DIMACS Series in Discrete Mathematics
and Theoretical Computer Science
Volume **37**, 1997

Dissimilarity Maps and Substitution Models: Some New Results

Vincent Moulton, Mike Steel, and Chris Tuffley

ABSTRACT. In part one we describe some new results on reconstructing trees from distance measures, based on results in Peter Buneman's pioneering (1971) paper. In part two we analyse a covarion-style model, of the type suggested by Walter Fitch (and colleagues) in 1971 that offers a plausible explanation for why sequence sites "appear" to evolve at different rates.

INTRODUCTION

This paper summarises some new results. Further details (and most of the proofs) will appear elsewhere (see [**28, 32**]). The paper is in two parts:

Part One: Trees with positively-weighted edges induce a natural metric on any subset of vertices, however not every metric is representable in this way. A problem arising in areas of classification, particularly in evolutionary biology, is how to approximate an arbitrary distance function by such a tree metric, and thereby estimate the underlying tree that generated the data. Such transformations, from distances to tree metrics (and thereby to edge-weighted trees) should have some basic properties such as continuity, so that a small change in the input data does not result in a drastically different tree, but this is lacking in several popular methods, for example (as pointed out by Buneman) in methods that attempt to find a closest fit tree metric, and (as we show) in the popular *neighbor joining* method. However known continuous transformations, such as Buneman's original construction, often produce uninteresting (unresolved) trees, which led Buneman to suggest that perhaps this was "the price paid for continuity". One way to extract more information (and continuously!) from distances is Bandelt and Dress' elegant *split decomposition* theory. Based on a modification of Buneman's construction,

1991 *Mathematics Subject Classification.* Primary: 92B10; Secondary: 05C05, 60J27, 92D15, 92D20.

Key words and phrases. Trees, tree metrics, isolation index, nucleotide substitution, covarion model, Markov processes, moment generating function.

The first author was supported in part by the NZ Lotteries Commission.

Correspondence should be directed to the second author.

Work on Part Two was funded by the New Zealand Marsden Fund, contract UOC 516. Thanks also to Dr David Penny, Dr Walter Fitch and Dr Boris Mirkin for their helpful comments.

this continuous map produces a much larger number of splits than Buneman's map, though these splits generally do not form a tree. Here we suggest an alternative modification to the Buneman construction that always leads to trees, and which are, in general, more resolved than those obtained via Buneman's construction. Yet we can achieve this goal without sacrificing continuity. This suggests the possibility of finding other such maps.

Part Two: A "covarion" model for nucleotide substitution which allows sites to turn "on" and "off" with time was proposed 25 years ago by Fitch and Markowitz. It has been argued that evidence supports such models over later, alternative models which postulate a static distribution of rates across sites. However, in contrast to these latter well-studied models, little is known about the analytic properties of the former model. Here we analyse a covarion-style model and show (i) how to obtain the evolutionary distance between two species from the expected proportion of sites where two species differ (ii) that the covarion model cannot be distinguished from a suitably chosen rates-across-sites model on pairs of taxa if only the trace of the joint probability matrix is considered (i.e. the probability that the two taxa are in the same state) and give conditions under which the two models may be distinguished if the full matrix is examined, (iii) that the two models can, in principle, be distinguished when there are at least four monophyletic groups of species. In particular, with a view to a possible test of the covarion hypothesis (against a rates-across-sites model) we construct a distance measure which is a tree metric under certain versions of the covarion model (satisfying a certain separability condition) but which, in general, will not be a tree metric under a rates-across-sites model. Such a measure may also be useful for reconstructing the tree on the monophyletic groups when the covarion model applies.

PART ONE: DISSIMILARITY MAPS

1. Tree metrics, edge-weighted S-trees and indices

Let $S := \{1, \ldots, n\}$, and define

$$\mathcal{D}(S) := \{d : S \times S \to \mathbb{R}_{\geq 0} \ : \ d_{xy} = d_{yx},\ d_{xx} = 0 \text{ for all } x, y \in S\}$$

to be the set of *distance functions* on S. A distance function which satisfies the triangle inequality ($d_{xy} \leq d_{xz} + d_{zy}$ for all $x, y, z \in S$) is said to be a *pseudo-metric*. Endow $\mathcal{D}(S)$ with the l^p norm, that is, set

$$\|d - d'\|_p = \begin{cases} (\Sigma_{i,j} |d_{ij} - d'_{ij}|^p)^{\frac{1}{p}} & p = 1, 2, \ldots \\ \max_{i,j} |d_{ij} - d'_{ij}| & p = \infty. \end{cases}$$

A distance function d on a finite set S is said to be a *tree metric* if there exists a tree $T = (V, E)$, a map $L : S \to V$, called a *labelling*, and a map $w : E \to \mathbb{R}_{>0}$, called an *edge weighting*, such that for all $x, y \in S$, d_{xy} is the sum of $w(e)$ over all edges e in the unique path in T connecting vertices $L(x)$ and $L(y)$.

We may assume that the tree T has no vertices in $V - L(S)$ of degree less than or equal to two, since, as is easily seen, any tree metric on S can be realized by such a tree with a suitable edge weighting. We call such a tree T (together with its associated labelling L) an *S-tree*. S-trees and tree metrics arise in many contexts, particularly in phylogenetic analysis in evolutionary biology (see, for example, [**2, 20**]).

Two fundamental results concerning the characterisation of tree metrics and their uniqueness of representation date back to the 1960s and work by the Russians Zaretsky [**36**] and Smolensky [**30**], and we recall these results and their recent extensions here. One classical result is that a tree metric can arise from only one triple (T, L, w) where T is an S-tree, and w is an edge weighting of T [**1, 6, 30, 36**]. Thus tree metrics are in a natural bijective correspondence with positively edge-weighted S-trees, and, furthermore, there exist fast algorithms for recovering the triple (T, L, w) from d (see, for example, [**1, 2, 19**]). We refer to T (with its associated labelling L) as the S-tree *associated with* d.

Given $d \in \mathcal{D}(S)$ and $\delta \geq 0$, d is said to be δ-*hyperbolic* if

$$d_{ij} + d_{kl} \leq \max\{d_{ik} + d_{jl}, d_{il} + d_{jk}\} + \delta$$

for all $i, j, k, l \in S$. This is a relaxation of the *four-point condition*, in which $\delta = 0$ (for a discussion of this point see [**9**]). A second classical result states that a pseudo-metric d is a tree metric if and only if d is 0-hyperbolic [**6, 30, 36**]. More generally, a result originally given in [**17**], and which is also described in [**5**], states that if a pseudo-metric d is δ-hyperbolic, then there exists a $d' \in \mathcal{T}(S)$ with

$$\|d - d'\|_\infty \leq (1 + \log_2 n)\delta,$$

where $n = |S|$. Thus, if δ is small, then d is close to a tree metric up to a term that grows slowly in n. For a different generalisation of the 0-hyperbolic result to "distance" functions taking values in a suitably structured Abelian monoid $\mathcal{A}$ (and the realisation of such functions by S-trees with edges weights in $\mathcal{A} \setminus \{0\}$) see [**4**].

Let $\mathcal{T}(S)$ be the subspace of $\mathcal{D}(S)$ consisting of tree metrics, and $\mathcal{S}(S)$ be the set of *splits* of S, that is, bipartitions of S. Note that each edge of an S-tree induces a split of S defined by the two non-empty subsets of S that label the two subtrees of T when e is deleted. We say that this split is a *split of* T and is *associated* to edge e. Notice also that any tree metric $d \in \mathcal{T}(S)$ can be conveniently written in the form

$$d = \Sigma_{\sigma \in \mathcal{S}(S)} \lambda_\sigma \cdot \delta_\sigma, \tag{1.1}$$

where

$$\lambda_\sigma = \lambda_\sigma(d) := \begin{cases} w(e) & \text{if } \sigma \text{ is associated to } e \\ 0 & \text{if } \sigma \text{ is not associated to any edge of } T, \end{cases}$$

and where

$$\delta_\sigma(i, j) := \begin{cases} 1 & \text{if } \sigma \text{ separates } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

($\sigma = \{A, B\}$ *separates* i and j if $i \neq j$, and $|\{i, j\} \cap A| = 1$).

Given $d \in \mathcal{D}(S)$ we can define some maps from $\mathcal{S}(S)$ into $\mathbb{R}$, which we call *indices*. We adopt the convenient shorthand xy for d_{xy}. Suppose that $\sigma = \{A, B\}$ is a split of S. Let

$$\mu_\sigma = \mu_\sigma(d) := \frac{1}{2} \cdot \min_{a,a' \in A, b,b' \in B} \{\min\{ab + a'b', ab' + a'b\} - (aa' + bb')\},$$

$$\mu_\sigma^+ = \mu_\sigma^+(d) := \max\{0, \mu_\sigma\},$$

$$\alpha_\sigma = \alpha_\sigma(d) := \frac{1}{2} \cdot \min_{a,a' \in A, b,b' \in B} \{\max\{ab + a'b', ab' + a'b\} - (aa' + bb')\},$$

$$\alpha_\sigma^+ = \alpha_\sigma^+(d) := \max\{0, \alpha_\sigma\}.$$

The map μ is the *Buneman index* [**6**], while α^+ is the *isolation index* [**3**]. Clearly, for any $\sigma \in \mathcal{S}(S)$, we have $\mu_\sigma \leq \alpha_\sigma$ and $\mu_\sigma^+ \leq \alpha_\sigma^+$. The proof of the following lemma can be found in [**3**] and [**6**].

LEMMA 1.1. *If d is an element of $\mathcal{T}(S)$ with $d = \Sigma_\sigma \lambda_\sigma \cdot \delta_\sigma$, then $\lambda_\sigma = \mu_\sigma^+ = \alpha_\sigma^+$ for all $\sigma \in \mathcal{S}(S)$.*

Let $\lambda(d)$ be the vector $[\lambda_\sigma(d)]$ which lies in $\mathbb{R}^{|\mathcal{S}(S)|}$,

$$\mathcal{W}(S) := \{\lambda(d) \,:\, d \in \mathcal{T}(S)\},$$

and endow $\mathcal{W}(S) \subseteq \mathbb{R}^{|\mathcal{S}(S)|}$ with the l^p norm. The l^1 norm on the the space $\mathcal{W}(S)$ was proposed in [**29**] as a natural metric for comparing edge-weighted trees. The following theorem shows that $\mathcal{W}(S)$ and $\mathcal{T}(S)$ are homeomorphic. In particular the question of whether or not a map of $\mathcal{D}(S)$ into $\mathcal{T}(S)$ is good does not depend on whether we view the output as a distance function or as an edge-weighted S-tree. The second inequality in Theorem 1.2 is also established, using a slightly different approach, in [**10**, Lemmas 6,7].

THEOREM 1.2. *For $d, d' \in \mathcal{T}(S)$, we have*

$$\begin{aligned} \|d - d'\|_\infty &\leq \|\lambda(d) - \lambda(d')\|_1, \\ \|\lambda(d) - \lambda(d')\|_\infty &\leq 2 \cdot \|d - d'\|_\infty, \end{aligned}$$

and both of these inequalities can be equalities for any S.

PROOF. Writing d, d' in the form of equation (1.1) we have

$$\begin{aligned} \|d - d'\|_\infty &= \max_{i,j} |d_{ij} - d'_{ij}| \\ &= \max_{i,j} |\Sigma_{\{\sigma \in \mathcal{S}(S)\}} (\lambda_\sigma - \lambda'_\sigma) \cdot \delta_\sigma(i,j)| \\ &\leq \max_{i,j} \Sigma_{\{\sigma \in \mathcal{S}(S)\}} |\lambda_\sigma - \lambda'_\sigma| \cdot \delta_\sigma(i,j) \\ &\leq \Sigma_{\{\sigma \in \mathcal{S}(S)\}} |\lambda_\sigma - \lambda'_\sigma| \cdot \max_{i,j} \{\delta_\sigma(i,j)\} \\ &= \|\lambda(d) - \lambda(d')\|_1 \ . \end{aligned}$$

To obtain the second inequality, we show that for any $\sigma \in \mathcal{S}(S)$,

$$|\lambda_\sigma - \lambda'_\sigma| \leq 2 \cdot \delta,$$

where $\delta = \|d - d'\|_\infty$.

Now

$$\begin{aligned} |\lambda_\sigma - \lambda'_\sigma| &= |\mu_\sigma^+(d) - \mu_\sigma^+(d')| \\ &\leq |\mu_\sigma(d) - \mu_\sigma(d')| \\ &\leq 2 \cdot \delta \end{aligned}$$

since, by definition of μ_σ and the triangle inequality

$$\mu_\sigma(d) \leq \mu_\sigma(d') + 2 \cdot \delta,$$

and

$$\mu_\sigma(d') \leq \mu_\sigma(d) + 2 \cdot \delta.$$

This establishes the the two inequalities in the Theorem. To see that they can both be equalities we give the following two examples.

For the first inequality let d be the tree metric induced by the S-tree given by labelling bijectively the degree one vertices of a star tree (a tree having just one

vertex of degree larger than 1) by the elements of S, and assigning weight α to each edge. Let d' be defined in the same way, except that we assign one of the edges weight β instead of α. Then we immediately see that

$$\|d - d'\|_\infty = \|\lambda(d) - \lambda(d')\|_1 = |\alpha - \beta|.$$

For the second inequality, take a tree with four leaves, labelled bijectively by S, and with five edges. Let d be the metric on S induced by assigning weight 2 to all five edges; let d' be the metric on S induced by assigning weight 1 to the central edge and $9/4$ to the other four edges. Then,

$$\|\lambda(d) - \lambda(d')\|_\infty = 1 = 2 \cdot \|d - d'\|_\infty.$$

This completes the proof.

□

2. Retractions

2.1. Preliminaries. A map $\varphi : \mathcal{D}(S) \to \mathcal{D}(S)$ is a *retraction* onto $\mathcal{T}(S)$ if

(i) φ is *continuous*,
(ii) $\varphi(d) \in \mathcal{T}(S)$ for all $d \in \mathcal{D}(S)$, and
(iii) $\varphi(d) = d$ for all $d \in \mathcal{T}(S)$.

Furthermore, if such a retraction φ is *homogeneous*, that is, if

$$\varphi(\lambda d) = \lambda \varphi(d)$$

for all $\lambda > 0$ and $d \in \mathcal{D}(S)$, and if φ is *equivariant*, that is, for all $\tau \in \Sigma_S$ (the permutation group on S)

$$\varphi(d^\tau) = \varphi(d)^\tau,$$

where

$$(d^\tau)_{ij} = d_{\tau(i)\tau(j)},$$

then we say that φ is *good*. These last two properties are desirable in applications in requiring the method to be independent of the units in which d is measured and the names given to the objects in S, respectively [**22, 34**].

Define a partial order on the set of retractions as follows. Given two retractions φ_1, φ_2 of $\mathcal{D}(S)$ onto $\mathcal{T}(S)$, and a metric $d \in \mathcal{D}(S)$, let

$$\varphi_i(d) = \Sigma_{\sigma \in \mathcal{S}(S)} \lambda^i_\sigma(d) \cdot \delta_\sigma, \quad i = 1, 2.$$

We say that φ_2 *refines* φ_1, written $\varphi_1 \preceq \varphi_2$, if and only if for all $d \in \mathcal{D}(S)$ we have

$$\lambda^1_\sigma(d) \leq \lambda^2_\sigma(d),$$

for all $\sigma \in \mathcal{S}(S)$. As can be easily verified, $\preceq$ is a partial order. Note that if $\varphi_1 \preceq \varphi_2$, and if T_1, T_2 are the S-trees associated with $\varphi_1(d)$, $\varphi_2(d)$, respectively, then T_2 is a *refinement* of T_1, in the sense that T_1 can be obtained from T_2 by collapsing a subset of edges.

As has been pointed out (see [**6, 34**]) many early maps for constructing tree metrics fail to be continuous. It can also be shown that the currently popular (in biology) "neighbour-joining" method (see [**20**, p. 488]) is also discontinuous, even when $n = 4$ [**28**].

2.2. The Buneman retraction. Two splits $\sigma = \{A, B\}, \sigma' = \{A', B'\}$ in $\mathcal{S}(S)$ are said to be *compatible* if at least one of the intersections $A \cap A'$, $A \cap B'$, $B \cap A'$, $A' \cap B'$ is empty. If two splits σ, σ' are not compatible then we say that they are *incompatible*, and denote this by writing $\sigma \perp \sigma'$. Clearly any S-tree gives a set of pairwise compatible splits: just take the set of splits induced by the set of edges of the tree. Moreover in [**6**] it is shown that a set of pairwise compatible splits gives rise to a unique tree.

THEOREM 2.1. [**6**] *The set $\{\sigma : \mu_\sigma > 0\}$ is a pairwise compatible collection of splits, and thus gives rise to a unique S-tree.*

The index μ is the basis for the following good map, which is given in [**6**]. We define the *Buneman retraction* $\varphi_B : \mathcal{D}(S) \to \mathcal{T}(S)$ by setting

$$\begin{aligned} \varphi_B(d) &:= \Sigma_{\{\sigma : \mu_\sigma > 0\}} \mu_\sigma \cdot \delta_\sigma \\ &= \Sigma_{\sigma \in \mathcal{S}(S)} \mu_\sigma^+ \cdot \delta_\sigma. \end{aligned}$$

By the previous corollary and the properties of the Buneman index μ, φ_B is a good map. In addition, from [**6**], $\varphi_B(d) \leq d$, in the sense that

$$\varphi_B(d)_{ij} \leq d_{ij}, \text{ for all } i, j \in S.$$

2.3. The Refined Buneman Retraction. In this section we define a new index map $\overline{\mu}_\sigma$ which refines the Buneman index, in the sense that $\overline{\mu}_\sigma \geq \mu_\sigma$ for all $\sigma \in \mathcal{S}(S)$, with strict inequality holding for certain cases. We assume throughout this section that $n \geq 4$.

Writing $q := ab|cd$ to denote the bipartition $\{\{a, b\}, \{c, d\}\}$ of the subset $\{a, b, c, d\}$ of S, let

$$\beta_q := \frac{1}{2}(\min\{ac + bd, ad + bc\} - (ab + cd)).$$

Thus, given a split $\sigma = \{A, B\}$ of S, the Buneman index of σ is given by

$$\mu_\sigma = \min_{a,a' \in A, b,b' \in B} \{\beta_{aa'|bb'}\}.$$

Let Q be the set of $q = aa'|bb'$ consisting of all unordered choices of $a, a' \in A$, and $b, b' \in B$, insisting, furthermore, that if $|A| \geq 2$, then $a \neq a'$ and if $|B| \geq 2$, then $b \neq b'$. Now let $q_1, \ldots, q_{|Q|}$ be an ordering of the elements in Q such that $\beta_{q_i} \leq \beta_{q_j}$ for all $1 \leq i \leq j \leq |Q|$, and define the *refined Buneman index* by

$$\overline{\mu}_\sigma := \frac{1}{n-3} \cdot \Sigma_{i=1}^{n-3} \beta_{q_i}.$$

Note that, by definition, $\overline{\mu}_\sigma \geq \mu_\sigma$ for all $\sigma \in \mathcal{S}(S)$.

LEMMA 2.2 ([**28**]). *If σ and σ' are incompatible splits then*

$$\mu_\sigma + \mu_{\sigma'} \leq 0, \quad \overline{\mu}_\sigma + \overline{\mu}_{\sigma'} \leq 0.$$

This generalises Theorem 2.1, and is a useful tool for establishing the next theorem.

THEOREM 2.3. [**28**] *The map*

$$\psi : d \mapsto \Sigma_{\{\sigma : \overline{\mu}_\sigma > 0\}} \overline{\mu}_\sigma \cdot \delta_\sigma,$$

is a good map, and $\varphi_B \preceq \psi$.

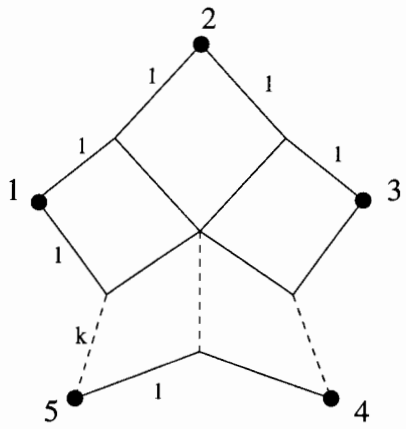

FIGURE 1. *All edges are weighted 1 except dotted edges, which are weighted k.*

We now give a simple example to illustrate that, in certain cases, the refined Buneman retraction gives us a tree which strictly refines the tree given by the Buneman retraction i.e. $\varphi_B \prec \psi$.

Consider the metric d_k on the set $\{1, \ldots, 5\}$ given by the shortest (weighted) path between vertices of the edge-weighted graph in Figure 1, where all edges have weight one except those which are dotted, which have weight k, for some $k \geq 0$.

The Buneman tree for d_k depends upon the value of k. For the case $0 \leq k \leq 2$ the Buneman tree is simply a vertex. If $k \geq 2$, then the Buneman tree consists of one edge of length $k-2$, with its endpoints labelled by $\{1,2,3\}$ and $\{4,5\}$. Thus, in either case, the Buneman tree is highly unresolved (in the sense of [**2**]).

However, in contrast to this, the refined Buneman tree (i.e. that given by using the refined Buneman index), the topology of which also depends upon k, and which is shown in Figure 2, is fully resolved for $k > 0$. Note that in the case where $k = 1$ we get, as might be expected, a star tree.

Note that, in biological applications at least, a desirable feature of a good map is that it be efficiently computable. We will address the computability of the refined Buneman retraction and applications of the refinement to biological data elsewhere [**21**].

2.4. Identifying S-trees using the Buneman retraction and its refinement. We show how the Buneman retraction (or its refinement) essentially identifies the underlying S-tree of a tree metric d' when applied to a distance function d that is close enough to d'. This is summarized in the following theorem.

THEOREM 2.4. *Let $\varphi = \varphi_B$ or ψ (the Buneman retraction or its refinement). Suppose that $d' \in \mathcal{T}(S)$ has associated S-tree T and edge weighting w. Let*

$$x := \min\{w(e) : e \in T\},$$

and suppose that for $d \in \mathcal{D}(S)$ one has

$$\|d - d'\|_\infty < \frac{x}{2}.$$

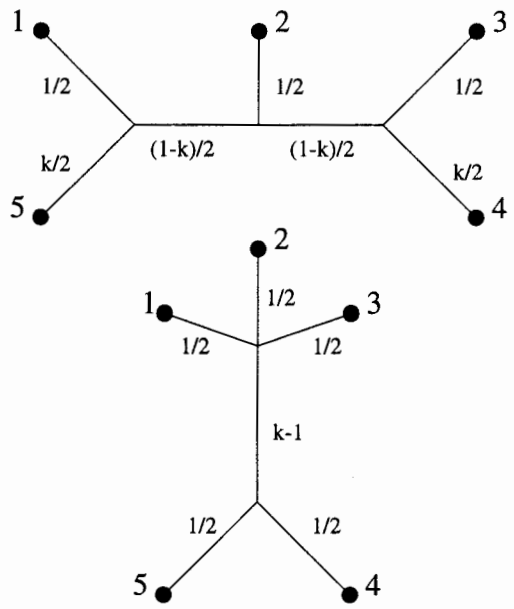

FIGURE 2. *The refined Buneman tree: the top tree is for the case* $0 \leq k \leq 1$ *and the bottom for the case* $1 \leq k$.

Then the S-tree, t, associated to $\varphi(d)$ *refines* T *and the weight of any edge in* t *that does not correspond to an edge of* T *is less than* x. *In particular, if* T *is a fully resolved tree (i.e. every vertex has degree* 1 *or* 3*), then* $t = T$.

PROOF. Suppose $\varphi = \varphi_B$ and let σ be a split of T corresponding to edge e. Then $\mu_\sigma(d') = w(e) \geq x$. Now

$$|\mu_\sigma(d) - \mu_\sigma(d')| \leq 2\delta, \tag{2.1}$$

where $\delta = ||d - d'||_\infty$, as in the proof of Theorem 1.2. Hence, since $\delta < x/2$,

$$\begin{aligned} \mu_\sigma(d) &\geq \mu_\sigma(d') - 2\delta \\ &> x - x = 0, \end{aligned}$$

and so σ is a split of t. Thus t refines T, and in particular, if T is fully resolved then $t = T$.

If T is not fully resolved and σ is a split of t but not T, then by (2.1)

$$\begin{aligned} \mu_\sigma(d) &\leq \mu_\sigma(d') + 2\delta \\ &< 0 + x, \end{aligned}$$

and we deduce that the edge e of t corresponding to σ has weight less than x.

The proof for $\varphi = \psi$ is exactly the same, except that the justification of the analogue of (2.1) is slightly more involved.

□

PART TWO: SUBSTITUTION MODELS

3. The models

In order to accurately reconstruct evolutionary trees and time scales from aligned nucleotide sequences it is helpful to model the mechanism by which the sequences came to differ. Such models can be used to devise new techniques for tree reconstruction and analysis, and also to determine cases where existing methods are likely to lead to erroneous results (see for example [**11**]).

The simplest and earliest models assume that each site evolves i.i.d. at the same rate, and according to simple Markov-style assumptions. However, this single-rate assumption appears to be unrealistic, and accordingly models incorporating some variation of rates across sites have been proposed and studied to take into account different functional constraints at different sites (see for example [**7, 31, 35**]). An alternative approach to accounting for differing selective constraints is Fitch and Markowitz's "concomitantly variable codons" or "covarion" hypothesis [**15**]. This is that at any given time, some sites are invariable due to functional or structural constraints, but that as mutations are fixed elsewhere in the sequence these constraints may change, so that sites that were previously invariable may become variable and vice versa. The pool of variable sites is therefore changing with time (see Figure 3). Since its proposal 25 years ago, it has been argued that evidence supports the covarion hypothesis, both on biochemical grounds, and by providing a better description of certain data [**13, 14, 27**]. However, in contrast to the rates-across-sites models, little is known about the analytic properties of covarion-style models.

Here we present and analyse a simple covarion-style model. Although the motivation for this model clearly says that the i.i.d. assumption is not valid, without it the mathematics becomes much more difficult. We therefore keep this assumption and model the behaviour only of a covarion-style process, with a two-state Markov process that acts as a "switch", turning sites "on" (variable) and "off" (invariable). We do not impose any restrictions on the Markov process that operates at the variable sites other than that it is stationary and reversible. Using techniques from the theory of Markov processes such a model may be analysed and compared with rates-across-sites models in terms of the expected frequencies of site patterns the models should generate.

3.1. A covarion-style model. We model a covarion-style process with two parts: a "switch" process, and an "observable" process, which operates while the switch is "on". Only the state of the observable process, and not that of the switch process, is able to be measured.

The switch is governed by a two state continuous time Markov process with state space $\mathcal{O} = \{\texttt{on}, \texttt{off}\}$ and rate matrix

$$S = \begin{pmatrix} -s_1 & s_1 \\ s_2 & -s_2 \end{pmatrix}$$

where $s_i > 0$ for each i. It is assumed to have the stationary initial distribution $\sigma = (\sigma_1, \sigma_2)$ where

$$\sigma_1 = \frac{s_2}{s_1 + s_2}, \quad \sigma_2 = \frac{s_1}{s_1 + s_2},$$

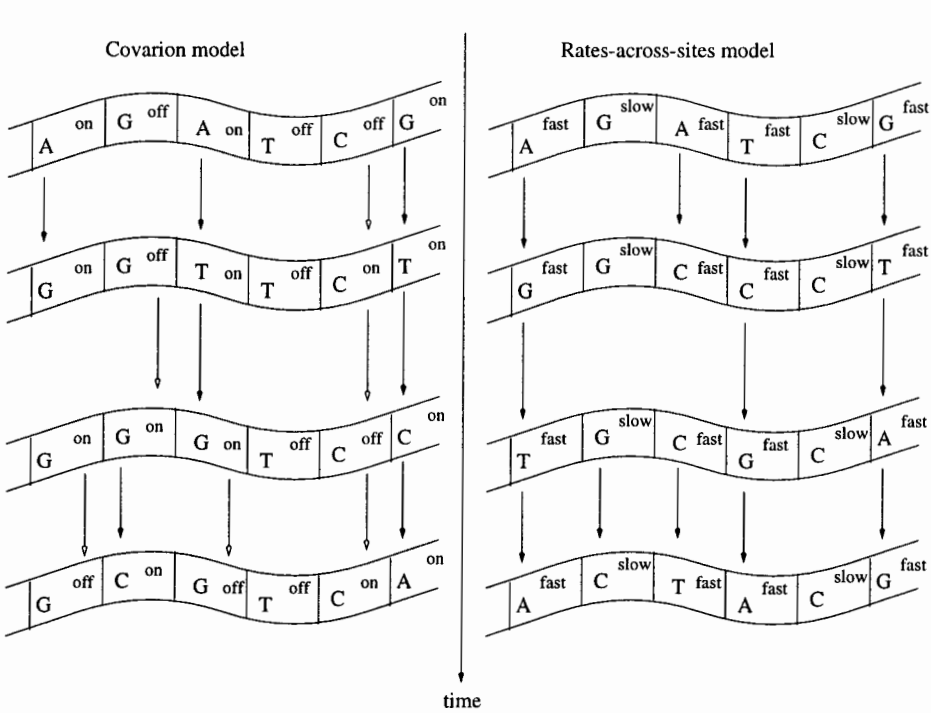

FIGURE 3. Contrasting a covarion style process and rates-across-sites. Under a covarion style process, each site is either "on" or "off". Sites that are off are unable to change character state, but may later turn on (due to character state changes elsewhere in the sequence) and be able to change. Under rates-across-sites, sites evolve at different rates (shown here as "fast" and "slow"), with faster sites changing more frequently than slower ones. The rate at a given site is assumed constant across the entire tree.

so that it is stationary and time-reversible. For a background in Markov processes, the reader is referred to [**16**] and [**24**].

While the switch is in state `off`, the observable process is unable to change state; however, when the switch is in state `on`, the observable process is governed by a second stationary and time reversible Markov process with state space $\mathcal{A} = \{1, \ldots, r\}$, rate matrix R satisfying $R_{ij} > 0$ if $i \neq j$, and initial distribution π. Stationarity and time-reversibility are equivalent to the conditions

$$\pi R = 0 \quad \text{and} \quad \pi_i R_{ij} = \pi_j R_{ji} \quad \text{for all } i, j.$$

In general, for positive integer n we denote the set $\{1, \ldots, n\}$ by $[n]$, and we write $C = (R, S)$ for the covarion model C with observable process rate matrix R and switch process rate matrix S.

This model may be alternatively formulated in terms of a single time-reversible Markov process with state space $\mathcal{A} \times \mathcal{O}$ (which we identify with $[2r]$ according to $(i, \texttt{on}) \mapsto i$, $(i, \texttt{off}) \mapsto i + r$), initial distribution $\pi' = (\sigma_1 \pi, \sigma_2 \pi)$ and $2r \times 2r$ rate

matrix

$$R' = \begin{pmatrix} R - s_1 I_r & s_1 I_r \\ s_2 I_r & -s_2 I_r \end{pmatrix},$$

where I_r denotes the $r \times r$ identity matrix. In this formulation we assume that we are unable to distinguish between the states $(i, \mathtt{on})$ and $(i, \mathtt{off})$.

It is easily checked that R' is stationary and time-reversible whenever R and S are. Further, both formulations lead to the same joint probability matrix for $\mathcal{A} \times \mathcal{A}$. Using this formulation we may also show that the resulting random process on $\mathcal{A}$ is not in general Markov, so the covarion model may not be analysed by simply treating it as a Markov process on $\mathcal{A}$.

3.2. Rates-across-sites. A *rates-across-sites* model $D = (Q, \mathcal{D})$ consists of a stationary and time-reversible continuous time Markov process with rate matrix Q and initial distribution θ, and a distribution $\mathcal{D}$ of rates ν, which may be either discrete or continuous. We denote the cumulative distribution function of $\mathcal{D}$ by $F_{\mathcal{D}}$.

Each site evolves according to rate matrix νQ where ν is chosen i.i.d. according to $\mathcal{D}$. The rate at a given site is assumed constant across the whole tree. This kind of model has been well studied, see for example [**7, 31, 35**].

4. The two taxa tree

Here we calculate the joint probability matrix for the two taxa tree (that is, the matrix whose ij entry is the probability that taxa 1 is in state i and taxa 2 is in state j), and give conditions under which a suitably chosen rates-across-sites model will agree with a covarion model on all two taxa trees. We also consider the limiting cases of the covarion model as the rate of the switch tends either to zero or to infinity.

4.1. Under the covarion model. Time reversibility implies we may assume the tree is rooted at either of the leaves. Let the process operate for time τ on the edge between the two taxa and write $J_C(\tau)$ for the joint probability matrix. We regard τ as the "length" of the edge. Put $\Pi = \mathrm{diag}(\pi)$ and let $J(t)$ be the joint probability matrix of the unswitched observable process (that is, the Markov process with rate matrix R and initial distribution π operating in the absence of the switch) for time t. If the occupation time of state on in time τ is the random variable $X(\tau)$, then, as far as the observable process is concerned, the edge has effective length $X(\tau)$. The joint probability matrix, given the value of $X(\tau)$, is then $J(X(\tau))$. It follows that

$$J_C(\tau) = \mathbb{E}[J(X(\tau))].$$

Reversibility allows us to obtain a spectral representation of $J(t)$ (see Keilson [**24**, pp. 32–35]). Since ΠR is symmetric so is $\Pi^{1/2} R \Pi^{-1/2}$ which therefore has real eigenvalues $\{\lambda_j\}$ and orthonormal eigenvectors $\{\mathrm{u}_j\}$ (related to the eigenvectors $\{v_j\}$ of R by $v_j = \Pi^{-1/2} u_j$ and $R v_j = \lambda_j v_j$). We then find that

$$J(t) = \sum_{j=1}^{r} e^{\lambda_j t} w_j w_j^{\mathrm{T}},$$

where $w_j = \Pi^{1/2} u_j$ and the superscripted T denotes transposition. Hence

$$\begin{aligned} J_C(\tau) &= \mathbb{E}\left[\sum_{j=1}^{r} e^{\lambda_j X(\tau)} w_j w_j^{\mathrm{T}}\right] \\ &= \sum_{j=1}^{r} \mathbb{E}[e^{\lambda_j X(\tau)}] w_j w_j^{\mathrm{T}}. \end{aligned}$$

From Darroch and Morris [8] we have

$$\mathbb{E}[e^{\lambda X(\tau)}] = \sigma^{\mathrm{T}} e^{\tau(S+\lambda D)} \mathbf{1}, \tag{4.1}$$

where $D = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and so, diagonalising $S + \lambda D$ we obtain

$$J_C(\tau) = \sum_{j=1}^{r} [c_j^+ e^{\mu_j^+ \tau} + c_j^- e^{\mu_j^- \tau}] w_j w_j^{\mathrm{T}}, \tag{4.2}$$

where μ_j^+ and μ_j^- are the positive and negative roots of

$$\mu^2 + (s_1 + s_2 - \lambda)\mu - s_2\lambda = 0$$

for $\lambda = \lambda_j$, and

$$c_j^+ = \frac{-(s_1 + s_2 + \mu_j^+)\mu_j^-}{(s_1 + s_2)(\mu_j^+ - \mu_j^-)} \quad \text{and} \quad c_j^- = \frac{(s_1 + s_2 + \mu_j^-)\mu_j^+}{(s_1 + s_2)(\mu_j^+ - \mu_j^-)} \quad .$$

These co-efficients may be shown to satisfy the following inequalities:

LEMMA 4.1.

1. $\lambda_j \leq 0$ *for* $j = 1, \ldots, r$, *with zero occurring as an eigenvalue exactly once.*
2. μ^+ *and* μ^- *are real increasing functions of* λ *satisfying* $\mu^- \leq -(s_1 + s_2) < -s_2 < \mu^+ \leq 0$ *on* $(-\infty, 0]$.
3. $c_j^+, c_j^- \geq 0$ *(with equality only for* $\lambda = 0$, *when* $c^- = 0$*) and* $c_j^+ + c_j^- = 1$.
4. *trace*$(w_j w_j^{\mathrm{T}}) > 0$ *and* $\sum_{j=1}^{r}$ *trace*$(w_j w_j^{\mathrm{T}}) = 1$.

An additional expression of interest is the trace of the joint probability matrix, which is the probability that the two species agree at a given site. Denoting the zero eigenvalue by λ_1, from equation (4.2) we obtain

$$\mathrm{trace}(J_C(\tau)) = \pi\pi^{\mathrm{T}} + \sum_{j=2}^{r} [c_j^+ e^{\mu_j^+ \tau} + c_j^- e^{\mu_j^- \tau}] \mathrm{trace}\left(w_j w_j^{\mathrm{T}}\right). \tag{4.3}$$

4.2. Under rates-across-sites. Put $\Theta = \mathrm{diag}(\theta)$ and let Q have eigenvalues $\{\alpha_j\}$. Arguing as for the covarion model, if $\Theta^{1/2} Q \Theta^{-1/2}$ has orthonormal eigenvectors $\{y_j\}$, then the joint probability matrix $J_D(\tau)$ of the rates-across-sites model D is given by

$$J_D(\tau) = \sum_{i=1}^{r} \mathbb{E}[e^{\alpha_j \nu \tau}] z_j z_j^{\mathrm{T}},$$

where $z_j = \Theta^{1/2} y_j$. We may write this as

$$J_D(\tau) = \sum_{i=1}^{r} M(\alpha_j \tau) z_j z_j^{\mathrm{T}} \tag{4.4}$$

where $M(x) = \mathbb{E}[e^{\nu x}]$ is the *moment generating function* of $\mathcal{D}$, given by the Lebesgue-Stieltjes integral

$$M(x) = \int_0^\infty e^{\nu x} dF_{\mathcal{D}}(\nu).$$

If $F_{\mathcal{D}}$ has a continuous derivative $f_{\mathcal{D}}$ (its probability density function) this is simply

$$M(x) = \int_0^\infty e^{\nu x} f_{\mathcal{D}}(\nu) d\nu,$$

while if $\mathcal{D}$ has only finitely many rates $\nu_1, \ldots, \nu_k$ and $\mathbb{P}[\nu = \nu_i] = p_i$ we have

$$M(x) = \sum_{i=1}^{k} p_i e^{\nu_i x}.$$

As in the covarion case we have

$$\text{trace}(J_{\mathcal{D}}(\tau)) = \theta\theta^{\mathrm{T}} + \sum_{i=2}^{r} M(\alpha_j \tau)\text{trace}(z_j z_j^{\mathrm{T}}). \tag{4.5}$$

4.3. Recovering the evolutionary distance under the two models. Equation (4.4) may be written

$$J_D(\tau) = \Theta M(\tau Q), \tag{4.6}$$

where M is the moment generating function of $\mathcal{D}$ applied to matrices. This expression has the advantage of enabling us to calculate the expected number of substitutions K between the two taxa without requiring knowledge of Q, via

$$K = -\text{trace}\left\{\Theta\left[M^{-1}\left(\Theta^{-1} J_D(\tau)\right)\right]\right\} \tag{4.7}$$

[**18, 33**]. Here M^{-1} is the inverse of the moment generating function, again applied to matrices. This expression gives a tree metric, and since row i of $J_D(\tau)$ sums to θ_i, requires knowledge only of $\mathcal{D}$ to reconstruct the tree from $J_D(\tau)$.

If both Q and $\mathcal{D}$ are known we may express K in terms of just the trace of $J_D(\tau)$ as

$$K = -\text{trace}(\Theta Q) f_D^{-1}(\text{trace}(J_D(\tau))), \tag{4.8}$$

where $f_D(\tau) = \text{trace}(J_D(\tau))$ is given by equation (4.5). Note that f_D^{-1} exists since f_D is monotone decreasing.

The property of (4.4) that allows it to be written in the form (4.6) (namely, M is applied to products of the form $\alpha_j \tau$) does not hold for (4.2), and it appears that a transformation analogous to (4.7) does not exist for the covarion model. However, if R and S are known (or estimated) then, as in (4.8), we may express K in terms of $\text{trace}(J_C(\tau))$ as

$$K = -\text{trace}(\Pi R)\sigma_1 f_C^{-1}(\text{trace}(J_C(\tau))),$$

where $f_C(\tau) = \text{trace}(J_C(\tau))$ is given by equation (4.3). Again f_C is monotone decreasing so f_C^{-1} exists.

Note that in applications, the joint probability matrix (J_C or J_D) is estimated from the observed joint frequency matrix $\hat{J}$. Since J_C and J_D are both symmetric, it is usual practice to take the symmetrised matrix $(\hat{J} + \hat{J}^{\mathrm{T}})/2$ as the estimate.

4.4. Distinguishing between the two models. A question of interest is whether it is possible to distinguish the covarion model from rates-across-sites from pair-wise comparisons of sequences. For a fixed $\tau = \tau_1$, if the rates are distributed according to the distribution of $X(\tau_1)/\tau_1$ then we have $J_C(\tau_1) = J_D(\tau_1)$ for $C = (R, S)$ and $D = (R, \mathcal{D})$, so we cannot tell the covarion model apart from rates-across-sites. However the distribution of $X(\tau)/\tau$ depends on τ which opens the possibility that the models may be distinguished if more than one pair is considered. A partial answer to this question is given by the following results:

THEOREM 4.2.

1. *For any covarion model C there is a rates-across-sites model D such that*

$$trace(J_D(\tau)) = trace(J_C(\tau))$$

for all $\tau \geq 0$.

2. *For a given covarion model $C = (R, S)$, there is a rates-across-sites model $D = (Q, \mathcal{D})$ such that*

$$J_C(\tau) = J_D(\tau)$$

for all $\tau \geq 0$ if and only if R has only one distinct non-zero eigenvalue, in which case $\mathcal{D}$ is a discrete two rate distribution and Q is a scalar multiple of R.

3. *For a given rates-across-sites model $D = (Q, \mathcal{D})$, there is a covarion model $C = (R, S)$ such that*

$$J_D(\tau) = J_C(\tau)$$

for all $\tau \geq 0$ if and only if Q has only one distinct non-zero eigenvalue and $\mathcal{D}$ is a discrete two rate distribution, with both rates greater than zero.

Furthermore stationary and reversible rate matrices with exactly one distinct non-zero eigenvalue and stationary distribution π may be completely characterised as scalar multiples of the matrix

$$R_\pi = \mathbf{1}\pi - I_r$$

where $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$.

It follows from Theorem 4.2 that the trace does not contain enough information to distinguish the covarion model from rates-across-sites models, in the sense that data generated by a covarion model could have been generated by a suitably chosen rates-across-sites model. Note however that parts (ii) and (iii) do not completely answer the question of when the two models may be distinguished on the basis of pairwise comparisons and without knowledge of τ. Firstly, the requirement that the times τ are the same is too strong: if $J_C(\tau) = J_D(f(\tau))$ for all $\tau \geq 0$ for an increasing function f such that $f(0) = 0$ and $f(\tau) \to \infty$ as $\tau \to \infty$ then we cannot distinguish between C and D. Secondly, since we can only ever compare finitely many sequences, it is important to know when we may have $J_C(\tau_i) = J_D(\tau_i')$ for times $\tau_1, \ldots, \tau_k, \tau_1', \ldots, \tau_k'$.

Section 5 gives an alternative approach to distinguishing between the covarion and rates-across-sites models.

We include a proof of Theorem 4.2 part (1). For proofs of the remaining parts of this theorem see [**32**].

PROOF. By (4.3) and Lemma 4.1, $\mathrm{trace}(J_C(\tau))$ has the form

$$\mathrm{trace}(J_C(\tau)) = \pi\pi^{\mathrm{T}} + \sum_{j=2}^{r}[c'^{+}_j e^{\mu^{+}_j \tau} + c'^{-}_j e^{\mu^{-}_j \tau}]$$

where $c'^{+}_j, c'^{-}_j > 0$ and $\sum_{j=2}^{r}[c'^{+}_j + c'^{-}_j] = 1 - \pi\pi^{\mathrm{T}}$. If R has k distinct non-zero eigenvalues we may collect terms in $e^{\mu^{\pm}\tau}$ for each eigenvalue, writing $\mathrm{trace}(J_C(\tau))$ in the form

$$\mathrm{trace}(J_C(\tau)) = a_0 + \sum_{i=1}^{2k} a_i e^{-\nu_i \tau},$$

where $a_i, \nu_i > 0$ for each i and $\sum_{i=0}^{2k} a_i = 1$.

Let $\mathcal{D}$ be the discrete distribution of rates such that

$$\mathbb{P}[\nu = \nu_i] = \frac{a_i}{1 - a_0} \qquad i = 1, \ldots, 2k.$$

Then $\mathcal{D}$ is well-defined, and if $D = (R_\pi, \mathcal{D})$ then by (4.5) and Lemma 4.1,

$$\begin{aligned}
\mathrm{trace}(J_D(\tau)) &= \pi\pi^{\mathrm{T}} + \sum_{j=2}^{r} M(-\tau)\mathrm{trace}(z_j z_j^{\mathrm{T}}) \\
&= \pi\pi^{\mathrm{T}} + M(-\tau)(1 - \pi\pi^{\mathrm{T}}) \\
&= a_0 + (1 - a_0)\sum_{i=1}^{2k} \frac{a_i}{1 - a_0} e^{-\nu_i \tau} \\
&= a_0 + \sum_{i=1}^{2k} a_i e^{-\nu_i \tau} \\
&= \mathrm{trace}(J_C(\tau)).
\end{aligned}$$

□

4.5. Limiting cases. We consider the limiting cases of the covarion model when the switch is very slow ($s_1, s_2 \to 0$) and very fast ($s_1, s_2 \to \infty$), keeping s_1/s_2 (the ratio of "off" sites to "on" sites) constant.

For very slow switches we expect few changes between the states on and off to occur, so that sites in state on will tend to remain in state on, and sites in state off will tend to remain in state off. In the limiting case $s_1, s_2 \to 0$ we expect σ_2 of the sites to be invariant and σ_1 of them to be variable. Calculating this limit we find

$$J_C(\tau) \to \sigma_2 J(0) + \sigma_1 J(\tau)$$

as expected.

For fast switches we expect sites to flip back and forth between on and off very rapidly, and each spend about the same amount of time in state on. Since the expected time in state on is $\sigma_1\tau$, in the limiting case $s_1, s_2 \to \infty$ with s_1/s_2 constant we expect

$$J_C(\tau) \to J(\sigma_1 \tau).$$

Calculating this limit we find this is indeed the case.

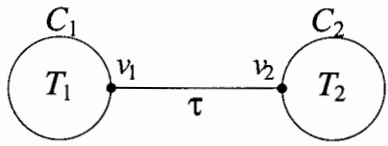

FIGURE 4. The tree joining two monophyletic groups of species C_1 and C_2. The circles marked T_1 and T_2 are rooted subtrees, with roots v_1 and v_2 respectively. The edge $\{v_1, v_2\}$ has length τ.

5. A tree-like measure on monophyletic groups under the covarion model

One approach to testing the covarion model against rates-across-sites models is to examine the sites that are varied and unvaried in two widely separated groups of closely related species. Under the rates-across-sites model, if a given site is in the same state for each member of a group of closely related taxa, then it is likely that the rate of evolution at that site is slow. Since the rate does not change across the tree, we might expect little change to occur in another group of closely related species that is widely separated from the first. On the other hand, under the covarion model if each species has the same state at a given site it seems likely that the site was off for much of the time. In a distant part of the tree the switch might be on so we no longer expect the unvaried sites in the two groups to match up. This observation was made by Fitch [**12**], and examined by Miyamoto and Fitch [**27**], who compared Cu, Zn superoxide dismutase (SOD) sequences from seven mammals and seven plants with simulated sequences generated under covarion and gamma distribution rates-across-sites models, finding that the covarion hypothesis explained the evolution of the protein better than rates-across-sites.

The following discussion is also motivated by Fitch's observation. For a certain class of events and parameters of a covarion model we obtain a tree-like distance measure between monophyletic groups of species that will not in general be tree-like under rates-across-sites models and so could lead to a test of the covarion model against rates-across-sites. The class of covarion models for which this is relevant includes those whose underlying observable process is based on the Kimura [**26**] three-substitution-type model (K3ST) or one of its submodels (the Kimura [**25**] two parameter (K2P) and Jukes-Cantor [**23**] (JC) models).

5.1. Separable events. We describe a class of events that give rise, under the covarion model, to a tree-like metric that is not in general tree-like under a rates-across-sites model.

Suppose E is an event involving an r-state character χ on a set C of species, for example the events

$$E^s = \text{“}\chi(i) \text{ is the same state for all } i \in C\text{”}$$

and

$$E^d = \text{“}\chi(i) \text{ is not the same state for all } i \in C\text{”}.$$

Given two monophyletic groups C_1 and C_2 of species, the tree joining them will be as shown in Figure 4. Let E_i be the event “E occurs for group C_i” for $i = 1, 2$. We

say that the event E is *separable* under the covarion model (R, S) if

$$\mathbb{P}[E_1 \wedge E_2 | 0_1 = o_1, 0_2 = o_2] = \mathbb{P}[E_1 | 0_1 = o_1]\mathbb{P}[E_2 | 0_2 = o_2] \tag{5.1}$$

for all $o_1, o_2 \in \{\mathtt{on}, \mathtt{off}\}$. Note that the separability of a given event may depend on R and S. An analogous condition that might be satisfied by a rates-across-sites model $(Q, \mathcal{D})$ is the following *independence condition:*

$$\mathbb{P}[E_1 \wedge E_2 | \nu] = \mathbb{P}[E_1 | \nu]\mathbb{P}[E_2 | \nu]. \tag{5.2}$$

Let

$$\begin{aligned} p_{12} &= \mathbb{P}[E_1 \wedge E_2], \\ p_i &= \mathbb{P}[E_i], \quad i = 1, 2 \end{aligned}$$

and further in the case of the covarion model let 0_i be the state on or off of the switch at the vertex v_i and

$$\begin{aligned} p_i^{\mathtt{on}} &= \mathbb{P}[E_i | 0_i = \mathtt{on}] \\ p_i^{\mathtt{off}} &= \mathbb{P}[E_i | 0_i = \mathtt{off}] \\ \delta_i &= p_i^{\mathtt{on}} - p_i^{\mathtt{off}} \end{aligned}$$

for $i = 1, 2$. Then under conditions (5.1) and (5.2) we have the following:

LEMMA 5.1.

1. *If E is separable under the covarion model (R, S) then*

$$p_{12} - p_1 p_2 = \sigma_1 \sigma_2 e^{-(s_1+s_2)\tau} \delta_1 \delta_2. \tag{5.3}$$

2. *If the independence condition holds for the rates-across-sites model $(Q, \mathcal{D})$ then $p_{12} - p_1 p_2$ does not depend on τ.*

THEOREM 5.2. *For a tree with several monophyletic groups $C_1, \ldots C_n$ at its tips the measure*

$$\rho_{ij} = -\ln |p_{ij} - p_i p_j|$$

is a tree metric realised by the under-lying tree under a covarion model for which E is separable, but in general is not a tree metric under a rates-across-sites model for which the independence condition holds.

PROOF OF LEMMA 5.1 AND THEOREM 5.2. In the covarion case,

$$\begin{aligned} p_{12} &= \sum_{o_1, o_2} \mathbb{P}[E_1 \wedge E_2 | 0_1 = o_1, 0_2 = o_2]\mathbb{P}[0_1 = o_1, 0_2 = o_2] \\ &= \sum_{o_1, o_2} \mathbb{P}[E_1 | 0_1 = o_1]\mathbb{P}[E_2 | 0_2 = o_2]\mathbb{P}[0_1 = o_1, 0_2 = o_2], \end{aligned}$$

since E is separable, and

$$p_1 p_2 = \sum_{o_1, o_2} \mathbb{P}[E_1 | 0_1 = o_1]\mathbb{P}[E_2 | 0_2 = o_2]\mathbb{P}[0_1 = o_1]\mathbb{P}[0_2 = o_2].$$

Thus

$$p_{12} - p_1 p_2 = \sum_{o_1, o_2} \mathbb{P}[E_1 | o_1]\mathbb{P}[E_2 | o_2] \left(\mathbb{P}[o_1, o_2] - \mathbb{P}[o_1]\mathbb{P}[o_2]\right). \tag{5.4}$$

Now the joint probability matrix for the switch operating for time τ is

$$(\mathbb{P}[0_1 = o_1, 0_2 = o_2]) = \sigma_1\sigma_2 \begin{pmatrix} \frac{s_2}{s_1} + e^{-(s_1+s_2)\tau} & 1 - e^{-(s_1+s_2)\tau} \\ 1 - e^{-(s_1+s_2)\tau} & \frac{s_1}{s_2} + e^{-(s_1+s_2)\tau} \end{pmatrix}$$

(see for example [**16**, p. 156]) so the matrix of $\mathbb{P}[0_1 = o_1, 0_2 = o_2] - \mathbb{P}[0_1 = o_1]\mathbb{P}[0_2 = o_2]$ is

$$\sigma_1\sigma_2 e^{-(s_1+s_2)\tau} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Hence, from (5.4),

$$p_{12} - p_1p_2 = \sigma_1\sigma_2 e^{-(s_1+s_2)\tau}(p_1^{\mathbf{on}} - p_1^{\mathbf{off}})(p_2^{\mathbf{on}} - p_2^{\mathbf{off}}) = \sigma_1\sigma_2 e^{-(s_1+s_2)\tau}\delta_1\delta_2$$

as claimed.

Under rates-across-sites with the independence condition holding,

$$\begin{aligned} \mathbb{P}[E_1 \wedge E_2] &= \int_0^\infty \mathbb{P}[E_1 \wedge E_2|\nu]dF_{\mathcal{D}}(\nu) \\ &= \int_0^\infty \mathbb{P}[E_1|\nu]\mathbb{P}[E_2|\nu]dF_{\mathcal{D}}(\nu) \end{aligned}$$

which does not depend on τ, and similarly

$$\mathbb{P}[E_i] = \int_0^\infty \mathbb{P}[E_i|\nu]dF_{\mathcal{D}}(\nu)$$

does not depend on τ, so that $p_{12} - p_1p_2 = \mathbb{P}[E_1 \wedge E_2] - \mathbb{P}[E_1]\mathbb{P}[E_2]$ does not depend on τ either.

Since ρ_{ij} does not depend on the length of the edge between T_i and T_j in the rates-across-sites case, we may rearrange the tree on the groups without changing the value of ρ_{ij}, so the tree on the groups is not uniquely determined by ρ. In the covarion case, if the edge between T_x and T_y has total length τ_{xy} then

$$\begin{aligned} \rho_{xy} &= -\ln|p_{xy} - p_xp_y| \\ &= -\ln(\sigma_1\sigma_2 e^{-(s_1+s_2)\tau_{xy}}|\delta_x||\delta_y|) \\ &= -\ln(\sigma_1\sigma_2) + (s_1+s_2)\tau_{xy} - \ln|\delta_x| - \ln|\delta_y|. \end{aligned}$$

Referring to Figure 5 we have $\tau_{ij} = \tau_i + \tau_j$, $\tau_{ik} = \tau_i + \tau_m + \tau_k$ etc., and Theorem 5.2 follows.

□

We conclude by giving some examples of separable events. Details and results in greater generality appear in [**32**].

THEOREM 5.3 (Some separable events). *The events E^s and E^d above are separable under the covarion model (R, S), and satisfy the independence condition under the rates-across-sites model $(R, \mathcal{D})$, if R has one of the following forms:*

1. *$R = R_K$, where*

$$R_K = \begin{pmatrix} -\delta & \alpha & \beta & \gamma \\ \alpha & -\delta & \gamma & \beta \\ \beta & \gamma & -\delta & \alpha \\ \gamma & \beta & \alpha & -\delta \end{pmatrix}$$

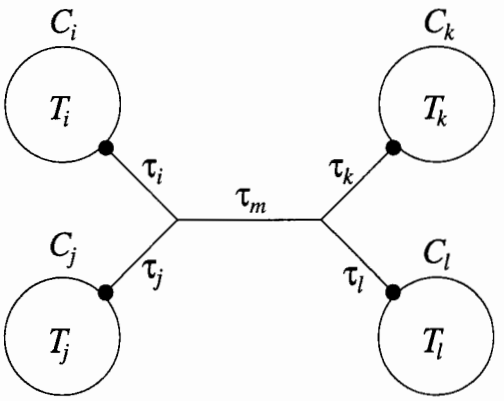

FIGURE 5. The tree on the four monophyletic groups of species C_i, C_j, C_k and C_l. The τ_x are the edge lengths.

and is the form of the matrix used in the K3ST model. This includes as special cases the Kimura two parameter model ($\beta = \gamma$) and the Jukes-Cantor model ($\alpha = \beta = \gamma$).

2. *R is the $r \times r$ matrix given by $R_{ij} = \alpha$ if $i \neq j$ and $R_{ii} = (1 - r)\alpha$, for any r. This gives the fully symmetric model, and includes as particular cases the Cavender-Farris model ($r = 2$) and the Jukes-Cantor model ($r = 4$).*

CONCLUSION

In this paper we have investigated some properties of maps for constructing tree metrics, and contrasted two models of nucleotide substitution. A feature of such models (including the two we described) is that it is often possible to construct distance functions on the sequences which converge in probability to a tree metric (realised by the underlying evolutionary tree), as the sequence length tends to infinity. This is useful because, provided the underlying tree is fully resolved, then any *good* map (as described in Part One) applied to these distance function will reconstruct the correct tree with high probability for sufficiently long sequences. While this might be interpreted as a further advertisement for the virtues of a map being *good*, one in fact requires continuity only at points of $\mathcal{T}(S)$ corresponding to fully resolved trees, and most methods, including neighbor-joining, satisfy this condition. (In case the underlying tree is not fully resolved, then the reconstructed tree will be, with high probability, a refinement of the underlying tree, for sufficiently long sequences, and the maximal weight assigned to any edges that are not in the underlying tree will tend to 0). An important additional issue is the *rate* of convergence of the distance function to an additive metric, and the consequent sequence length required to reconstruct the underlying tree with high probability. Such issues have been addressed by other authors in this book.

References

[1] H.-J. Bandelt, *Recognition of tree metrics*, SIAM J. Discrete Math. **3** (1990), 1–6.

[2] H.-J. Bandelt and A. Dress, *Reconstructing the shape of a tree from observed dissimilarity data*, Advances in Applied Mathematics **7** (1986), 309–343.
[3] ———, *A canonical decomposition theory for metrics on a finite set*, Advances in Mathematics **92** (1992), 47–105.
[4] H.-J. Bandelt and M. A. Steel, *Symmetric matrices representable by weighted trees over a cancellative abelian monoid*, SIAM J. Discrete Math. **8** (1995), no. 4, 517–525.
[5] B. Bowditch, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, Group Theory from a Geometrical Viewpoint (E. Ghys, A. Haefliger, and A. Verjovsky, eds.), World Scientific Publishing Co. Pte. Ltd., Singapore, New Jersey, London, Hong Kong, 1991, pp. 64–167.
[6] P. Buneman, *The recovery of trees from measures of dissimilarity*, Mathematics in the archaeological and historical sciences (F. R. Hodson, D. G. Kendall, and P. Tautu, eds.), Edinburgh University Press, 1971, pp. 387–395.
[7] J. T. Chang, *Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters*, Mathematical Biosciences **134** (1996), 189–215.
[8] J. N. Darroch and K. W. Morris, *Passage time generating functions for continuous-time finite Markov chains*, Journal of Applied Probability **5** (1968), 414–426.
[9] A. Dress, V. Moulton, and W. Terhalle, *T-theory – an overview*, The European Journal of Combinatorics **17** (1996), 161–175.
[10] M. Farach and S. Kannan, *Efficient algorithms for inverting evolution*, Proceedings of the 1996 ACM Symposium on the Foundations of Computer Science, 1996.
[11] J. Felsenstein, *Cases in which parsimony or compatibility will be positively misleading*, Systematic Zoology **27** (1978), 401–410.
[12] W. M. Fitch, *The nonidentity of invariable positions in the cytochrome* c *of different species*, Biochemical Genetics **5** (1971), 231–241.
[13] ———, *Rate of change of concomitantly variable codons*, J. Mol. Evol. **1** (1971), 84–96.
[14] W. M. Fitch and F. J. Ayala, *The superoxide dismutase molecular clock revisited*, Proc. Natl. Acad. Sci. USA **91** (1994), 6802–6807.
[15] W. M. Fitch and E. Markowitz, *An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution*, Biochemical Genetics **4** (1970), 579–593.
[16] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*, Clarendon Press, Oxford, 1982.
[17] M. Gromov, *Hyperbolic groups*, Essays in Group Theory (S. Gerstin, ed.), MSRI Publ., no. 8, Springer, 1987, pp. 75–263.
[18] X. Gu and W.-H. Li, *A general additive distance with time-reversibility and rate variation among nucleotide sites*, Proc. Natl. Acad. Sci. USA **93** (1996), 4671–4676.
[19] D. Gusfield, *Efficient algorithms for inferring evolutionary trees*, Networks **21** (1991), 19–28.
[20] D. Hillis, C. Moritz, and K. Barbara, *Molecular systematics*, second ed., Sinauer Associates Inc., 1996.
[21] D. Huson, V. Moulton, and M. Steel, *Retractions as a tool for analyzing distance functions in biology*, in preparation.
[22] N. Jardine and R. Sibson, *The construction of hierarchic and non-hierarchic classifications*, The Computer Journal **11** (1968), 177–184.
[23] T. H. Jukes and C. R. Cantor, *Evolution of protein molecules*, Mammalian protein metabolism (H. N. Munro, ed.), Academic Press, New York, 1969, pp. 21–132.
[24] J. Keilson, *Markov chain models—rarity and exponentiality*, Applied Mathematical Sciences, vol. 28, Springer-Verlag, 1979.
[25] M. Kimura, *A simple method of estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences*, J. Mol. Evol. **16** (1980), 111–120.
[26] ———, *Estimation of evolutionary distances between homologous nucleotide sequences*, Proc. Nat. Acad. Sci. USA **78** (1981), 454–458.
[27] M. M. Miyamoto and W. M. Fitch, *Testing the covarion hypothesis of molecular evolution*, Mol. Biol. Evol. **12** (1995), no. 3, 503–513.
[28] V. Moulton and M. Steel, *Retractions of finite distance functions onto tree metrics*, submitted to SIAM J. Discrete Math.
[29] D. Robinson and L. Foulds, *Comparison of weighted labelled trees*, Combinatorial Mathematics VI, Lecture Notes in Mathematics, vol. 748, Springer, 1979, pp. 119–129.

[30] Y. A. Smolensky, *A method for linear recording of graphs*, USSR Comput. Math. Phys. **2** (1969), 396–397.
[31] M. A. Steel, L. A. Székely, and M. D. Hendy, *Reconstructing trees when sequence sites evolve at variable rates*, Journal of Computational Biology **1** (1994), no. 2, 153–163.
[32] C. Tuffley and M. Steel, *Modelling the covarion hypothesis of nucleotide substitution*, submitted to Mathematical Biosciences, 1996.
[33] P. J. Waddell and M. A. Steel, *General time reversible distances with unequal rates across sites*, Research report 143, Dept. of Mathematics and Statistics, University of Canterbury, New Zealand, 1996.
[34] K. Wolf and P. Degens, *On properties of additive tree algorithms*, Conceptual and Numerical Analysis of Data, Proc. of the 13th conf. of the Gesellschaft für Klasification, Springer-Verlag, 1989.
[35] Z. Yang, *Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ across sites*, Mol. Biol. Evol. **10** (1993), 1396–1401.
[36] K. A. Zaretsky, *Reconstruction of a tree from the distances between its pendant vertices*, Uspekhi Math. Nauk (Russian Mathematical Surveys) **20** (1965), 90–92 (Russian).

BIOMATHEMATICS RESEARCH CENTRE, UNIVERSITY OF CANTERBURY, CHRISTCHURCH, NEW ZEALAND

BIOMATHEMATICS RESEARCH CENTRE, UNIVERSITY OF CANTERBURY, CHRISTCHURCH, NEW ZEALAND
E-mail address: m.steel@math.canterbury.ac.nz

BIOMATHEMATICS RESEARCH CENTRE, UNIVERSITY OF CANTERBURY, CHRISTCHURCH, NEW ZEALAND