models and/or larger sets of input sequences is potentially of high practical importance. Similarly, modifying existing distance based reconstruction algorithms so that they select SR functions which are good for the given input has the potential to significantly improve their performance.

- A subcase of the previous item which is of special interest is that of non-homogeneous model trees. When the model tree is homogeneous, the different metrics induced on the tree by different SR functions are all proportional to each other. Hence, the only thing that makes a difference in that case is the noise induced by the SR function (normalized by some scaling factor). However, when the model tree is non-homogeneous, then different SR functions might induce non-proportional metrics. This further enhances the difference between different SR functions.

Acknowledgements

We thank Danny Geiger for pointing [1] to us. The second author would like to thank Mike Hendy, John Sved and Marianne Frommer for interesting discussions, which inspired parts of this research.

References

[1] J. Aczél. *Functional equations and their applications.* Academic press, 1966.

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

[3] P. Buneman. The recovery of trees from measures of dissimilarity. *Mathematics in the Archeological and Historical Sciences*, pages 387–395, 1971.

[4] J. Cavender. Taxonomy with confidence. *Math Biosci*, 40:271–280, 1978.

[5] C. Daskalakis, E. Mossel, and S. Roch. Phylogenies without branch bounds: Contracting the short, pruning the deep. Extended abstract to appear in proceedings of RECOMB, 2009.

[6] P. Erdos, M. Steel, L. Szekely, and T. Warnow. A few logs suffice to build (almost) all trees (I). *Random Structures and Algorithms*, 14:153–184, 1999.

[7] M. Evans, N. Hastings, and B. Peacock. *Statistical distributions*. Wiley interscience, 3rd edition, 2000.

[8] J. Farris. A probability model for inferring evolutionary trees. *Systematic Zoology*, 22:250–256, 1973.

[9] I. Gronau. Bayesian analysis of distance estimation in the K2P model. Technical Report CS-2009-06, Technion, March 2009.
http://www.cs.technion.ac.il/users/wwwb/cgi-bin/tr-info.cgi/2009/CS/CS-2009-06.

[10] I. Gronau and S. Moran. Neighbor joining algorithms for inferring phylogenies via LCA-distances. *J Comp Biol*, 14(1):1–15, 2007.

[11] I. Gronau, S. Moran, and S. Snir. Fast and Reliable Reconstruction of Phylogenetic Trees with Very Short Edges. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2008.

[12] X. Gu and W. H. Li. Estimation of evolutionary distances under stationary and nonstationary models of nucleotide substitution. *Proc Natl Acad Sci*, 95:5899–5905, 1998.

[13] M. Hasegawa, H. Kishino, and T. Yano. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*, 22(2):160–174, October 1985.

[14] Penny D. Steel M. Hendy, U. A discrete Fourier analysis for evolutionary trees. *Proc Natl Acad Sci*, 91(8):3339–3343, april 1994.

[15] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge, 1999.

[16] D. Huson, S. Nettles, and T. Warnow. Disk-Covering, a fast-converging method for phylogenetic tree reconstruction. *J Comp Biol*, 6:369–386, 1999.

[17] T. Jukes and C. Cantor. Evolution of protein molecules. In H.N. Munro, editor, *Mammalian Protein Metabolism*, pages 21–132. Academic Press, New York, 1969.

[18] M. Kimura. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16(2):111–120, December 1980.

[19] J. Lake. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc Natl Acad Sci*, 91:1455–1459, 1994.

[20] P. Lio and N. Goldman. Models of molecular evolution and phylogeny. *Genome Research*, 8(12):1233–1244, dec 1998.

[21] P. Lockhart, M. Steel, M. Hendy, and D. Penny. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol*, 11(4):605–612, 1994.

[22] E. Mossel. Phase transitions in phylogeny. *Trans Amer Math Soc*, 356:2379–2404, 2004.

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

[24] Gary W. Oehlert. A note on the delta method. *The American Statistician*, 46(1):27–29, 1992.

[25] A. Papoulis and S. U. Pillali. *Probability, Random Variables andstochastic processes*. McGraw Hill, 4th edition, 2002.

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

[27] S. Sattath and A. Tversky. Additive similarity trees. *Psychometrica*, 42(3):319–345, 1977.

[28] M. Steel. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl Math Lett*, 7(2):19–24, march 1994.

[29] K. Tamura and M. Nei. Estimation of the number of nucleotide substitutions in the control region of mitochondrial dna in humans and chimpanzees. *Mol Biol Evol*, 10(3):512–526, May 1993.

[30] S. Tavaré. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17:57–86, 1986.

A Proof of Lemma 5.1.

Lemma 5.1. *Let $\alpha, \beta > 0$ be the K2P parameters associated with a certain evolutionary path. Then the SR-function which minimizes $\mathsf{NMSE}[\widetilde{d}]$ along this path is given by SR coefficients c_1, c_2, s.t.*

$$\frac{c_1}{c_2} \;=\; c_{\mathtt{OPT}} \;\triangleq\; \frac{\frac{e^{4\beta}+1}{e^{4\beta}-1}(e^{4\alpha}-1)\beta \;-\; \alpha}{(e^{4\beta}-1)\beta \;+\; (e^{4\beta}+1)\alpha} \tag{20}$$

Proof. Any SR function which minimizes $\mathsf{NMSE}[\widetilde{d}]$ is given by SR coefficients c_1, c_2 s.t. $c = \frac{c_1}{c_2}$ minimizes the expression for $\mathsf{NMSE}[\widetilde{d}]$ given in §17:

$$\mathsf{NMSE}[\widetilde{d}] \;=\; \frac{(e^{8\beta}-1)\;c^2 \;+\; 2(e^{4\beta}-1)\;c \;+\; \frac{1}{2}(e^{4\alpha}(e^{4\beta}+1)-2)}{k(4\beta\;c \;+\; 2(\alpha+\beta))^2}\;.$$

Before turning to show that $c_{\mathtt{OPT}}$ is the global minimum of $\mathsf{NMSE}[\widetilde{d}]$, we show that $c_{\mathtt{OPT}} > 0$, implying that it corresponds to a *valid* SR function. Since the denominator of §20 is clearly positive, all we have to do is show that the numerator is positive. This is established by the fact that $e^{4\alpha} - 1 > 4\alpha$ for all $\alpha > 0$, and $\frac{e^{4\beta}+1}{e^{4\beta}-1}\beta > \frac{1}{2}$ for all $\beta > 0$, which imply together that the numerator is always greater than α. Now in order to find the minimal point of $\mathsf{NMSE}[\widetilde{d}]$ as a function of c, we take the partial derivative of $\mathsf{NMSE}[\widetilde{d}]$ according to c:

$$\begin{aligned}
\frac{\partial(\mathsf{NMSE}[\widetilde{d}])}{\partial c} \;&=\; \frac{(4\beta\;c \;+\; 2(\alpha+\beta))^2\left(2(e^{8\beta}-1)\;c \;+\; 2(e^{4\beta}-1)\right)}{k(4\beta\;c \;+\; 2(\alpha+\beta))^4} \;- \\
&\qquad \frac{8\beta(4\beta\;c \;+\; 2(\alpha+\beta))\left((e^{8\beta}-1)\;c^2 \;+\; 2(e^{4\beta}-1)\;c \;+\; \frac{1}{2}(e^{4\alpha}(e^{4\beta}+1)-2)\right)}{k(4\beta\;c \;+\; 2(\alpha+\beta))^4} \\
&=\; \frac{(2\beta\;c \;+\; \alpha+\beta)\left((e^{8\beta}-1)\;c \;+\; (e^{4\beta}-1)\right)}{2k(2\beta\;c \;+\; \alpha+\beta)^3} \;- \\
&\qquad \frac{2\beta\left((e^{8\beta}-1)\;c^2 \;+\; 2(e^{4\beta}-1)\;c \;+\; \frac{1}{2}(e^{4\alpha}(e^{4\beta}+1)-2)\right)}{2k(2\beta\;c \;+\; \alpha+\beta)^3} \\
&=\; \frac{2\beta(e^{8\beta}-1)\;c^2 \;+\; \left((\alpha+\beta)(e^{8\beta}-1)+2\beta(e^{4\beta}-1)\right)\;c \;+\; (\alpha+\beta)(e^{4\beta}-1)}{2k(2\beta\;c \;+\; \alpha+\beta)^3} \;- \\
&\qquad \frac{2\beta(e^{8\beta}-1)\;c^2 \;+\; 4\beta(e^{4\beta}-1)\;c \;+\; \beta(e^{4\alpha}(e^{4\beta}+1)-2)}{2k(2\beta\;c \;+\; \alpha+\beta)^3} \\
&=\; \frac{\left((\alpha+\beta)(e^{8\beta}-1)-2\beta(e^{4\beta}-1)\right)\;c \;-\; \left(\beta(e^{4\alpha}(e^{4\beta}+1)-2)-(\alpha+\beta)(e^{4\beta}-1)\right)}{2k(2\beta\;c \;+\; \alpha+\beta)^3} \\
&=\; \frac{\left((e^{8\beta}-2e^{4\beta}+1)\;\beta \;+\; (e^{8\beta}-1)\;\alpha\right)\;c \;-\; \left((e^{4\alpha+4\beta}+e^{4\alpha}-e^{4\beta}-1)\;\beta \;-\; (e^{4\beta}-1)\;\alpha\right)}{2k(2\beta\;c \;+\; \alpha+\beta)^3} \\
&=\; \frac{\left((e^{4\beta}-1)^2\beta \;+\; (e^{8\beta}-1)\alpha\right)\;c \;-\; \left((e^{4\beta}+1)(e^{4\alpha}-1)\beta \;-\; (e^{4\beta}-1)\alpha\right)}{2k(2\beta\;c \;+\; \alpha+\beta)^3}\;.
\end{aligned} \tag{21}$$

Clearly, $c_{\mathtt{OPT}}$ of §20 is the only point where this derivative is nullified. Also note that $\mathsf{NMSE}[\widetilde{d}]$ is a continuous and differentiable function of c in the range of c which leads to positive substitution rates ($4\beta c + 2(\alpha+\beta) > 0$). Therefore, in order to establish that $c_{\mathtt{OPT}}$ is a global minimum, all we have to do is show it is a local minimum. The derivative in §21 is negative when $c = 0$, due to the positivity of the numerator of §20 which was established earlier. Furthermore, when c goes to ∞ the derivative also goes to ∞, due to the positivity of the denominator of §20. Therefore, $c_{\mathtt{OPT}}$ provides a local, and hence a global, minimum for $\mathsf{NMSE}[\widetilde{d}]$. □