ELSEVIER

Modeling the Covarion Hypothesis of Nucleotide Substitution

CHRIS TUFFLEY AND MIKE STEEL
Biomathematics Research Centre, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

Received 31 October 1996; revised 10 June 1997

ABSTRACT

A "covarion" model for nucleotide substitution that allows sites to turn "on" and "off" with time was proposed in 1970 by Fitch and Markowitz. It has been argued recently that evidence supports such models over later, alternative models that postulate a static distribution of rates across sites. However, in contrast with these latter well-studied models, little is known about the analytic properties of the former model. Here we analyze 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 gives identical results to a suitably chosen rates-across-sites model if several sequences are compared in pairs by using only the expected proportion of sites at which they differ, (iii) conditions under which the two models will give identical results if the full joint probability matrix is examined, and (iv) that the two models can, in principle, be distinguished when there are at least four monophyletic groups of species. This last result is based on a distance measure that is tree additive under certain versions of the covarion model but, in general, will not be additive under a rates-across-sites model. The measure constructed does not require knowledge of the parameters of the model and so shows that sequences generated by the covarion model do in fact contain information about the underlying tree.

1. INTRODUCTION

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 to determine cases for which existing methods are likely to lead to erroneous results —see, for example, [1].

The simplest and earliest models assume that each site evolves independently and is identically distributed (i.i.d.) at the same rate and according to simple Markov-style assumptions. However, this single-rate

MATHEMATICAL BIOSCIENCES 147:63–91 (1998)

655 Avenue of the Americas, New York, NY 10010
0025-5564/98/$19.00
PII S0025-5564(97)00081-3

assumption appears to be unrealistic, and accordingly models incorporating some variation of rates across sites have been proposed and studied [2–5] to take into account different functional constraints at different sites. An alternative approach to accounting for differing selective constraints is Fitch and Markowitz's "concomitantly variable codons," or "covarion," hypothesis [6]. This approach proposes that, at any given time, some sites are invariable owing to functional or structural constraints but, as mutations are fixed elsewhere in the sequence, these constraints may change, so sites that were previously invariable may become variable and vice versa. The pool of variable sites is therefore changing with time (Figure 1). Since 1970, it has been argued that evidence supports the covarion hypothesis, both on biochemical grounds and by providing a better description of certain data [7–9]. However, in contrast with the rates-across-sites models, little is known about the analytic properties of covarion-style models.

In this paper, we present and analyze 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 only the behavior 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. With the use of techniques from the theory of Markov processes, such a model can be analyzed and compared with rates-across-sites models in terms of the expected frequencies of site patterns that the models should generate. This is the first step in comparing the two models because, if they cannot be distinguished with the use of infinite sequences, there is no prospect of distinguishing between them with finite sequences.

The i.i.d. assumption may be justified as an approximation to the covarion hypothesis by the following remarks. We are concerned here with the limiting frequency of site patterns in sequences as the sequence length becomes large and without reference to the order in which the patterns occur along the sequence. If the dependency between sites is spatially localized (perhaps under some reordering of the sites), then the frequencies of the patterns will converge toward those generated under an i.i.d. model. This follows from an argument similar to the proof of Bernstein's theorem—see, for example, Rényi [10], page 379—which requires only that the correlation between the sites, reordered if necessary, falls off sufficiently quickly. In our setting, the assumption of local dependency between sites is reasonable. This type of approach is already commonly employed (albeit tacitly) when modeling a distribution of rates across sites. In real sequences, high rates are

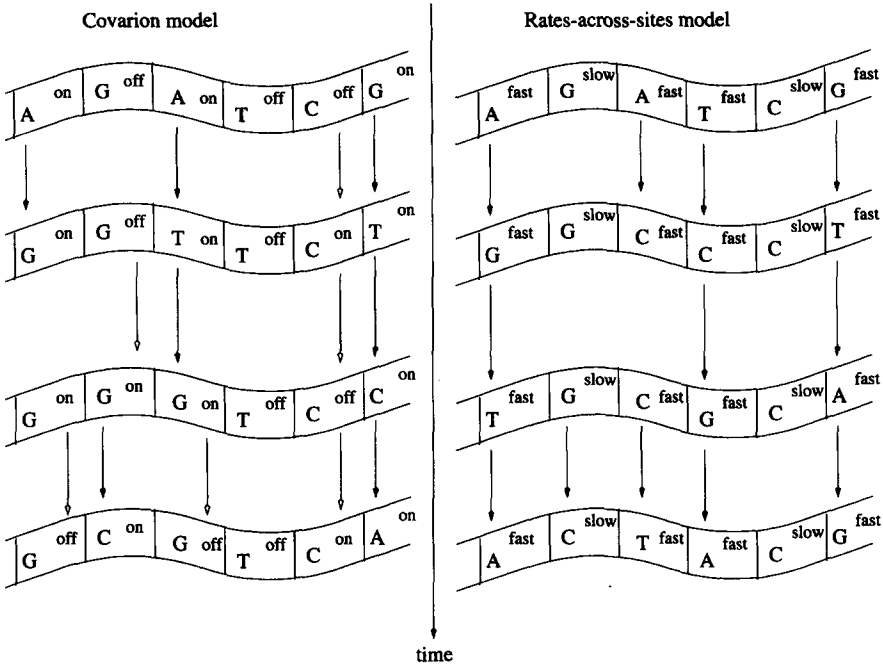

FIG. 1. Contrasting a covarion-style process and rates across sites. Under a covarion-style process, each site is either "on" or "off." Sites that are off are unable to change state but may later turn on (owing to state changes elsewhere in the sequence) and be able to change. Under rates across sites, 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.

often associated with particular positions in the sequence (such as the third position in a codon) or proximity to other high rate sites (as in hypervariable regions), so the sites are clearly neither independent nor identically distributed. Nevertheless, because the dependency is local, it is usual to suppose that the rate at each site is chosen i.i.d. from some distribution, and the resulting i.i.d. model produces pattern frequencies indistinguishable from those of the original model as the sequence length tends to infinity.

In Section 3.1, we find an expression for the joint probability matrix of states for two species separated by an evolutionary distance τ. This allows τ to be determined from the expected proportion of sites where two species differ. Such relationships are useful to the biologist, because the evolutionary distance between pairs of species allows for both the underlying evolutionary tree and its edge lengths to be recovered. We

then compare the joint probability matrix with the equivalent expression under a rates-across-sites model in Section 3.4 to address the question of whether the covarion model will give different results from those of rates across sites when several sequences are analyzed by comparing each pair in turn. We show that the covarion model gives identical results with those of a suitably chosen rates-across-sites model if only the trace of the joint probability matrix (i.e., the probability that the two species are in the same state at a given site) is considered and give a partial answer to the question of when the two models can give identical results if the full joint probability matrix is considered.

In Section 4, we show that the two models can, in principle, be distinguished when there are at least four monophyletic groups of species. This result is based on the construction of a distance measure that is tree additive under certain versions of the covarion model but, in general, will not be additive under a rates-across-sites model. The measure constructed does not require knowledge of the parameters of the model and so shows that sequences generated by the covarion model do in fact contain information about the structure of the underlying tree.

2. THE MODELS

2.1. *A COVARION-STYLE MODEL*

We model a covarion-style process with two parts: (1) a "switch" process and (2) 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 $\mathscr{S} = \{\texttt{on}, \texttt{off}\}$ 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

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

so it is stationary and time reversible. For a background in Markov processes, the reader is referred to [11] and [12].

While the switch is in state "off," the observable process is unable to change state; however, when the switch is in state "on," the observ-

able process is governed by a second stationary and time-reversible Markov process with state space $\mathscr{A} = \{1, \ldots, r\}$, rate matrix R satisfying $R_{ij} > 0$ if $i \neq j$, and initial distribution π. Stationarity and time reversibility are equivalent to the conditions

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

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 $\mathscr{A} \times \mathscr{O}$ [which we identify with $[2r]$ according to $(i, \texttt{on}) \mapsto i$, $(i, \texttt{off}) \mapsto i + r$], initial distribution $\pi' = (\sigma_1 \pi_1, \ldots, \sigma_1 \pi_r, \sigma_2 \pi_1, \ldots, \sigma_2 \pi_r)$ 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. We assume that we are unable to distinguish between the states $(i, \texttt{on})$ and $(i, \texttt{off})$. With the use of this formulation, the probability of observing each site pattern (given a tree and the values of the parameters) can be calculated for such purposes as maximum likelihood estimation. As usual, if each edge e of the tree is given a nonnegative weight τ_e, the transition matrices P^e are given by

$$P^e = \exp(\tau_e R').$$

The probability of generating a particular pattern is given by a sum over all possible assignments of states in $\mathscr{A} \times \mathscr{O}$ to the remaining vertices of the tree. In practice, this can be found quickly by using a simple modification of the usual dynamic programming technique.

It is easy to check that R' is stationary and time reversible whenever R and S are. Both formulations lead to the same random process with state space $\mathscr{A}$, and this random process is not itself Markov.

2.2. *RATES ACROSS SITES*

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

Each site evolves according to rate matrix νQ where ν is chosen i.i.d. according to $\mathscr{D}$. The rate at a given site is assumed constant across the whole tree. This kind of model has been well studied; see, for example, [2–5].

3. THE TWO-TAXA TREE

Here we calculate the joint probability matrix for the two-taxa tree (i.e., the matrix whose ij entry is the probability that taxon 1 is in state i and taxon 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.

3.1. *UNDER THE COVARION MODEL*

The joint probability matrix may be calculated by using either of the two formulations of the covarion model. We present the calculation by using the first formulation. With the use of the second formulation the ij entry of this matrix is found by summing the probability that taxa 1 is in state (i, o_1) and taxa 2 is in state (j, o_2) for $o_i = \mathtt{on}$, $o_i = \mathtt{off}$, $i = 1,2$.

Time reversibility implies that we may assume the tree to be 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 (i.e., 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

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

Reversibility allows us to obtain a spectral representation of $J(t)$—see Keilson [12], pages 32–35. Because ΠR is symmetric, so is $\Pi^{1/2} R \Pi^{-1/2}$, which therefore has real eigenvalues $\{\lambda_j\}$ and orthonormal eigenvectors $\{u_j\}$ (related to the eigenvectors $\{v_j\}$ of R by $v_j = \Pi^{-1/2} u_j$ and $Rv_j = \lambda_j v_j$). 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

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

Darroch and Morris [13] give the moment generating function $\mathsf{E}[e^{\lambda X(\tau)}]$ of $X(\tau)$ by

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

where $D = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Diagonalizing $S + \lambda D$ we obtain the following lemma.

LEMMA 1

The joint probability matrix $J_C(\tau)$ is given by

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

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

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

and

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

We note that by examining $I + \frac{1}{k}R$, where $k > \max\{|R_{ii}|\}$, and using the Perron-Frobenius theorem [11], page 134, it can be shown that the eigenvalues of R are nonnegative, with zero occurring as an eigenvalue exactly once.

A common measure of the extent to which two sequences differ is the proportion of sites at which they disagree, known as the *dissimilarity*. The expected proportion of such sites is given by one minus the trace of the joint probability matrix. From Equation (1) we obtain

$$\operatorname{trace}(J_C(\tau)) = \sum_{j=1}^{r} \left[c_j^+ e^{\mu_j^+ \tau} + c_j^- e^{\mu_j^- \tau}\right] \operatorname{trace}(w_j w_j^{\mathrm{T}}).$$

For the zero eigenvalue $\lambda_1 = 0$ we have $\mu_1^+ = 0$, $\mu_1^- = -(s_1 + s_2)$ and $w_1 = \pi^T$, so we have the following lemma.

LEMMA 2

The probability that two sequences have the same state at a given site is given by

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

To proceed any further with this calculation, we need to be able to calculate $\text{trace}(w_j w_j^T) = \text{trace}(\Pi u_j u_j^T)$ for the remaining eigenvalues, which requires some knowledge of R. However, in the case of the equifrequent stationary distribution $\pi = (1/r, \ldots, 1/r)$ we have simply $\text{trace}(w_j w_j^T) = \text{trace}(\frac{1}{r} u_j u_j^T) = 1/r$, so

$$\text{trace}(J_C(\tau)) = \frac{1}{r} + \frac{1}{r} \sum_{j=2}^{r} \left[c_j^+ e^{\mu_j^+ \tau} + c_j^- e^{\mu_j^- \tau}\right].$$

We conclude this section by establishing some properties of the coefficients in Equation (1) that are helpful in determining the behavior of the covarion model.

LEMMA 3

(i) μ^+ *and* μ^- *are real increasing functions of* λ *satisfying* $\mu^- \leqslant -(s_1 + s_2) < -s_2 < \mu^+ \leqslant 0$ *on* $(-\infty, 0]$.

(ii) $c_j^+, c_j^- \geqslant 0$ (*with equality only for* $\lambda = 0$, *when* $c^- = 0$) *and* $c_j^+ + c_j^- = 1$.

(iii) $\text{trace}(w_j w_j^T) > 0$ *and* $\sum_{j=1}^{r} \text{trace}(w_j w_j^T) = 1$.

Proof. (i) Suppose $\lambda_1 < \lambda_2$ and consider the functions

$$f_{\lambda_j}(\mu) = \mu^2 + (s_1 + s_2 - \lambda_j)\mu - s_2 \lambda_j.$$

If $f_{\lambda_1}(\mu) = f_{\lambda_2}(\mu)$ then we find $\mu = -s_2$, at which point $f_{\lambda_j}(-s_2) = -s_1 s_2$ and $f'_{\lambda_j}(-s_2) = s_1 - s_2 - \lambda_j$. Thus the situation is as illustrated in Figure 2, and, because μ_j^+ and μ_j^- are the roots of $f_{\lambda_j} = 0$ it follows that $\mu_1^- < \mu_2^-$ and $\mu_1^+ < \mu_2^+$. For the inequalities, we have $\mu^- = -(s_1 + s_2)$, $\mu^+ = 0$ when $\lambda = 0$, and $f_{\lambda_j}(-s_2) = -s_1 s_2 < 0$ so $\mu^- < -s_2 < \mu^+$, because f_{λ_j} are right-way-up parabolas.

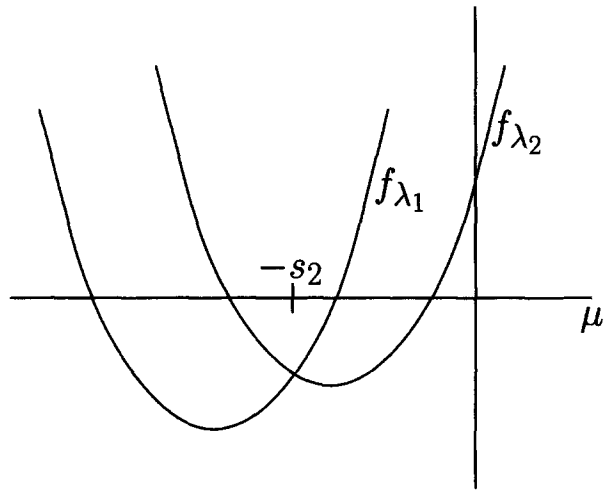

FIG. 2. The function f_{λ_1} lies above f_{λ_2} on $(-s_2,\infty)$ and below on $(-\infty,-s_2)$, with $f_{\lambda_1}(-s_2)=f_{\lambda_2}(-s_2)<0$. Hence $\mu_1^- < \mu_2^- < \mu_1^+ < \mu_2^+$.

The inequalities in (ii) follow from (i), and the equality $c_j^+ + c_j^- = 1$ may be verified directly. For (iii), we have $\text{trace}(w_j w_j^{\mathrm{T}}) = w_j^{\mathrm{T}} = |w_j|^2 > 0$ and

$$\sum_{j=1}^{r} \text{trace}\left(w_j w_j^{\mathrm{T}}\right) = \text{trace} \sum_{j=1}^{r} w_j w_j^{\mathrm{T}} = \text{trace}(J(0)) = \text{trace}(\Pi) = 1. \quad \blacksquare$$

3.2. *UNDER RATES ACROSS SITES*

In the rates-across-sites case, put $\Theta = \text{diag}(\theta)$ and let Q have eigenvalues $\{\alpha_j\}$. Argued 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

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

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

$$J_D(\tau) = \sum_{j=1}^{r} M(\alpha_j \tau) z_j z_j^{\mathrm{T}} \tag{3}$$

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

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

This calculation is not new, and similar or equivalent calculations appear in other papers dealing with rates-across-sites models, such as [2, 14, 15].

As in the covarion case, the probability that the two sequences have the same state at a given site is given by

$$\operatorname{trace}(J_D(\tau)) = \theta\theta^{\mathrm{T}} + \sum_{j=2}^{r} M(\alpha_j \tau) \operatorname{trace}\left(z_j z_j^{\mathrm{T}}\right); \tag{4}$$

in the equifrequent stationary distribution case $\theta = (1/r, \ldots, 1/r)$, this is

$$\operatorname{trace}(J_D(\tau)) = \frac{1}{r} + \frac{1}{r} \sum_{j=2}^{r} M(\alpha_j \tau).$$

3.3. *RECOVERING THE EVOLUTIONARY DISTANCE UNDER THE TWO MODELS*

Equation (3) may be written

$$J_D(\tau) = \Theta M(\tau Q), \tag{5}$$

where M is the moment generating function of $\mathscr{D}$ 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 by

$$K = -\operatorname{trace}\left\{\Theta\left[M^{-1}\left(\Theta^{-1} J_D(\tau)\right)\right]\right\} \tag{6}$$

[14, 15]. Here M^{-1} is the inverse of the moment-generating function, again applied to matrices. This expression gives a treelike distance and, because row i of $J_D(\tau)$ sums to θ_i, requires knowledge only of $\mathscr{D}$ to reconstruct the tree from $J_D(\tau)$.

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

$$K = -\operatorname{trace}(\Theta Q) f_D^{-1}(\operatorname{trace}(J_D(\tau)), \tag{7}$$

where $f_D(\tau) = \text{trace}(J_D(\tau))$ is given by Equation (4). Note that f_D^{-1} exists because f_D is monotone decreasing.

The property of Equation (3) that allows it to be written in the form of Equation (5)—namely, M is applied to products of the form $\alpha_j\tau$—does not hold for Equation (1), and it appears that a transformation analogous to Equation (6) does not exist for the covarion model. However, if R and S are known (or estimated), then, as in Equation (7), 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 (2). Again f_C is monotone decreasing (by Lemma 3), 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 $\hat{J}$. Because J_C and J_D are both symmetric, it is usual practice to take the symmetrized matrix $(\hat{J} + \hat{J}^{\mathrm{T}})/2$ as the estimate.

3.4. *PAIR-BY-PAIR COMPARISONS OF SEQUENCES*

Simultaneous pair-by-pair comparisons of several sequences are frequently used as a method of building trees—for example, through the construction of treelike distances. Here we address the issue of whether the covarion model will give different results to rates across sites when making such comparisons. 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 $D = (R, \mathscr{D})$, so the covarion model gives results identical with those of a suitably chosen rates-across-sites model if only one pair of sequences is examined. However, the distribution of $X(\tau)/\tau$ depends on τ, which opens the possibility that the models may give different results if more than one pair is considered.

A common measure of the dissimilarity of two sequences is one minus the trace of the joint probability matrix, which is the probability that they disagree at a given site. In applications, this is estimated from the proportion of sites at which the aligned sequences from the two taxa differ. We show here (Theorem 5) that, given any covarion model, there is always a rates-across-sites model that will generate exactly the same data if only the trace is considered. We also characterize the conditions (Theorem 6) under which $C = (R, S)$ and $D = (Q, \mathscr{D})$ satisfy $J_C(\tau) = J_D(\tau)$ for all τ. Models satisfying this equality will give identical results under any form of pair-by-pair comparison and on any tree; however, models that do not satisfy this equality may still give identical results on certain trees.

A related question is whether it is possible to distinguish between the covarion model and rates across sites on the basis of simultaneous pair-by-pair comparisons of several sequences. On this question, the results of this section are largely negative and suggest that pair-by-pair comparisons are inadequate for distinguishing the covarion model from rates across sites. Thus, a test of the two models will probably require the simultaneous comparison of three or more sequences. Section 4 gives an alternative approach to distinguishing between the covarion and rates-across-sites models based on such a comparison.

We begin with a preliminary result. Stationary and reversible rate matrices with exactly one distinct nonzero eigenvalue will be of relevance to us in what follows, so we give a characterization of them here.

LEMMA 4

(i) *Given a distribution π of states, there is a stationary and reversible rate matrix R_π having π as its stationary distribution and possessing exactly one distinct nonzero eigenvalue — namely,*

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

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

(ii) *If the stationary and reversible rate matrix R with stationary distribution π and $|R_{ij}| > 0$ for all i, j has exactly one distinct nonzero eigenvalue $-\lambda$, then $R = \lambda R_\pi$.*

Proof. (i) We have $(\mathbf{1}\pi - I_r)\cdot\mathbf{1} = \mathbf{1}(\pi\mathbf{1}) = \mathbf{1} - \mathbf{1} = \mathbf{0}$, so the rows of R_π sum to zero. All the off-diagonal entries are positive and therefore R_π is a rate matrix. $\pi R_\pi = \pi(\mathbf{1}\pi - I_r) = (\pi\mathbf{1})\pi - \pi = \pi - \pi = \mathbf{0}^{\mathrm{T}}$, so R_π has stationary distribution π, and if $i \neq j$ then $(R_\pi)_{ij} = \pi_j$, so $\pi_i(R_\pi)_{ij} = \pi_i\pi_j = \pi_j(R_\pi)_{ji}$. Hence R_π is reversible.

The matrix $\mathbf{1}\pi$ has rank 1 and hence null space of dimension $r-1$, so 0 is an eigenvalue of multiplicity $r-1$. The remaining eigenvalue is 1 because $(\mathbf{1}\pi)\mathbf{1} = \mathbf{1}$. Hence R_π has eigenvalues -1 (multiplicity $r-1$) and 0 (multiplicity 1).

(ii) Consider the matrix $Q = I_r + \frac{1}{\lambda}R$, which has eigenvalues 0 (multiplicity $r-1$) and 1 (multiplicity 1). By the reversibility assumption, R has a full complement of eigenvectors, so Q has null space of dimension $r-1$ and hence rank 1. Further, Q has row sums equal to 1, from which $Q = \mathbf{1}v$, where $\sum_{i=1}^{r} v_i = 1$. In fact, we must have $v = \pi$, because the left eigenvector corresponding to 1 is π. Hence $R = \lambda(Q - I_r) = \lambda(\mathbf{1}\pi - I_r) = \lambda R_\pi$.

THEOREM 5

For any covarion model C, there is a rates-across-sites model D such that

$$\operatorname{trace}(J_D(\tau)) = \operatorname{trace}(J_C(\tau))$$

for all $\tau \geqslant 0$.

Note that this result does not imply that, given a rates-across-sites model D, there is necessarily a covarion model C for which the preceding equality holds.

Proof. By Equation (2) and Lemma 3, $\operatorname{trace}(J_C(\tau))$ has the form

$$\operatorname{trace}(J_C(\tau)) = \pi\pi^{\mathrm{T}} + \sum_{j=2}^{r} \left[c_j'^{+} e^{\mu_j^{+}\tau} + c_j'^{-} e^{\mu_j^{-}\tau} \right],$$

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 nonzero eigenvalues, we may collect terms in $e^{\mu^{\pm}\tau}$ for each eigenvalue, writing $\operatorname{trace}(J_C(\tau))$ in the form

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

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

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

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

Then $\mathscr{D}$ is well defined and, if $D = (R_\pi, \mathscr{D})$, then, by Equation (4) and Lemmas 3 and 4,

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

■

THEOREM 6

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

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

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

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

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

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

Proof. Suppose $J_C(\tau) = J_D(\tau)$ for all τ. Because they agree for $\tau = 0$, when $J_C(\tau) = \Pi$ and $J_D(\pi) = \Theta$, we must have $\theta = \pi$. Multiply $J_C(\tau) = J_D(\tau)$ on the left and right by $\Pi^{-1/2}$ to get

$$\sum_{j=1}^{r} C_j(\tau) u_j u_j^{\mathrm{T}} = \sum_{j=1}^{r} M(\alpha_j \tau) y_j y_j^{\mathrm{T}},$$

where $C_j(\tau) = c_j^+ e^{\mu_j^+ \tau} + c_j^- e^{\mu_j^- \tau}$. Now $u_j^{\mathrm{T}} u_k = \delta_{jk}$ implying $\Pi^{-1/2} J_C(\tau) \Pi^{-1/2}$ has eigenvalues $\{C_j(\tau)\}$ and corresponding eigenvectors $\{u_j\}$. Similarly $\Pi^{-1/2} J_D(\tau) \Pi^{-1/2}$ has eigenvalues $\{M(\alpha_j \tau)\}$. So there must be some ordering for which $C_j(\tau) = M(\alpha_j \tau)$ for each j. We will suppose that the functions have been ordered in this way.

Write $M_j(\tau)$ for $M(\alpha_j \tau)$. For the zero eigenvalue ($j = 1$), we have $C_1(\tau) = 1 = M_1(\tau)$, so we need worry about only the nonzero eigenvalues ($j \geqslant 2$). The M_j ($j \geqslant 2$) have the property that $M_k(\alpha_l \tau / \alpha_k) = M_l(\tau)$; that is, we may transform from one to another simply by rescaling τ. Clearly the C_j must satisfy this also. Suppose $C_k(\gamma\tau) = C_l(\tau)$. Then

$$C_k(\gamma\tau) = c_k^+ e^{\mu_k^+ \gamma\tau} + c_k^- e^{\mu_k^- \gamma\tau} = C_l(\tau) = c_l^+ e^{\mu_l^+ \tau} + c_l^- e^{\mu_l^- \tau},$$

so we must have $\gamma\mu_k^+ = \mu_l^+$, $\gamma\mu_k^- = \mu_l^-$, $c_k^+ = c_l^+$, and $c_k^- = c_l^-$ because exponential functions are independent (note that $\mu_j^- < \mu_j^+$ which precludes the possibility of matching $\gamma\mu_k^+$ with μ_l^-, etc.). Hence, from the

definition of c_l^-, we have

$$\begin{aligned} c_l^- &= \frac{(s_1 + s_2 + \mu_l^-)\mu_l^+}{(s_1 + s_2)(\mu_l^+ - \mu_l^-)} \\ &= \frac{(s_1 + s_2 + \gamma\mu_k^-)\gamma\mu_k^+}{(s_1 + s_2)(\gamma\mu_k^+ - \gamma\mu_k^-)} \\ &= \frac{(s_1 + s_2 + \gamma\mu_k^-)\mu_k^+}{(s_1 + s_2)(\mu_k^+ - \mu_k^-)}, \end{aligned}$$

which by hypothesis equals

$$c_k^- = \frac{(s_1 + s_2 + \mu_k^-)\mu_k^+}{(s_1 + s_2)(\mu_k^+ - \mu_k^-)}.$$

Hence $\gamma = 1$. Lemma 3 (i) then implies that $\lambda_k = \lambda_l$, and it follows that R has only one distinct nonzero eigenvalue λ. Now $M_k(\tau) = C_k(\tau) = C_j(\tau) = M_j(\tau), 2 \leqslant j, k \leqslant r$, so Q has only one distinct nonzero eigenvalue also and, because both R and Q have stationary distribution π, by Lemma 4 both are scalar multiples of R_π. $\mathscr{D}$ has moment-generating function $M(\tau) = c_\lambda^+ e^{\mu_\lambda^+ \tau} + c_\lambda^- e^{\mu_\lambda^- \tau}$ and so is two rate with both rates greater than zero.

Conversely, if $R = -\lambda R_\pi$, then

$$J_C(\tau) = \pi^{\mathrm{T}}\pi + \left[c_\lambda^+ e^{\mu_\lambda^+ \tau} + c_\lambda^- e^{\mu_\lambda^- \tau}\right] \sum_{j=2}^{r} w_j w_j^{\mathrm{T}}.$$

Let $\mathscr{D}$ be the two-rate distribution such that

$$\mathsf{P}[\nu = |\mu_\lambda^*|] = c_\lambda^*, \qquad * = +, -.$$

Then $\mathscr{D}$ is well defined and, if $D = (R_\pi, \mathscr{D})$, we have

$$\begin{aligned} J_D(\tau) &= \pi^{\mathrm{T}}\pi + \sum_{j=2}^{r} M(-\tau) w_j w_j^{\mathrm{T}} \\ &= \pi^{\mathrm{T}}\pi + \left[c_\lambda^+ e^{\mu_\lambda^+ \tau} + c_\lambda^- e^{\mu_\lambda^- \tau}\right] \sum_{j=2}^{r} w_j w_j^{\mathrm{T}} \\ &= J_C(\tau). \end{aligned}$$

It remains to show that, if $D = (\gamma R_\pi, \mathscr{D})$, where $\mathscr{D}$ is a two-rate distribution such that

$$\mathsf{P}[\nu = \nu_i] = \rho_i, \qquad i = 1, 2,$$

then we may choose a covarion model $C = (R, S)$ such that $J_C(\tau) = J_D(\tau)$ for all τ. By scaling ν_1 and ν_2, if necessary we may assume that $\gamma = 1$, and $0 < \nu_1 < \nu_2$. We must then find $\lambda < 0$ and $s_1, s_2 > 0$ such that

$$\mu_\lambda^+ = -\nu_1, \quad \mu_\lambda^- = -\nu_2 \quad \text{and} \quad c_\lambda^- = \rho_2$$

and then take $R = -\lambda R_\pi$ (note that the third condition implies $c_\lambda^+ = \rho_1$). Using

$$(\mu + \nu_1)(\mu + \nu_2) = (\mu - \mu_\lambda^+)(\mu - \mu_\lambda^-) = \mu^2 + (s_1 + s_2 - \lambda)\mu - s_2\lambda,$$

we obtain the system of equations

$$\nu_1\nu_2 = -s_2\lambda,$$

$$\nu_1 + \nu_2 = s_1 + s_2 - \lambda,$$

$$\rho_2 = \frac{(s_1 + s_2 - \nu_2)\nu_1}{(\nu_1 - \nu_2)(s_1 + s_2)},$$

which may be solved (uniquely) to give

$$\lambda = -\frac{\nu_1^2\rho_1 + \nu_2^2\rho_2}{\nu_1\rho_1 + \nu_2\rho_2},$$

$$s_1 = \frac{\rho_1\rho_2\nu_1\nu_2(\nu_1 - \nu_2)^2}{(\nu_1\rho_1 + \nu_2\rho_2)(\nu_1^2\rho_1 + \nu_2^2\rho_2)},$$

$$s_2 = \frac{\nu_1\nu_2(\nu_1\rho_1 + \nu_2\rho_2)}{\nu_1^2\rho_1 + \nu_2^2\rho_2}.$$

This defines the required covarion model.

■

3.5. *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 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 invariable

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 to spend about the same amount of time in state "on." The expected time in state "on" is $\sigma_1\tau$ and, in the limiting case $s_1, s_2 \to \infty$ with s_1/s_2 constant, we find that

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

4. A TREELIKE 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. Because 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 [16] and examined by Miyamoto and Fitch [9], who compared Cu, Zn superoxide dismutase 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 also is motivated by Fitch's observation. For a certain class of events and parameters of a covarion model, we obtain a treelike distance measure between monophyletic groups of species that will not in general be treelike under rates-across-sites models. This shows, first, that infinite sequences can in fact distinguish between the two models and, second, that infinite sequences do contain information about the tree without requiring knowledge of the parameters of the model. Standard statistical techniques (such as maximum likelihood for tree reconstruction) may then be used to address these questions, given finite sequences.

The class of covarion models for which this is relevant includes those whose underlying observable process is based on the Kimura [17] three-substitution-type model (K3ST) or one of its submodels—the Kimura [18] two parameter (K2P) and Jukes-Cantor [19] (JC) models.

4.1. *TREELIKE DISTANCES*

Trees with positively weighted edges induce a natural distance function ρ_{ij} between leaves i and j of the tree—simply sum the edge weights along the path from leaf i to leaf j. Such a distance function has all the properties of a metric, as well as an additional property known as the *four-point condition.* That is, if the subtree induced by leaves i, j, k, and l is as shown in Figure 3, where τ_{xy} is the net sum of the edge weights along the path from vertex x to vertex y, then

$$\rho_{ij} + \rho_{kl} \leqslant \rho_{ik} + \rho_{jl} = \rho_{il} + \rho_{jk}. \tag{8}$$

The distance between u and v also may be recovered, because

$$\tau_{uv} = \frac{1}{2}(\rho_{ik} + \rho_{jl} - \rho_{ij} - \rho_{kl}).$$

Furthermore, the four-point condition [Eq. (8)] characterizes metrics that may be realized as edge-weighted trees: if the metric function ρ satisfies the four-point condition, then there is a unique tree T and a unique edge weighting w such that (T,w) realizes ρ [20]. Such metrics

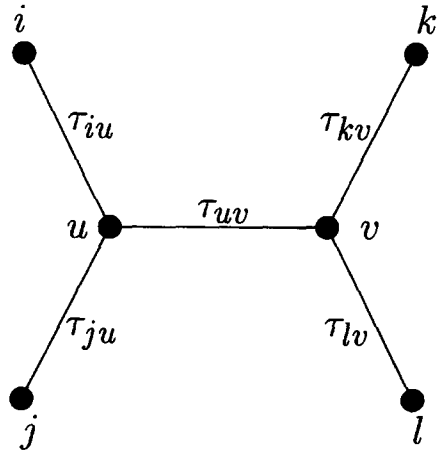

FIG. 3. The subtree induced by leaves i, j, k, and l.

are called *treelike* and are accordingly of interest in phylogeny, because they enable the tree to be recovered uniquely and quickly (by an algorithm whose running time grows polynomially in the number of taxa).

4.2. *SEPARABLE EVENTS*

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

Suppose E is an event involving an r-state site pattern 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 with corresponding rooted trees T_1 and T_2, 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 event E is *separable* under the covarion model (R,S) if

$$\mathsf{P}[E_1 \wedge E_2 \mid \mathrm{o}_1 = \mathrm{o}_1, \mathrm{o}_2 = \mathrm{o}_2] = \mathsf{P}[E_1 \mid \mathrm{o}_1 = \mathrm{o}_1]\mathsf{P}[E_2 \mid \mathrm{o}_2 = \mathrm{o}_2] \quad (9)$$

for all $\mathrm{o}_1, \mathrm{o}_2 \in \{\mathtt{on}, \mathtt{off}\}$. 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 $(Q, \mathscr{D})$ is the following *independence condition*:

$$\mathsf{P}[E_1 \wedge E_2 \mid \nu] = \mathsf{P}[E_1 \mid \nu]\mathsf{P}[E_2 \mid \nu]. \quad (10)$$

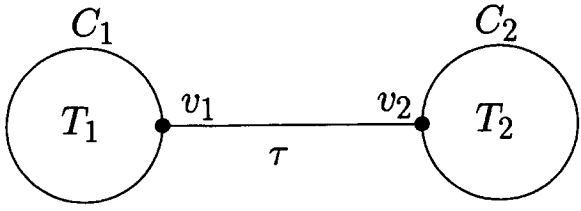

FIG. 4. The tree joining two monophyletic groups of species C_1 and C_2. The circles denote the rooted subtrees T_1 and T_2, the roots being v_1 and v_2, respectively. The edge $\{v_1, v_2\}$ has length τ.

Let

$$p_{12} = \mathsf{P}[E_1 \wedge E_2],$$
$$p_i = \mathsf{P}[E_i], \qquad i = 1,2$$

and, further, in the case of the covarion model, let $\mathtt{o}_i$ be the state "on" or "off" of the switch at the vertex v_i and

$$p_i^{\mathtt{on}} = \mathsf{P}[E_i \mid \mathtt{o}_i = \mathtt{on}],$$

$$p_i^{\mathtt{off}} = \mathsf{P}[E_i \mid \mathtt{o}_i = \mathtt{off}],$$

$$\delta_i = p_i^{\mathtt{on}} - p_i^{\mathtt{off}}$$

for $i = 1,2$. Then under conditions (9) and (10), we have the following lemma.

LEMMA 7

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

$$p_{12} - p_1 p_2 = \sigma_1 \sigma_2 e^{-(s_1+s_2)\tau} \delta_1 \delta_2. \tag{11}$$

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

THEOREM 8

For a tree with several monophyletic groups $C_1,\ldots,C_n$ ($|C_i| \geqslant 2$ for each i) at its tips, the measure

$$\rho_{ij} = -\ln|p_{ij} - p_i p_j| \qquad i \neq j$$

is treelike under a covarion model for which E is separable but, in general, is not under a rates-across-sites model for which the independence condition holds.

Proof of Lemma 7 and Theorem 8. In the covarion case,

$$p_{12} = \sum_{\mathtt{o}_1,\mathtt{o}_2} \mathsf{P}[E_1 \wedge E_2 \mid \mathtt{o}_1 = \mathtt{o}_1, \mathtt{o}_2 = \mathtt{o}_2]\mathsf{P}[\mathtt{o}_1 = \mathtt{o}_1, \mathtt{o}_2 = \mathtt{o}_2]$$
$$= \sum_{\mathtt{o}_1,\mathtt{o}_2} \mathsf{P}[E_1 \mid \mathtt{o}_1 = \mathtt{o}_1]\mathsf{P}[E_2 \mid \mathtt{o}_2 = \mathtt{o}_2]\mathsf{P}[\mathtt{o}_1 = \mathtt{o}_1, \mathtt{o}_2 = \mathtt{o}_2]$$

because E is separable, and

$$p_1 p_2 = \sum_{o_1, o_2} \mathsf{P}[E_1 \mid \mathrm{O}_1 = o_1] \times \mathsf{P}[E_2 \mid \mathrm{O}_2 = o_2]\mathsf{P}[\mathrm{O}_1 = o_1]\mathsf{P}[\mathrm{O}_2 = o_2].$$

Thus

$$p_{12} - p_1 p_2 = \sum_{o_1, o_2} \mathsf{P}[E_1 \mid \mathrm{O}_1 = o_1] \times \mathsf{P}[E_2 \mid \mathrm{O}_2 = o_2](\mathsf{P}[\mathrm{O}_1 = o_1, \mathrm{O}_2 = o_2] - \mathsf{P}[\mathrm{O}_1 = o_1]\mathsf{P}[\mathrm{O}_2 = o_2]) \tag{12}$$

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

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

see, for example, [11] page 156; so the matrix of $\mathsf{P}[\mathrm{O}_1 = o_1, \mathrm{O}_2 = o_2] - \mathsf{P}[\mathrm{O}_1 = o_1]\mathsf{P}[\mathrm{O}_2 = o_2]$ is

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

Hence, from Equation (12),

$$\begin{aligned} p_{12} - p_1 p_2 &= \sigma_1 \sigma_2 e^{-(s_1+s_2)\tau} \left(p_1^{\mathrm{on}} - p_1^{\mathrm{off}} \right) \left(p_2^{\mathrm{on}} - p_2^{\mathrm{off}} \right) \\ &= \sigma_1 \sigma_2 e^{-(s_1+s_2)\tau} \delta_1 \delta_2, \end{aligned}$$

as claimed.

Under rates across sites with the independence condition holding,

$$\begin{aligned} \mathsf{P}[E_1 \wedge E_2] &= \int_0^\infty \mathsf{P}[E_1 \wedge E_2 \mid \nu] \, dF_{\mathscr{D}}(\nu) \\ &= \int_0^\infty \mathsf{P}[E_1 \mid \nu]\mathsf{P}[E_2 \mid \nu] \, dF_{\mathscr{D}}(\nu), \end{aligned}$$

which does not depend on τ, and, similarly,

$$\mathsf{P}[E_i] = \int_0^\infty \mathsf{P}[E_i \mid \nu] \, dF_{\mathscr{D}}(\nu)$$

does not depend on τ, so $p_{12} - p_1 p_2 = \mathsf{P}[E_1 \wedge E_2] - \mathsf{P}[E_1]\mathsf{P}[E_2]$ does not depend on τ either.

Because 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

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

Referring to Figure 5, we have $\tau_{ij} = \tau_i + \tau_j$, $\tau_{ik} = \tau_i + \tau_m + \tau_k$, etc., and Theorem 8 follows.

■

Note that the set of equations

$$\rho_{ij} = -\ln(\sigma_1\sigma_2) + (s_1+s_2)\tau_{ij} - \ln|\delta_i| - \ln|\delta_j|,$$

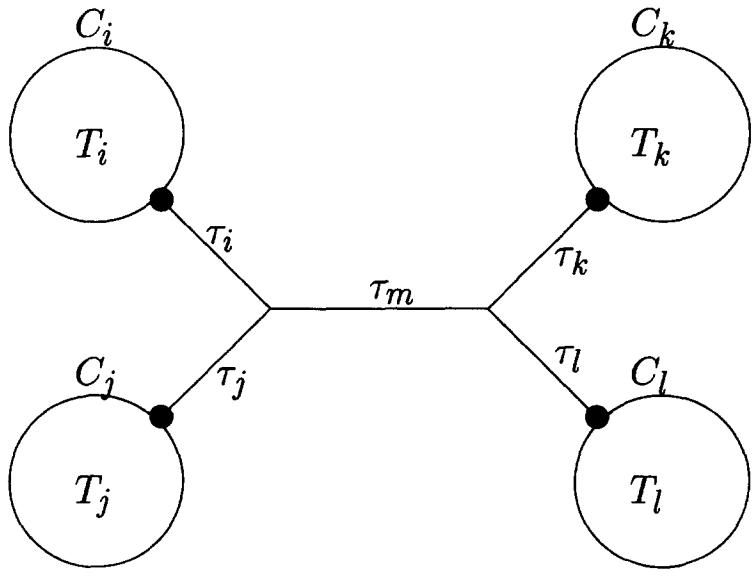

FIG. 5. The tree on the four monophyletic groups of species C_i, C_j, C_k, and C_l. The τ_x are the edge lengths.

where $1 \leqslant i < j \leqslant 4$ is a system of six linear equations in the six unknowns $\ln(\sigma_1\sigma_2)$, $(s_1+s_2)\tau_k - \ln|\delta_k|$, $1 \leqslant k \leqslant 4$ and $(s_1+s_2)\tau_5$. However, this system is singular (because $\rho_{13}+\rho_{24} = \rho_{14}+\rho_{23}$) and only $(s_1+s_2)\tau_5$ may be solved for uniquely if only the ρ_{ij} are known.

Further, although ρ_{ij} is not in general additive under a rates-across-sites model, the four-point condition may still hold, albeit in the form

$$\rho_{ij}+\rho_{kl} = \rho_{ik}+\rho_{jl} = \rho_{il}+\rho_{jk}. \tag{13}$$

For example, if trees T_i, T_j, T_k, and T_l are exactly the same, this will certainly be the case. Less restrictively, if the T_x are all two-taxa trees with their leaves separated by a distance τ_x and we assume the fully symmetric model ($Q_{ij} = \alpha$ if $i \neq j$ and $Q_{ii} = (1-r)\alpha$), then the ρ_{xy} may be calculated relatively easily and it appears that Equation (13) holds for any choice of τ_i, τ_j, τ_k, and τ_l if and only if $\mathscr{D}$ is a discrete one- or two-rate distribution.

4.3. *EXAMPLES OF SEPARABLE EVENTS*

We begin by giving a sufficient condition for separability, which will allow us to show that, under a model that regards the states somewhat interchangeably, any event that respects that interchangeability will be separable. We will then be able to find some examples of separable events.

Let A_i be the state of the observable process at vertex v_i.

LEMMA 9

(i) *Under the covarion model (R,S), if E_i is independent of A_i for* $\mathrm{O}_i = \mathtt{on}$ *and* $\mathrm{O}_i = \mathtt{off}$, *then E is separable.*

(ii) *Under the rates-across-sites model $(Q,\mathscr{D})$, if E_i is conditionally independent of A_i given ν, then the independence condition holds.*

Proof. The proofs of parts (i) and (ii) are entirely similar, so we prove only (i). For any reversible Markov tree model, we have

$$\mathsf{P}[E_1 \wedge E_2 \mid \mathrm{O}_1 \wedge \mathrm{O}_2, A_1 \wedge A_2] = \mathsf{P}[E_1 \mid \mathrm{O}_1 \wedge A_1]\mathsf{P}[E_2 \mid \mathrm{O}_2 \wedge A_2]. \tag{14}$$

Let $p_i(a_i) = \mathsf{P}[E_i \mid \mathrm{O}_i \wedge (A_i = a_i)]$. Then, from Equation (14),

$$\mathsf{P}[E_1 \wedge E_2 \mid \mathrm{O}_1 \wedge \mathrm{O}_2] = \sum_{a_1, a_2 \in \mathscr{A}} p_1(a_1)p_2(a_2)\mathsf{P}[(A_1 = a_1) \wedge (A_2 = a_2) \mid \mathrm{O}_1 \wedge \mathrm{O}_2].$$

Also,

$$\mathsf{P}[E_i \mid \mathsf{O}_i] = \sum_{a_i \in \mathscr{A}} p_i(a_i)\mathsf{P}[A_i = a_i \mid \mathsf{O}_i],$$

so

$$\begin{aligned}\Delta &= \mathsf{P}[E_1 \wedge E_2 \mid \mathsf{O}_1 \wedge \mathsf{O}_2] - \mathsf{P}[E_1 \mid \mathsf{O}_1]\mathsf{P}[E_2 \mid \mathsf{O}_2] \\ &= \sum_{a_1, a_2 \in \mathscr{A}} p_1(a_1)p_2(a_2)\big(\mathsf{P}[(A_1 = a_1) \wedge (A_2 = a_2) \mid \mathsf{O}_1 \wedge \mathsf{O}_2] \\ &\qquad - \mathsf{P}[A_1 = a_1 \mid \mathsf{O}_1]\mathsf{P}[A_2 = a_2 \mid \mathsf{O}_2]\big).\end{aligned}$$

Now, if E_i is independent of A_i for $\mathsf{O}_i = \text{on}$ and for $\mathsf{O}_i = \text{off}$, then we may write $p_i(a_i) = p_i$, and so

$$\begin{aligned}\Delta &= p_1 p_2\Big(\sum_{a_1, a_2 \in \mathscr{A}} \mathsf{P}[(A_1 = a_1) \wedge (A_2 = a_2) \mid \mathsf{O}_1 \wedge \mathsf{O}_2] \\ &\qquad - \sum_{a_1, a_2 \in \mathscr{A}} \mathsf{P}[A_1 = a_1 \mid \mathsf{O}_1]\mathsf{P}[A_2 = a_2 \mid \mathsf{O}_2]\Big) \\ &= p_1 p_2(1-1) = 0.\end{aligned}$$

Hence E is separable.

■

Given a permutation $\sigma \in S_r$ (the symmetric group on r objects), the permutation matrix P_σ corresponding to σ is the matrix whose $i\sigma(i)$-entry is 1, with all other entries being zero. If R is an $r \times r$ matrix, then $P_\sigma R$ is the matrix that results if the rows of R are swapped according to σ, whereas RP_σ^{T} is the matrix that results if the columns of R are swapped according to σ. Consequently $P_\sigma RP_\sigma^{\mathrm{T}}$ is the result of swapping both the rows and columns.

The map $R \mapsto \sigma R = P_\sigma RP_\sigma^{\mathrm{T}}$ defines a group action on the set of all $r \times r$ rate matrices. If $R = P_\sigma RP_\sigma^{\mathrm{T}}$, we say that R is *invariant* under σ; further, if G is a subgroup of S_r and R is invariant under σ for all $\sigma \in G$, we say that R is invariant under the action of G. We note that, because $P_\sigma^{\mathrm{T}} = P_{\sigma^{-1}} = P_\sigma^{-1}$, the set of matrices invariant under the action of G is closed under multiplication.

As an example, consider the matrices

$$R_K = \begin{pmatrix} -\delta & \alpha & \beta & \gamma \\ \alpha & -\delta & \gamma & \beta \\ \beta & \gamma & -\delta & \alpha \\ \gamma & \beta & \alpha & -\delta \end{pmatrix} \quad \text{and} \quad R_C = \begin{pmatrix} -\delta & \alpha & \beta & \gamma \\ \gamma & -\delta & \alpha & \beta \\ \beta & \gamma & -\delta & \alpha \\ \alpha & \beta & \gamma & -\delta \end{pmatrix} \tag{15}$$

where $\delta = \alpha + \beta + \gamma$. It is easily checked that R_K (which is the matrix used in the K3ST model) is invariant under the action of

$$K_4 = \{\mathrm{id}, (1\ 2)(3\ 4), (1\ 3)(2\ 4), (1\ 4)(2\ 3)\},$$

which is isomorphic to the Klein 4-group, whereas R_C is invariant under the action of

$$C_4 = \{\mathrm{id}, (1\ 2\ 3\ 4), (1\ 3)(2\ 4), (1\ 4\ 3\ 2)\},$$

which is isomorphic to the cyclic group Z_4.

In a similar way, we may define a group action on site patterns by $\chi \mapsto \sigma\chi$, where $\sigma\chi(i) = \sigma(\chi(i))$. Because events concerning states of the taxa are sets of site patterns, this extends to an action on such events by

$$\sigma E = \{\sigma\chi \mid \chi \in E\}.$$

Again, if $\sigma E = E$ for all $\sigma \in G$, we say that E is invariant under the action of G. As an example, if C is a set of species, then the event E^s that a given site takes the same state at each species is invariant under the action of any subgroup of S_r. For a less-trivial example, consider the events on two species with four states (1, 2, 3, and 4) given by

$$E^2 = \text{“the states differ by 2” (e.g., 1 and 3 or 4 and 2)}$$

and

$$E^{1,3} = \text{“the states differ by 1 or 3” (e.g., 1 and 2 or 1 and 4)}.$$

Again, it is easily checked that E^2 and $E^{1,3}$ are invariant under the action of both K_4 and C_4. Note that, if the states are the nucleotides A, C, G, and T in that order, then E^2 is the event “the states differ by a transition,” whereas $E^{1,3}$ is the event “the states differ by a transversion.”

The usefulness of these concepts in the present context is given by the following theorem. Invariance of the rate matrix under the action of a group G breaks the state space up into classes of states that “look the same.” If there is just one class of states that “look the same” (i.e., if there is just one orbit under the action of G; such an action is called transitive), then any event invariant under G will be separable.

THEOREM 10

Let R be a stationary and time-reversible $r \times r$ rate matrix, and let E be an event involving r-state site patterns on monophyletic sets of species. If

both R and E are invariant under the action of some $G \subseteq S_r$ that acts transitively on $[r]$, then

(i) *E is separable under the covarion model (R, S) for any switch matrix S.*

(ii) *The independence condition holds for E under the rates-across-sites model $(R, \mathscr{D})$ for any distribution $\mathscr{D}$.*

Proof. The result follows from a simple symmetry argument. For part (ii), the transition matrices $P^e = \exp(\tau_e R)$ inherit invariance under the action of G from R, so

$$P^e_{\sigma(\alpha)\sigma(\beta)} = P^e_{\alpha\beta}$$

for all $\alpha, \beta \in [r]$ and all edges e of T_i. It follows that, on renaming all states according to $\sigma \in G$, we have $\mathsf{P}[E_i \mid A_i = a_i] = \mathsf{P}[E_i \mid A_i = \sigma(a_i)]$. Because G acts transitively on $[r]$, (ii) now follows from Lemma 9 (ii).

For part (i), we use the second formulation of the covarion model and argue similarly to part (ii). We again have

$$P^e_{(\sigma(\alpha),\mathrm{o})(\sigma(\beta),\mathrm{o}')} = P^e_{(\alpha,\mathrm{o})(\beta,\mathrm{o}')}$$

for all $\alpha, \beta \in [r]$, $\mathrm{o}, \mathrm{o}' \in \{\mathtt{on}, \mathtt{off}\}$ and all edges e of T_i, so, on renaming all states according to σ, we get

$$\mathsf{P}[E_i \mid (A_i, \mathrm{O}_i) = (a_i, \mathrm{o}_i)] = \mathsf{P}[E_i \mid (A_i, \mathrm{O}_i) = (\sigma(a_i), \mathrm{o}_i)].$$

Because G acts transitively on $[r]$, E_i is independent of A_i for $\mathrm{O}_i = \mathtt{on}$ and $\mathrm{O}_i = \mathtt{off}$, so E is separable by Lemma 9 (i). ∎

Theorem 10 will allow the construction of many examples of separable events, and we state some examples as a corollary below. Part (ii) is of interest in its own right. Fu and Li [21], in constructing certain quadratic invariants, showed that the heretofore defined events E^s, E^2, and $E^{1,3}$ satisfy the independence condition on four taxa trees if all transition matrices have the form R_K, without placing any conditions on the location of the root or the distribution of states there. The proof of part (ii) of Theorem 10 requires only that all transition matrices be invariant under the action of G (they need not be generated by a single continuous-time Markov process); no requirements are placed on the distribution of states at the root, the number of taxa, or the number of states. This extends Fu and Li's result, fitting it into a much broader framework.

We will say that R is *permutable* if it is invariant under the action of some $G \subseteq S_r$ that acts transitively on $[r]$.

COROLLARY 11 (Some examples of separable events)

1. *The events E^s and E^d referred to earlier are separable under the covarion model (R, S) and satisfy the independence condition under the rates-across-sites model $(R, \mathscr{D})$ whenever R is permutable. In particular, R is permutable if it has one of the following forms*:

(i) $R = R_K$, *where* R_K *is as given in Equation* (15) *and is the form of the matrix used in the K3ST model. This includes as special cases the K2P model* ($\beta = \gamma$) *and the JC model* ($\alpha = \beta = \gamma$).

(ii) *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 JC model* ($r = 4$).

2. *The events on pairs of species E^2 (differ by a transition) and $E^{1,3}$ (differ by a transversion) are separable under the covarion model (R, S) and satisfy the independence condition under the rates-across-sites model $(R, \mathscr{D})$ whenever R is of the form R_K.*

(Note that the matrix R_C is not time reversible unless $\alpha = \gamma$, in which case it is of the form R_K).

5. CONCLUSION

We have presented and analyzed a simple covarion-style model, comparing it with the better known rates-across-sites models.

We have shown that, even for infinitely long sequences, a covarion model will give results identical with those of a suitably chosen rates-across-sites model when making simultaneous comparisons of pair-by-pair dissimilarities between a collection of sequences. Consequently, if one wishes to test between these two models by using real (finite length) sequences, it is necessary to consider further properties of the data than just pair-by-pair dissimilarities. We also showed how the expected pair-by-pair dissimilarities could be transformed so as to estimate the evolutionary distance between the two sequences; however, this required knowledge of the underlying rate matrices R and S.

In Section 4, by following an observation of Fitch [16], we showed how certain versions of the covarion model could be used to construct a treelike distance on monophyletic groups of species—again for infinitely long sequences but this time without using knowledge of the underlying rate matrices R and S. The significance of this result for real sequences is twofold. First, it shows that treelike information can be recovered from sufficiently long sequences under the covarion-style model, given knowledge of monophyletic groupings. The particular treelike distance described could be used directly on real sequences, provided they are reasonably long, in much the same way as similar

logarithmic transformations are routinely used in phylogenetics. Alternatively, more powerful (but also more computationally intensive) statistical techniques such as maximum likelihood could be employed—our result simply shows that treelike information is there in the sequences to be recovered.

Second, because the treelike measure is not, in general, treelike under the rates-across-sites model, this shows that the two models can indeed be distinguished, given sufficiently long sequences. A useful project for future work would be the development of such tests. A test that did not depend on restrictions to the model such as the separability condition of Equation (9) would be particularly desirable.

We thank the New Zealand Marsden Fund (Contract No. UOC 516) for funding this research. We also thank Dr. David Penny, Dr. Walter Fitch, and two anonymous referees for their helpful comments.

REFERENCES

1 J. Felsenstein, Cases in which parsimony or compatibility will be positively misleading. *Syst. Zool.* 27:401–410 (1978).
2 C. Kelly and J. Rice, Modeling nucleotide evolution: a heterogeneous rate analysis. *Math. Biosci.* 133:85–109 (1996).
3 J. T. Chang, Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* 134:189–215 (1996).
4 M. A. Steel, L. A. Székely, and M. D. Hendy, Reconstructing trees when sequence sites evolve at variable rates. *J. Comput. Biol.* 1(2):153–163 (1994).
5 Z. Yang, Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ across sites. *Mol. Biol. Evol.* 10:1396–1401 (1993).
6 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. *Biochem. Genet.* 4:579–593 (1970).
7 W. M. Fitch, Rate of change of concomitantly variable codons. *J. Mol. Evol.* 1:84–96 (1971).
8 W. M. Fitch and F. J. Ayala, The superoxide dismutase molecular clock revisited. *Proc. Natl. Acad. Sci. USA* 91:6802–6807 (1994).
9 M. M. Miyamoto and W. M. Fitch, Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.* 12(3):503–513 (1995).
10 A. Rényi, *Probability Theory*. Vol. 10 of *North-Holland Series in Applied Mathematics and Mechanics*, North-Holland, Amsterdam, 1970. English translation by L. Vekerdi.
11 G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*. Clarendon Press, Oxford, 1982.
12 J. Keilson, *Markov Chain Models: Rarity and Exponentiality*. Vol. 28 of *Applied Mathematical Sciences*, Springer-Verlag, 1979.
13 J. N. Darroch and K. W. Morris, Passage time generating functions for continuous-time finite Markov chains. *J. Appl. Probability* 5:414–426 (1968).

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

15 P. J. Waddell and M. A. Steel, General time reversible distances with unequal rates across sites. Research Report No. 143, Dept. of Mathematics and Statistics, Univ. of Canterbury, New Zealand (1996).

16 W. M. Fitch, The nonidentity of invariable positions in the cytochrome *c* of different species. *Biochem. Genet.* 5:231–241 (1971).

17 M. Kimura, Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Nat. Acad. Sci. USA* 78:454–458 (1981).

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

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

20 P. Buneman, The recovery of trees from measures of dissimilarity. In *Mathematics in the Archaeological and Historical Sciences*, F. R. Hodson, D. G. Kendall, and P. Tautu, eds., Edinburgh Univ. Press, 1971, pp. 387–395.

21 Y.-X. Fu and W.-H. Li, Necessary and sufficient conditions for the existence of certain quadratic invariants under a phylogenetic tree. *Math. Biosci.* 105:229–238 (1991).