The algebra of the general Markov model on phylogenetic trees and networks

J. G. Sumner, B. H. Holland[*], and P. D. Jarvis[†]

School of Mathematics and Physics, University of Tasmania, Australia

Abstract

It is known that the Kimura 3ST model of sequence evolution on phylogenetic trees can be extended quite naturally to arbitrary split systems. However, this extension relies heavily on mathematical peculiarities of the K3ST model, and providing an analogous augmentation of the general Markov model has thus far been elusive. In this paper we rectify this shortcoming by showing how to extend the general Markov model on trees to to include arbitrary splits; and even further to more general network models. This is achieved by exploring the algebra of the generators of the continuous-time Markov chain together with the "splitting" operator that generates the branching process on phylogenetic trees. For simplicity we proceed by discussing the two state case and note that our results are easily extended to more states with little complication. Intriguingly, upon restriction of the two state general Markov model to the parameter space of the binary symmetric model, our extension is indistinguishable from the previous approach only on trees; as soon as any incompatible splits are introduced the two approaches give rise to differing probability distributions with disparate structure. Through exploration of a simple example, we give a tentative argument that our approach to extending to more general networks has desirable properties that the previous approaches do not share. In particular, our construction allows for the possibility of convergent evolution of previously divergent lineages; a property that is of significant interest for biological applications.

1 Introduction

Phylogenetic methods seek to infer the prior evolutionary relationships of extant taxa. Classically, this was achieved by comparing morphological features, but modern methods focus on molecular data such as DNA. Harking back to sketches in Darwin's early notebooks, it has also been assumed that evolutionary history resembles a tree structure. However, it is now well known that evolutionary processes such as hybridisation, deep coalescence (incomplete lineage-sorting), horizontal gene transfer and recombination cannot be accurately modelled as a tree. Even when the underlying historical signal fits a tree, there may be conflicting non-historical signals caused by sampling error, long-branch attraction, nucleotide composition bias, or changes in the substitution rate at individual sites across the tree, as well as alignment or misreading errors. Incorrect or over-parameterized models of sequence mutations can lead to high statistical support for splits that are incompatible with a single tree.

It is clear that imposing a strictly treelike evolutionary history may be inappropriate in the situations described above, hence methods that can assist in identifying and understanding conflict in phylogenetic data are essential. One class of methods that have proved useful in this respect are weighted split-systems and their corresponding visualization as networks. Split networks initially arose as means of visualizing the split decomposition of a distance metric as defined by Bandelt & Dress (1992). In that work, the authors gave a decomposition that provides a way of assessing whether the structure of a distance matrix is treelike or if it contains other conflicting signals. However, the idea of a weighted split-system is very general and has arisen in many phylogenetic contexts. These include (1) median networks (Bandelt, 1994), where splits and their weights are derived from a binary coding of a sequence alignment; (2) Hadamard (or spectral) analysis (Hendy & Penny, 1989), which defines an invertible relationship between site patterns and a split spectrum under certain simple models (K3ST and subclasses); (3) Neighbor-Net (Bryant & Moulton, 2004), a distance-based method which applies a greedy agglomerative algorithm to find a circular ordering of taxa and then a least-squares approach to find weights for the corresponding set of circular splits; (4) Consensus networks (Holland & Moulton, 2004), which take a set of trees and define a weighted split-system based on the number of trees that display a particular edge, possibly incorporating edge weight information (Holland *et al.*, 2006).

With the exception of spectral analysis Hendy & Penny (1989) these methods are all combinatorial and/or distance-based; there is currently no way to infer a weighted split-system in the likelihood setting using general Markov models of sequence evolution. That said, there has been some previous work on calculating likelihood scores for particular phylogenetic networks under special models. von Haeseler & Churchill (1993) developed a framework for computing the likelihood of a split-system for binary sequences under a Cavender-Farris model with invariable sites. Strimmer & Moulton (2000) developed a Bayesian approach that calculated the likelihood of a given directed acyclic graph (DAG) for more complex models of sequence evolution. More recently Jin *et al.* (2006) defined a likelihood score for phylogenetic networks as a weighted mixture of tree likelihoods. All of these methods begin with a given phylogenetic network and then attempt to calculate its likelihood under some model. This is very different from the Hadamard based approach – and the new approach we explore here – which begin with the data and a model and infer a weighted split-system.

Given their importance, there is a distinct lack of an extension of the standard Markov models on trees to arbitrary split systems. In recent work, Bryant (2005a, 2009) has re-examined the nature of the Kimura 3ST and binary-symmetric model as, under a simple extension, these models permit the inclusion of arbitrary splits over and above those that come from a single tree. However, these so called "group-based" models of sequence evolution are not motivated by biological considerations and hence their validity in applied studies must be scrutinized carefully. The primary motivation for these models is mathematical elegance and simplicity, so that it is not necessarily the case that the underlying assumptions have biological relevance. Specifically, all group-based models have doubly-stochastic rate matrices and thus uniform stationary distributions. This is clearly inappropriate given that varying GC content is known to be of crucial importance in phylogenetics (Jermiin *et al.*, 2004). It appears that to date it has not been possible
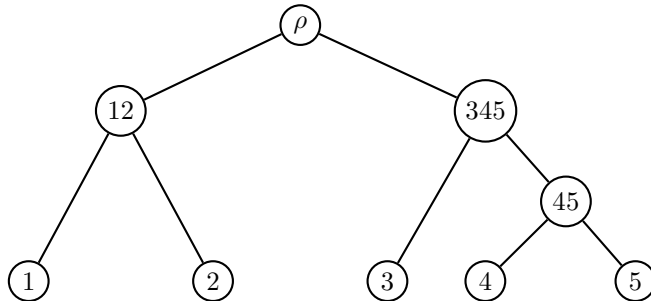
Figure 1: A rooted tree.

to employ general model-based methods to infer split networks from phylogenetic data sets.

In this article we show how Markov models of phylogenetic evolution on trees (thought of as *compatible* split systems) can be generalized to the case of *arbitrary* split systems. In the binary case of two character states, we achieve this by studying the algebra of the generators of the continuous time Markov chain together with the "splitting" operator that generates the branching process on phylogenetic trees. The resulting presentation of the general Markov rate matrix model on a tree is such that it can be generalized in a natural way to include arbitrary splits; including those that are incompatible with any tree. This results in a very general model that contains the standard tree model as a special case, but has the potential to associate an individual weight and rate matrix to *any* additional splits that we wish to include. Additionally, we show by example that our approach gives rise to the possibility of Markov models on much more general networks, with phylogenetic evolution proceeding in a series of "epochs" consisting of divergence or *convergence* of arbitrary groups of taxa (ie. lineages). As part of the discussion, we note that our results are fully generalizable to any number of character states with complication of detail only. Intriguingly, we will also show that under a restriction of the parameter space to the binary-symmetric case that this model is *not* consistent with the Hadamard based approach given by Bryant (2009). We close with a simple example that shows our construction has the ability to model convergent evolution of lineages; a property that is simply not available to the Hadamard based approach.

2   Preliminaries

In this article, a *tree* with vertex set X is an acyclic graph with vertices chosen from X such that all vertices have valence of exactly 3 or 1. A *rooted tree* is an acyclic graph as above but with a single vertex ρ (the root) having valence 2. Thus, a rooted tree is a collection of *edges* $e \in \binom{X}{2}$, and can be made to be a *directed* graph by considering $e = (u, v)$ as an ordered pair of adjacent vertices, where u lies on the path from v to ρ. Vertices of valence 1 are referred to as *leaves* and we label the leaves from elements of the set $[n] := \{1, 2, \ldots, n\}$. For a rooted tree, we label each non-leaf vertex v by the subset of leaves such that the path from each of these leaves to ρ contains v. Additionally, we label each edge $e = (u, v)$ by the subset that labels the vertex v. In this way, the edges are labelled by subsets of $[n]$, with pendant edges labelled by singletons. See Figure 1 for a graphical representation of a rooted tree.

Consider the vector space $V \cong \mathbb{C}^2$ with the ordered basis[1]

$$\left\{ |0\rangle \equiv e_0 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle \equiv e_1 := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

With respect to this basis, we can define "Markov generators" as the zero column-sum matrices

$$L_\alpha = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad L_\beta = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

---

[1]We use "Dirac notation", where a vector is represented by a "ket" $|\rangle$, as this notation is particularly elegant when it comes to more general phylogenetic character patterns.

In this way, the general rate matrix for a continuous-time Markov chain on two states can be expressed as the linear combination[2]

$$Q = \alpha L_\alpha + \beta L_\beta = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}. \tag{1}$$

The associated transition matrix $[M(t)]_{ij}$, representing the probability of a transistion $j \to i$ at time t, is then given by the exponential map

$$M(t) = \exp[Qt] := \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!}.$$

By noting that the generators satisfy the relations

$$L_\alpha^2 = L_\beta L_\alpha = -L_\alpha, \qquad L_\beta^2 = L_\alpha L_\beta = -L_\beta, \tag{2}$$

it is easy to show that

$$(\alpha L_\alpha + \beta L_\beta)^n = (-1)^{n-1}(\alpha + \beta)^{n-1} (\alpha L_\alpha + \beta L_\beta).$$

Thus

$$\begin{aligned} M(t) = \exp[Qt] = \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!} &= 1 - \frac{1}{\alpha + \beta} \left(e^{-(\alpha+\beta)t} - 1 \right) (\alpha L_\alpha + \beta L_\beta) \\ &= 1 - \frac{1}{\alpha + \beta} \left(e^{-(\alpha+\beta)t} - 1 \right) Q. \end{aligned} \tag{3}$$

As $M(t)$ is invariant under the reparameterization[3]

$$t \to t' = \lambda t, \quad \alpha \to \alpha' = \lambda^{-1}\alpha, \quad \beta \to \beta' = \lambda^{-1}\beta,$$

we see that we can "scale out" t by choosing $\lambda = t^{-1}$. As, in a practical context, α, β and even t are unknown parameters that must be inferred from observed data using some statistical estimation procedure, we see that we can take

$$M(\alpha, \beta) = e^Q = e^{(\alpha L_\alpha + \beta L_\beta)},$$

as completely equivalent to (3). If we think of $M(\alpha, \beta)$ as a two-dimensional manifold (in the sense of a Lie group (Procesi, 2007)), then we see that the Markov generators are none other than the basis vectors of the tangent space at the identity:

$$L_\alpha \equiv \left. \frac{\partial}{\partial \alpha} M(\alpha, \beta) \right|_{\alpha=\beta=0}, \qquad L_\beta \equiv \left. \frac{\partial}{\partial \beta} M(\alpha, \beta) \right|_{\alpha=\beta=0},$$

with algebraic closure across the "Lie bracket" $[L_\alpha, L_\beta] := L_\alpha L_\beta - L_\beta L_\alpha = L_\alpha - L_\beta$ ensuring "closure" of the corresponding Markov model (as is discussed in Jarvis & Sumner (2010)). This connection between continuous time Markov chains and Lie groups is an important one and seems to have been first noted by Johnson (1985). This point of view is needed in order to extend the results of the present article to the case of character state spaces of arbitrary size. Having given this perspective into the meaning of the Markov generators, we from will nevertheless take the more usual representation (3) of transition matrices in all that follows below.

---

[2] An amusing aside: Our rate matrices have zero *column-* rather than *row-*sum, as we like to conform with physicists' notation of right matrix multiplication. This sits well psychologically with the physical picture that a linear operator "hits" a vector and it "moves", which, in turn, is in tune with the left-to-right direction that one reads printed English.

[3] For further discussion of *local* time-reparameterization in phylogenetics see Jarvis *et al.* (2005) or, in the context of a changing rate of mutation, see Penny (2005)

In Sumner & Jarvis (2005) it was shown that the Markov models of phylogenetics in standard use can be represented in an abstract setting using the tensor product space $V \otimes V \otimes \ldots \otimes V$, where $\dim(V) = k$ is the number of character states and the number of copies of V is equal to the number of taxa under consideration. In these models, it is usual to impose conditional independence across the branches of the tree and this can be formalized using a linear operator $\delta : V \to V \otimes V$ to generate speciation events. This is referred to as the "splitting operator" and is defined, using our chosen basis, as

$$\delta \cdot |i\rangle = |i\rangle \otimes |i\rangle,$$

which, expressed at the level of the probability distributions, corresponds exactly to the duplication of a sequence of molecular units. That is, if we select state i from the initial sequence with probability p_i, then immediately after duplication – and assuming the sequences remain aligned – the probability of observing the pattern ij at a given site is $p_i \delta_{ij}$. By defining the vector $p := \sum_i p_i |i\rangle \in V$ and noting that the tensors $\{|ij\rangle := |i\rangle \otimes |j\rangle\}_{0 \le i,j \le 1}$ form a basis for $V \otimes V$, the splitting operator achieves this notion of speciation in the abstract setting:

$$\delta \cdot p = \delta \cdot \left(\sum_i p_i |i\rangle \right) = \sum_i p_i \delta \cdot |i\rangle = \sum_i p_i |ii\rangle = \sum_{i,j} p_i \delta_{ij} |ij\rangle,$$

where $P := \delta \cdot p \in V \otimes V$ has components $p_i \delta_{ij}$ and is referred to as a "phylogenetic tensor". Subsequently, given two rate matrices Q_1 and Q_2, and two edge weights τ_1 and τ_2, the phylogenetic tensor evolves to

$$P' = e^{Q_1 \tau_1} \otimes e^{Q_2 \tau_2} \cdot P,$$

which, up to first order terms in the edge weights, is

$$P' = [(1 + \alpha_1 \tau_1 L_\alpha + \beta_1 \tau_1 L_\beta + \ldots) \otimes (1 + \alpha_2 \tau_2 L_\alpha + \beta_2 \tau_2 L_\beta + \ldots)] \cdot P$$
$$= [1 + \alpha_1 \tau_1 L_\alpha \otimes 1 + \alpha_2 \tau_2 1 \otimes L_\alpha + \beta_1 \tau_1 L_\beta \otimes 1 + \beta_2 \tau_2 1 \otimes L_\beta + \ldots] \cdot P.$$

In this way, the splitting operator can be thought of as the generator of the branching pattern of the phylogenetic tree, while L_α and L_β are the generators of the Markov process. (For more details of this formalism see Bashford $et~al.$ (2004); Sumner & Jarvis (2005) and Sumner $et~al.$ (2008), and for an even more general setting see Jarvis $et~al.$ (2005)). Presently, we are concerned with the algebra resulting from application of these two types of generators.

3 Some helpful lemmas

The action of the Markov generators on the basis vectors is

$$
\begin{array}{llr}
L_\alpha |0\rangle = |1\rangle - |0\rangle, & L_\beta |0\rangle = 0, & \\
L_\alpha |1\rangle = 0; & L_\beta |1\rangle = |0\rangle - |1\rangle. & (4)
\end{array}
$$

By comparing the relations

$$\delta \cdot L_\alpha |0\rangle = |11\rangle - |00\rangle, \qquad \delta \cdot L_\alpha |1\rangle = 0;$$

and

$$
\begin{aligned}
L_\alpha \otimes L_\alpha |00\rangle &= |11\rangle - |01\rangle - |10\rangle + |00\rangle, \\
L_\alpha \otimes L_\alpha |11\rangle &= 0, \\
L_\alpha \otimes 1 |00\rangle &= |10\rangle - |00\rangle, \\
L_\alpha \otimes 1 |11\rangle &= 0, \\
1 \otimes L_\alpha |00\rangle &= |01\rangle - |00\rangle, \\
1 \otimes L_\alpha |11\rangle &= 0;
\end{aligned}
$$

with similar for L_β, we find that the Markov generators can be "pushed through" past the splitting operator:

4

Lemma 3.1. *As operators from V to $V \otimes V$, we have*

$$\delta \cdot L_\alpha = (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) \cdot \delta,$$
$$\delta \cdot L_\beta = (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta) \cdot \delta.$$

In the terminology of group actions (Procesi, 2007), this lemma tells us the rule for how the two operators "intertwine". It is exactly this relation that we will exploit to show how to generalize from a Markov model on a tree to a model on a general split system.

Given a linear operator X on V, we define a linear operator $X^{(i)}$ on $V \otimes V \otimes \ldots \otimes V$ as the tensor product

$$X^{(i)} := 1 \otimes 1 \otimes \ldots \otimes X \otimes 1 \otimes \ldots \otimes 1,$$

where X appears in the ith slot of the tensor product. Further, for a subset $A \subseteq [n] := \{1, 2, \ldots, n\}$, we define

$$X^{(A)} := \prod_{i \in A} X^{(i)}.$$

For example, if we take $n = 5$, we have

$$X^{(25)} = (1 \otimes X \otimes 1 \otimes 1 \otimes 1) \cdot (1 \otimes 1 \otimes 1 \otimes 1 \otimes X) = 1 \otimes X \otimes 1 \otimes 1 \otimes X.$$

Presently we will show how the interaction of δ with L_α naturally produces terms such as $L_\alpha^{(A)}$ (and similar for L_β).

Lemma 3.2. *As linear operators from V to $V \otimes V \otimes V$,*

$$1 \otimes \delta \cdot \delta = \delta \otimes 1 \cdot \delta.$$

Proof. We have

$$1 \otimes \delta \cdot \delta \cdot |i\rangle = 1 \otimes \delta \cdot (|i\rangle \otimes |i\rangle) = |i\rangle \otimes (|i\rangle \otimes |i\rangle) = |i\rangle \otimes |i\rangle \otimes |i\rangle := |iii\rangle \,.$$

Similarly,

$$\delta \otimes 1 \cdot \delta \cdot |i\rangle = \delta \otimes 1 \cdot (|i\rangle \otimes |i\rangle) = (|i\rangle \otimes |i\rangle) \otimes |i\rangle = |i\rangle \otimes |i\rangle \otimes |i\rangle = |iii\rangle \,.$$

$\square$

Using this lemma, we can recursively define

$$\delta^{i+1} := \delta \otimes 1 \otimes 1 \otimes \ldots 1 \cdot \delta^i \equiv 1 \otimes \delta \otimes 1 \otimes \ldots \otimes 1 \cdot \delta^i \equiv \ldots \equiv 1 \otimes 1 \otimes \ldots \otimes 1 \otimes \delta \cdot \delta^i,$$

with $\delta^1 := \delta$. The action of the operator δ^{n-1} taking V to $V \otimes V \otimes \ldots \otimes V$ generates exactly the "n-taxon process" as defined in Bryant (2009), which, in turn, is completely equivalent to the formalism given in Bashford *et al.* (2004).

If we note that $\delta \cdot 1 = 1 \otimes 1 \cdot \delta$ and consider Lemma 3.1, we see that, for $x = \alpha$ or β, we have

$$
\begin{aligned}
\delta^2 \cdot L_x :&= \delta \otimes 1 \cdot \delta \cdot L_x \\
&= \delta \otimes 1 \cdot (L_x \otimes L_x + 1 \otimes L_x + L_x \otimes 1) \cdot \delta \\
&= [(L_x \otimes L_x \otimes L_x + L_x \otimes 1 \otimes L_x + 1 \otimes L_x \otimes L_x) + \\
&\qquad 1 \otimes 1 \otimes L_x + (L_x \otimes L_x \otimes 1 + L_x \otimes 1 \otimes 1 + 1 \otimes L_x \otimes 1)] \cdot \delta^2 \\
&= \left(\sum_{A \subseteq \{1,2,3\}, A \neq \emptyset} L_x^{(A)} \right) \cdot \delta^2.
\end{aligned}
$$

Generalizing this result we have:

Lemma 3.3.

$$\delta^{n-1} \cdot L_x = \left(\sum_{A \subseteq [n], A \neq \emptyset} L_x^{(A)} \right) \cdot \delta^{n-1}. \tag{5}$$

Proof. The proof is by induction. We have shown that the result is true for $n = 3$. Assuming that it is true for some $n > 3$, we have

$$\delta^n \cdot L_x = (\delta \otimes 1 \otimes \ldots \otimes 1) \cdot \left(\sum_{A \subseteq [n], A \neq \emptyset} L_x^{(A)} \right) \cdot \delta^{n-1}$$

$$= (\delta \otimes 1 \otimes \ldots \otimes 1) \cdot \left(\sum_{A \subseteq [n-1], A \neq \emptyset} 1 \otimes L_x^{(A)} + \sum_{A \subseteq [n-1]} L_x \otimes L_x^{(A)} \right) \cdot \delta^{n-1}$$

$$= \left(\sum_{A \subseteq [n-1], A \neq \emptyset} 1 \otimes 1 \otimes L_x^{(A)} + \right.$$

$$\left. \sum_{A \subseteq [n-1]} \left(L_x \otimes L_x \otimes L_x^{(A)} + L_x \otimes 1 \otimes L_x^{(A)} + 1 \otimes L_x \otimes L_x^{(A)} \right) \right) \cdot \delta^n$$

$$= \left(\sum_{A \subseteq [n+1], A \neq \emptyset} L_x^{(A)} \right) \cdot \delta^n.$$

$\square$

For the two state model, recall that for $n \geq 1$ we have

$$Q^n = (\alpha L_\alpha + \beta L_\beta)^n = (-1)^{n-1}(\alpha + \beta)^{n-1} (\alpha L_\alpha + \beta L_\beta).$$

If we define

$$\mathfrak{L}_x^{[n]} := \sum_{A \subseteq [n], A \neq \emptyset} L_x^{(A)}, \tag{6}$$

then Lemma 3.3 implies that we have the intertwining

$$\delta^{n-1} \cdot L_x = \mathfrak{L}_x^{[n]} \cdot \delta^{n-1}.$$

We also note that we have the recursion

$$\mathfrak{L}_x^{[n]} = L_x \otimes \mathfrak{L}_x^{[n-1]} + 1 \otimes \mathfrak{L}_x^{[n-1]} + L_x \otimes 1 \otimes 1 \otimes \ldots \otimes 1,$$

and (after a little effort) it follows by induction on n that

$$\left(\mathfrak{L}_\alpha^{[n]} \right)^2 = \mathfrak{L}_\beta^{[n]} \mathfrak{L}_\alpha^{[n]} = -\mathfrak{L}_\alpha^{[n]}, \qquad \left(\mathfrak{L}_\beta^{[n]} \right)^2 = \mathfrak{L}_\alpha^{[n]} \mathfrak{L}_\beta^{[n]} = -\mathfrak{L}_\beta^{[n]}.$$

Inspection reveals that, for all n, $\mathfrak{L}_\alpha^{[n]}$ and $\mathfrak{L}_\beta^{[n]}$ satisfy exactly the same algebraic relations as L_α and L_β that were given in (2). It follows immediately that, for $n \geq 1$ we have

$$\left(\alpha \mathfrak{L}_\alpha^{[n]} + \beta \mathfrak{L}_\beta^{[n]} \right)^n = (-1)^{n-1}(\alpha + \beta)^{n-1} \left(\alpha \mathfrak{L}_\alpha^{[n]} + \beta \mathfrak{L}_\beta^{[n]} \right),$$

and

$$\delta^{n-1} \cdot (\alpha L_\alpha + \beta L_\beta)^n = (-1)^{n-1}(\alpha + \beta)^{n-1} \delta^{n-1} \cdot (\alpha L_\alpha + \beta L_\beta)$$

$$= (-1)^{n-1}(\alpha + \beta)^{n-1} \left(\alpha \mathfrak{L}_\alpha^{[n]} + \beta \mathfrak{L}_\beta^{[n]} \right) \cdot \delta^{n-1}$$

$$= \left(\alpha \mathfrak{L}_\alpha^{[n]} + \beta \mathfrak{L}_\beta^{[n]} \right)^n \cdot \delta^{n-1}.$$
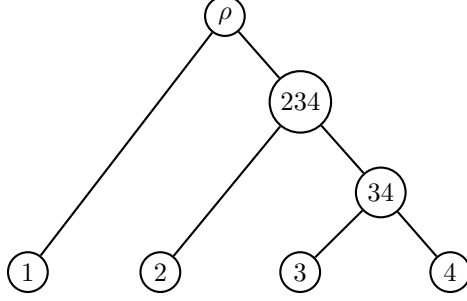
Putting this together we see that

Figure 2: A rooted tree on four leaves.

Lemma 3.4.

$$\delta^{n-1} \cdot \exp\left[\alpha L_\alpha + \beta L_\beta\right] = \exp\left[\alpha \mathfrak{L}_\alpha^{[n]} + \beta \mathfrak{L}_\beta^{[n]}\right] \cdot \delta^{n-1}.$$

In order to put all of the above to work, we require one final lemma regarding tensor products and the exponential map:

Lemma 3.5. *Given any two linear operators X and Y, we have*

$$e^X \otimes e^Y = e^{X \otimes 1 + 1 \otimes Y}.$$

Proof. Consider

$$e^X \otimes 1 = \left(1 + X + \tfrac{1}{2}X^2 + \ldots\right) \otimes 1 = 1 \otimes 1 + X \otimes 1 + \tfrac{1}{2}\left(X \otimes 1\right)^2 + \ldots = e^{X \otimes 1},$$

which implies that

$$e^X \otimes e^Y = e^X \otimes 1 \cdot 1 \otimes e^Y = e^{X \otimes 1} \cdot e^{1 \otimes Y} = e^{X \otimes 1 + 1 \otimes Y},$$

where the last identity follows because $X \otimes 1$ and $1 \otimes Y$ commute. $\square$

4 Alternative presentation of Markov models on trees

Consider the tree presented in Figure 2. Suppose we are given a root distribution $|\pi\rangle := \sum_i \pi_i |i\rangle$, a rate matrix $Q = \alpha L_\alpha + \beta L_\beta$, and edge weights $\tau_1, \tau_2, \tau_3, \tau_{34}$ and τ_{234}. The phylogenetic tensor corresponding to this tree can be generated as

$$P = e^{Q\tau_1} \otimes e^{Q\tau_2} \otimes e^{Q\tau_3} \otimes e^{Q\tau_4} \cdot 1 \otimes 1 \otimes \delta \cdot 1 \otimes 1 \otimes e^{Q\tau_{34}} \cdot 1 \otimes \delta \cdot 1 \otimes e^{Q\tau_{234}} \cdot \delta \cdot |\pi\rangle.$$

If we set $P = \sum_{i,j,k,l} p_{ijkl} |ijkl\rangle$ and interpret p_{ijkl} as the probability of observing the pattern $ijkl$ at the leaves of the tree, we see that this tensor is equivalent to specifying a joint distribution in the normal way (see Semple & Steel (2003) for example).

By setting $Q = \alpha L_\alpha + \beta L_\beta$ and applying Lemma 3.4, we find that

$$1 \otimes 1 \otimes \delta \cdot 1 \otimes 1 \otimes e^{Q\tau_{34}} = 1 \otimes 1 \otimes \left(\delta \cdot e^{Q\tau_{34}}\right) = 1 \otimes 1 \otimes e^{\left(\alpha \mathfrak{L}_\alpha^{[2]} + \beta \mathfrak{L}_\beta^{[2]}\right)\tau_{34}} \cdot 1 \otimes 1 \otimes \delta.$$

Now, we note that

$$1 \otimes 1 \otimes \delta \cdot 1 \otimes \delta = \left[1 \otimes (1 \otimes \delta)\right] \cdot 1 \otimes \delta = 1 \otimes (1 \otimes \delta \cdot \delta) = 1 \otimes \delta^2,$$

so again applying Lemma 3.4 we find that

$$1 \otimes \delta^2 \cdot 1 \otimes e^{Q\tau_{234}} = 1 \otimes e^{\left(\alpha \mathfrak{L}_\alpha^{[3]} + \beta \mathfrak{L}_\beta^{[3]}\right)\tau_{234}} \cdot 1 \otimes \delta^2.$$

We also set

$$1 \otimes \delta^2 \cdot \delta = \delta^3.$$

Thus

$$P = e^{Q\tau_1} \otimes e^{Q\tau_2} \otimes e^{Q\tau_3} \otimes e^{Q\tau_4} \cdot 1 \otimes 1 \otimes e^{\left(\alpha \mathfrak{L}_\alpha^{[2]} + \beta \mathfrak{L}_\beta^{[2]}\right)\tau_{34}} \cdot 1 \otimes e^{\left(\alpha \mathfrak{L}_\alpha^{[3]} + \beta \mathfrak{L}_\beta^{[3]}\right)\tau_{234}} \cdot \left|\delta^3 \pi\right\rangle,$$

with $\left|\delta^3 \pi\right\rangle := \delta^3 \cdot |\pi\rangle = p_0 |0000\rangle + p_1 |1111\rangle$. Finally, by applying Lemma 3.5 multiple times, we find that we can write

$$P = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{34} \mathcal{R}_{34}\right] \cdot \exp\left[\tau_{234} \mathcal{R}_{234}\right] \cdot \left|\delta^3 \pi\right\rangle, \tag{7}$$

where

$$\mathcal{R}_1 = \alpha \left(L_\alpha \otimes 1 \otimes 1 \otimes 1\right) + \beta \left(L_\beta \otimes 1 \otimes 1 \otimes 1\right),$$
$$\mathcal{R}_2 = \alpha \left(1 \otimes L_\alpha \otimes 1 \otimes 1\right) + \beta \left(1 \otimes L_\beta \otimes 1 \otimes 1\right),$$
$$\mathcal{R}_3 = \alpha \left(1 \otimes 1 \otimes L_\alpha \otimes 1\right) + \beta \left(1 \otimes 1 \otimes L_\beta \otimes 1\right),$$
$$\mathcal{R}_4 = \alpha \left(1 \otimes 1 \otimes 1 \otimes L_\alpha\right) + \beta \left(1 \otimes 1 \otimes 1 \otimes L_\beta\right),$$
$$\mathcal{R}_{34} = \alpha \left(1 \otimes 1 \otimes \mathfrak{L}_\alpha^{[2]}\right) + \beta \left(1 \otimes 1 \otimes \mathfrak{L}_\beta^{[2]}\right),$$
$$\mathcal{R}_{234} = \alpha \left(1 \otimes \mathfrak{L}_\alpha^{[3]}\right) + \beta \left(1 \otimes \mathfrak{L}_\beta^{[3]}\right).$$

Suppose instead of the tree above we considered the quartet given in Figure 3. The phylogenetic tensor corresponding to this tree is given by

$$P' = e^{Q\tau_1} \otimes e^{Q\tau_2} \otimes e^{Q\tau_3} \otimes e^{Q\tau_4} \cdot \delta \otimes \delta \cdot e^{Q\tau_{12}} \otimes e^{Q\tau_{34}} \cdot \delta \cdot |\pi\rangle.$$

By a similar argument to the one just given, it is possible to show that this tensor can be re-expressed as

$$P' = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{12} \mathcal{R}_{12} + \tau_{34} \mathcal{R}_{34}\right] \cdot \left|\delta^3 \pi\right\rangle,$$

with

$$\mathcal{R}_{12} = \alpha \left(\mathfrak{L}_\alpha^{[2]} \otimes 1 \otimes 1\right) + \beta \left(\mathfrak{L}_\beta^{[2]} \otimes 1 \otimes 1\right).$$

We can extend our definition (6) of $\mathfrak{L}_x^{[n]}$ to arbitrary subsets by taking

$$\mathfrak{L}_x^A := \sum_{A \subseteq B, A \neq \emptyset} L_x^{(A)},$$

for all $A \subseteq [n]$. Now label the elements of A as $A = \{a_1, a_2, \ldots, a_{|A|}\}$ and consider a permutation $\sigma \in \mathfrak{S}_n$ such that $\sigma(a_i) = i$ (obviously such a permutation always exists). If we allow σ to act on $V^{\otimes n}$ by permuting tensor factors, it is clear that

$$\sigma\left(\mathfrak{L}_x^A\right) = \mathfrak{L}_x^{[|A|]} \otimes 1^{([n-|A|])}.$$

From this we can conclude that for fixed A the operators $\left\{\mathfrak{L}_\alpha^A, \mathfrak{L}_\beta^A\right\}$ also satisfy the exact same algebra as $\{L_\alpha, L_\beta\}$:

$$\left(\mathfrak{L}_\alpha^A\right)^2 = \mathfrak{L}_\beta^A \mathfrak{L}_\alpha^A = -\mathfrak{L}_\alpha^A, \qquad \left(\mathfrak{L}_\beta^A\right)^2 = \mathfrak{L}_\alpha^A \mathfrak{L}_\beta^A = -\mathfrak{L}_\beta^A,$$

for all $A \subseteq [n]$. Using this result, we can unify the expressions for the rate matrices above by defining

$$\mathcal{R}_A := \alpha \mathfrak{L}_\alpha^A + \beta \mathfrak{L}_\beta^A.$$
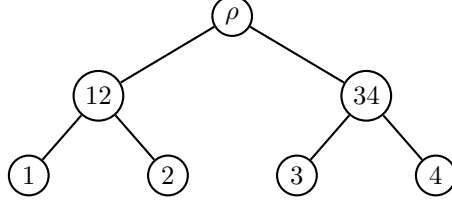
8

Figure 3: Alternative tree on four leaves.

Evidently $A \cap B$ implies that $[\mathcal{R}_A, \mathcal{R}_B] = 0$ and we see that there are several ways we can express our two phylogenetic tensors. For instance the following presentations are all equally valid:

$$P = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{34} \mathcal{R}_{34}\right] \cdot \exp\left[\tau_{234} \mathcal{R}_{234}\right] \cdot \left|\delta^3 \pi\right\rangle$$
$$= \exp\left[\tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_1 \mathcal{R}_1 + \tau_{34} \mathcal{R}_{34}\right] \cdot \exp\left[\tau_{234} \mathcal{R}_{234}\right] \cdot \left|\delta^3 \pi\right\rangle$$
$$= \exp\left[\tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{34} \mathcal{R}_{34}\right] \cdot \exp\left[\tau_1 \mathcal{R}_1 + \tau_{234} \mathcal{R}_{234}\right] \cdot \left|\delta^3 \pi\right\rangle$$
$$= \exp\left[\tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{34} \mathcal{R}_{34}\right] \cdot \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_{234} \mathcal{R}_{234}\right] \cdot \left|\delta^3 \pi\right\rangle.$$

The content of our main theorem is that there is a canonical choice of presentation of a phylogenetic tensor for arbitrary trees. Consider the presentations of the quartet tensors discussed above:

$$P = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{34} \mathcal{R}_{34}\right] \cdot \exp\left[\tau_{234} \mathcal{R}_{234}\right] \cdot \left|\delta^3 \pi\right\rangle,$$
$$P' = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 + \tau_4 \mathcal{R}_4\right] \cdot \exp\left[\tau_{12} \mathcal{R}_{12} + \tau_{34} \mathcal{R}_{34}\right] \cdot \left|\delta^3 \pi\right\rangle.$$

Theorem 4.1. *Consider a rooted tree with n leaves $\mathcal{T} = \{A_1, A_2, \ldots, A_{2n+2}\}$, where $A_i \subset [n]$. Given a root distribution π, any rate parameters α and β and weights $\{\tau_{A_1}, \tau_{A_2}, \ldots, \tau_{A_{2n+2}}\}$, a phylogenetic tensor (or joint distribution) P at the leaves of this split system can be expressed as*

$$P = \exp\left[\mathcal{X}_1\right] \cdot \exp\left[\mathcal{X}_2\right] \cdot \ldots \cdot \exp\left[\mathcal{X}_{n-1}\right] \cdot \left|\delta^{n-1} \pi\right\rangle,$$

with

$$\left|\delta^{n-1} \pi\right\rangle := \delta^{n-1} \cdot \left|\pi\right\rangle = p_0 \left|00\ldots0\right\rangle + p_1 \left|11\ldots1\right\rangle,$$

and

$$\mathcal{X}_i = \sum_{A, |A|=i} \tau_A \mathcal{R}_A.$$

Proof. The proof is by induction. Clearly a phylogenetic tensor on two leaves can be placed into the required form:

$$P = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2\right] \cdot \left|\delta \pi\right\rangle.$$

Assume that for $k > 2$ any phylogenetic tensor on k leaves $P^{(k)}$ can be expressed in the required form:

$$P^{(k)} = \exp\left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \ldots + \tau_n \mathcal{R}_n\right] \cdot \exp\left[\sum_{A, |A|=2} \tau_A \mathcal{R}_A\right] \cdot \ldots \cdot \exp\left[\sum_{A, |A|=k-1} \tau_A \mathcal{R}_A\right] \cdot \left|\delta^{k-1} \pi\right\rangle.$$

Without loss of generality we generate a phylogenetic tensor on $k+1$ leaves, $P^{(k+1)}$, by branching $P^{(k)}$ at the leaf k. This is expressed algebraically by

$$P^{(k)} \rightarrow P^{(k+1)} = (1 \otimes 1 \otimes \ldots \otimes 1 \otimes \delta) \cdot P^{(k)}.$$

For an arbitrary subset $A \subset [k]$, we have

$$1 \otimes 1 \otimes \ldots \otimes 1 \otimes \delta \cdot \mathcal{R}_A = \mathcal{R}_{A'}$$

9

where
$$A' = \begin{cases} A \cup \{k+1\} \text{ , if } k \in A, \\ A \text{ , otherwise.} \end{cases}$$

Pushing right through we have

$$P^{(k+1)} = \exp\left[\sum_{A,|A|=1} \tau_A \mathcal{R}_{A'}\right] \cdot \exp\left[\sum_{A,|A|=2} \tau_A \mathcal{R}_{A'}\right] \cdot \ldots \cdot \exp\left[\sum_{A,|A|=n-1} \tau_A \mathcal{R}_{A'}\right] \cdot \left|\delta^k \pi\right\rangle.$$

Now consider, for a given $1 \leq i < k-1$, the term

$$\exp\left[\sum_{A,|A|=i} \tau_A \mathcal{R}_{A'}\right].$$

As we are dealing with a tree $\mathcal{T}$, it is only possible that one (and only one) of the subsets (B_{i+1}, say) in the summation has cardinality $i+1$. Additionally, B is disjoint from all of the other subsets, so we may write

$$\exp\left[\sum_{A,|A|=i,A\neq\emptyset} \tau_A \mathcal{R}_{A'}\right] = \exp\left[\sum_{A,|A|=i,A\neq\emptyset,A\neq B} \tau_A \mathcal{R}_A\right] \exp\left[\tau_B \mathcal{R}_{B_{i+1}}\right].$$

Exactly the same argument is valid for the $i+1$ term:

$$\exp\left[\sum_{A,|A|=i+1,A\neq\emptyset} \tau_A \mathcal{R}_{A'}\right] = \exp\left[\sum_{A,|A|=i+1,A\neq\emptyset,A\neq B_{i+2}} \tau_A \mathcal{R}_A\right] \exp\left[\tau_B \mathcal{R}_{B+2}\right].$$

Thus we can express the product of these two terms as

$$\exp\left[\sum_{A,|A|=i} \tau_A \mathcal{R}_{A'}\right] \cdot \exp\left[\sum_{A,|A|=i+1} \tau_A \mathcal{R}_{A'}\right]$$
$$= \exp\left[\sum_{A,|A|=i,A\neq\emptyset,A\neq B} \tau_A \mathcal{R}_A\right] \exp\left[\tau_B + \mathcal{R}_{B_{i+1}} + \sum_{A,|A|=i+1,A\neq\emptyset,A\neq B_{i+2}} \tau_A \mathcal{R}_A\right]$$
$$\cdot \exp\left[\tau_B \mathcal{R}_{B+2}\right].$$

Continuing in this way we can place $P^{(k+1)}$ in the required form and the theorem follows by induction on k. ☐

Recall that, in the basis $\{|i_1 i_2 \ldots i_n\rangle\}$, a phylogenetic tensor P can be expressed as

$$P = \sum_{i_1,i_2,\ldots i_n} p_{i_1 i_2 \ldots i_n} |i_1 i_2 \ldots i_n\rangle,$$

where $p_{i_1 i_2 \ldots i_n}$ is the probability of observing the pattern $i_1 i_2 \ldots i_n$ at the leaf vertices.

From a biological perspective, it is apparent that the form given in Theorem 4.1 that utilizes a cardinality ordering is somewhat mysterious. However a little thought using commutivity (or otherwise) of the various operators shows that it is not so much the cardinality that matters, but it is that the operators that arise independently across branches of the tree are necessarily commutative and, conversely, those that do not commmute necessarily have non-zero intersection and hence are not independent. Noting that there is some freedom in the final expression, we see that the cardinality ordering is simply a nice way of unifying the description for arbitrary trees.

It is clear that if we take P as in Theorem 4.1 in the case that the split system $\mathcal{S}$ is compatible, we have a probability distribution on a tree identical to the standard presentation usually given in phylogenetics. However, the construction we have given naturally generalizes to a model on the most general split systems with trees occuring as sub-models (set the weights for incompatible splits equal to zero). There is also an obvious generalization to the case where each split has a unique rate matrix – simply give additional split labels to the rate parameters: $\alpha \to \alpha_A$ and $\beta \to \beta_A$.

In the next section we explore the $\alpha = \beta$ case in detail.

5   (In)consistency with the Hadamard conjugation

In this section we consider the "binary-symmetric" case where $\alpha = \beta = \frac{1}{2}$, so that we can write

$$Q = \frac{1}{2}\left(L_\alpha + L_\beta\right) = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2}\left(-1 + K\right),$$

where $K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the permutation matrix taking $|0\rangle \rightleftharpoons |1\rangle$.

For any tree $\mathcal{T}$ represented as a split system, it is shown in Bashford *et al.* (2004) that we can write

$$P = \exp\left[\sum_{x \in \mathcal{T}} w_x \frac{1}{2}\left(K^{(x)} - 1 \otimes 1 \otimes \ldots \otimes 1\right)\right]\left|\delta^{n-1}\pi\right\rangle,$$

where $\{w_x\}_{x \in \mathcal{T}}$ is any set of edge weights on $\mathcal{T}$. We will refer to this constructing of a phylogenetic tensor as the "K-representation".

On the other hand, we have shown in Theorem 4.1 that for $\alpha = \beta = \frac{1}{2}$ we can write

$$P = \exp\left[\mathcal{X}_1\right] \cdot \exp\left[\mathcal{X}_2\right] \cdot \ldots \cdot \exp\left[\mathcal{X}_n\right]\left|\delta^{n-1}\pi\right\rangle,$$

with

$$\mathcal{X}_i = \sum_{A,|A|=i} \tau_A \mathcal{R}_A,$$

and

$$\mathcal{R}_A = \sum_{B \subseteq A} \tfrac{1}{2}\left(\mathfrak{L}_\alpha^B + \mathfrak{L}_\beta^B\right).$$

We will refer to this construction of a phylogenetic tensor as the "$\mathfrak{L}$-representation". We will show that for a tree these two representations are exactly equal, but for arbitrary split systems this is not the case.

We will find it convenient to label the vector $|i_1 i_2 \ldots i_n\rangle$ by the subset $A \subset [n]$ defined by setting $j \in A$ if and only if $i_j = 1$. For example if $n = 6$, we have

$$|\emptyset\rangle = |000000\rangle,$$
$$|[6]\rangle = |\{1,2,3,4,5,6\}\rangle = |111111\rangle,$$
$$|\{2,3,4\}\rangle = |011100\rangle,$$
$$|\{5\}\rangle = |000010\rangle.$$

Lemma 5.1. *If*

(a.) $A \subseteq B$, it follows that $\mathcal{R}_A |B\rangle = |B - A\rangle - |B\rangle$,

11

(b.) $A \cap B = \emptyset$, *it follows that* $\mathcal{R}_A \left| B \right\rangle = \left| B \cup A \right\rangle - \left| B \right\rangle$.

If either case (a.) or (b.), we have

$$\mathcal{R}_A \left| B \right\rangle = \left(K^{(A)} - 1 \otimes 1 \otimes \ldots \otimes 1 \right) \left| B \right\rangle .$$

Proof. For all $A \subseteq B$ we have

$$K^{(A)} \left| B \right\rangle = \left| B - A \right\rangle .$$

On the other hand we know that

$$\mathcal{R}_{[n]} \left| \emptyset \right\rangle = \left| [n] \right\rangle - \left| \emptyset \right\rangle ,$$
$$\mathcal{R}_{[n]} \left| [n] \right\rangle = \left| \emptyset \right\rangle - \left| [n] \right\rangle .$$

For $A \subseteq B \subseteq [n]$ it is always possible to permute tensor factors to write

$$\left| B \right\rangle \cong \left| [n_A] \right\rangle \otimes \left| B' \right\rangle ,$$

where $n_A = |A|$ and $B' \subset [n] - A$ with $B' := B - A$. Thus

$$\begin{aligned}
\mathcal{R}_A \left| B \right\rangle &\cong \left(\mathcal{R}_{[n_A]} \left| [n_A] \right\rangle \right) \otimes \left| B' \right\rangle \\
&= \left(\left| \emptyset \right\rangle - \left| [n_a] \right\rangle \right) \otimes \left| B' \right\rangle \\
&\cong \left| B - A \right\rangle - \left| (B - A) \cup A \right\rangle \\
&= \left| B - A \right\rangle - \left| B \right\rangle ,
\end{aligned}$$

which proves the lemma for $A \subseteq B$. The $A \cap B = \emptyset$ case follows from a similar argument. $\square$

For two subsets A, B taken from a compatible split system with $|A| = |B|$ it is the case that $A \cap B = \emptyset$, which in turn implies that $[\mathcal{R}_A, \mathcal{R}_B] = 0$. Thus we can make the replacement

$$\exp \left[\mathcal{X}_i \right] := \exp \left[\sum_{A, |A| = i} \tau_A \mathcal{R}_A \right] = \prod_{A, |A| = i} \exp \left[\tau_A \mathcal{R}_A \right] ,$$

where by commutivity the product can be ordered in any way we please.

Thus, in the $\mathfrak{L}$-representation, we see that for a tree we can write

$$P = \ldots \exp \left[\tau_{A_3} \mathcal{R}_{A_3} \right] \exp \left[\tau_{A_2} \mathcal{R}_{A_2} \right] \exp \left[\tau_{A_1} \mathcal{R}_{A_1} \right] \left| \delta^{n-1} \pi \right\rangle ,$$

with, due to the fact we are dealing with a tree, *either* $A_i \cap A_{i+1} = \emptyset$ *or* $A_{i+1} \subset A_i$. Noting that $\left| \delta^{n-1} \pi \right\rangle = \pi_0 \left| \emptyset \right\rangle + \pi_1 \left| [n] \right\rangle$ and repeated application of Lemma 5.1 then gives:

Theorem 5.2. *For compatible split systems, the "$\mathfrak{L}$-representation" and the "K-representation" give rise to identical phylogenetic tensors.*

For *arbitrary* split systems, however, this is not true, as the example in the next section shows.

6 Phylogenetic networks and "epochs"

Consider the three taxa phylogenetic tree given in Figure 6 (a). As was proved above, if we take the binary symmetric model we get an identical probability distribution if we use either the $\mathfrak{L}$-representation $P_\ell = \exp \left[\tau_1 \mathcal{R}_1 + \tau_2 \mathcal{R}_2 + \tau_3 \mathcal{R}_3 \right] \cdot \exp \left[\tau_{12} \mathcal{R}_{12} \right] \left| \delta^2 \pi \right\rangle$ or the K-representation $P_k = e^{-\lambda} \exp \left[\tau_1 K^{(1)} + \tau_2 K^{(2)} + \tau_3 K^{(3)} + \tau_{12} K^{(12)} \right] \left| \delta^2 \pi \right\rangle$, with $\lambda = \tau_1 + \tau_2 + \tau_3 + \tau_{12}$.

We would like to introduce the additional parameter τ_{23} associated with the "imaginary" split $1|23$ to these probability distributions. We will do this in a way which is consistent with the design
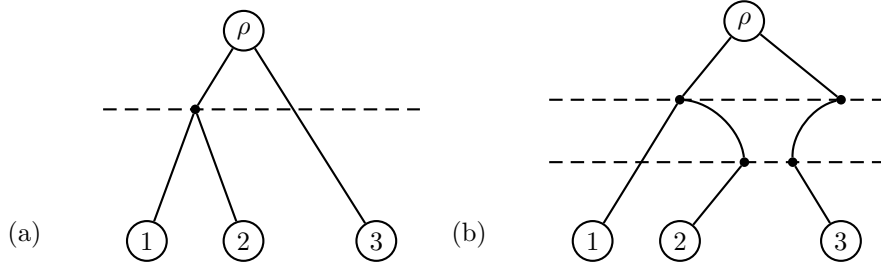
Figure 4: A three taxa tree (a.) is modified by the introduction of the addtional split $\{23\}$ in (b).

given in Figure 6 (b), where the evolutionary history in broken up into three epochs: a. divergence of taxa 3 away from 1 and 2, b. convergent evolution of taxa 2 and 3, with independent divergence of taxa 3, and c. independent divergence of all taxa.

To model this situation we must introduce the additional edge to each representation. To make the presentation as simple as possible, we set a molecular clock on the model such that $\tau_2 = \tau_1$ and $\tau_3 = \tau_{12} + \tau_1$, and we introduce a scaling parameter $\theta \in [0,1]$ to control the length of the second epoch as a proportion of the third epoch by setting $\tau_{23} = \tau_1\theta$. As all operators in the K-representation commute, the only choice available in this case is to take

$$P'_K = e^{-\lambda} \exp\left[\tau_1 K^{(1)} + \tau_1(1-\theta)K^{(2)} + (\tau_{12} + \tau_1(1-\theta)) K^{(3)} + \tau_{12}K^{(12)} + \tau_1\theta K^{(23)}\right] \left|\delta^2\pi\right\rangle .$$

This is exactly consistent with the generalizations given in Bryant (2005b, 2009).

For the $\mathfrak{L}$-representation we do *not* have commutivity of the operators $\mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_{12}$ with the new operator $\mathcal{R}_{23}$, thus we need to proceed more carefully as there is some choice in how the extra edge is introduced. Using the diagram and its three epochs as a guide, we take

$$P'_{\mathfrak{L}} = \exp\left[\tau_1(1-\theta)\left(\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3\right)\right] \cdot \exp\left[\tau_1\theta\left(\mathcal{R}_1 + \mathcal{R}_{23}\right)\right] \cdot \exp\left[\tau_{12}\left(\mathcal{R}_3 + \mathcal{R}_{12}\right)\right] \left|\delta^2\pi\right\rangle .$$

For clarity of comparison, we write the K-representation in epoch form:

$$P'_K = e^{-\lambda} \exp\left[\tau_1(1-\theta)\left(K^{(1)} + K^{(2)} + K^{(3)}\right)\right] \cdot \exp\left[\tau_1\theta\left(K^{(1)} + K^{(23)}\right)\right]$$
$$\cdot \exp\left[\tau_{12}\left(K^{(3)} + K^{(12)}\right)\right] \left|\delta^2\pi\right\rangle .$$

Now consider the state of the probability distribution at the beginning of epoch 2. As we are dealing with the binary symmetric model, it is clear that the probability of the any state $|ijk\rangle$ is invariant to permutation of the states $0 \rightleftharpoons 1$. Also, the structure of the tree up to the start of epoch 2 implies that the probability of any state of the form $|ijk\rangle$ where $i \neq j$ is of probability zero. Thus we can assume at the start of epoch 2 that the distribution is of the form

$$P = (1-q)\tfrac{1}{2}\left(|000\rangle + |111\rangle\right) + q\tfrac{1}{2}\left(|001\rangle + |110\rangle\right),$$

where, because we are considering a continuous-time model, $q \in \left[0, \tfrac{1}{2}\right)$.

Considering the definitions

$$\mathcal{R}_{23} = 1 \otimes \left(L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha + L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta\right),$$

and

$$K^{(23)} = 1 \otimes K \otimes K = \mathcal{R}_{23} + 1 \otimes \left(L_\alpha \otimes L_\beta + L_\beta \otimes L_\alpha\right),$$

it follows that transition rates between the four existing states in the two cases are given by the two graphs in Figure 6, where all transition rates are equal. The crucial thing to note is that $\mathcal{R}_{23}$ "corrects" patterns that are inconsistent with the split $1|23$, whereas $K^{(23)}$ simply *premutes* these two states.
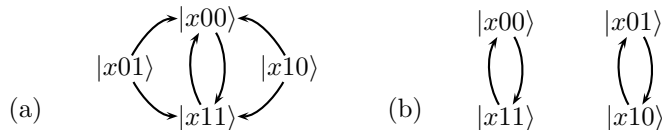
Figure 5: Transition rates for states $|xyz\rangle$ under (a.) $\mathcal{R}_{23}$, and (b.) $K^{(23)} - 1 \otimes 1 \otimes 1$.

It is now clear the there is quite a marked difference between the $\mathfrak{L}$- and K-representations. The $\mathfrak{L}$ representations introduces a natural notion of the "coming together" of taxa. In fact it is easy to see directly from the diagram that in the limit of extension of the edge τ_{23} to infinity that the probability distribution will converge to

$$P = \left(\left(1 - \tfrac{1}{2}q\right)\tfrac{1}{2} + \tfrac{1}{4}q\right)\left(|000\rangle + |111\rangle\right) + \tfrac{1}{2}q\tfrac{1}{2}\left(|011\rangle + |100\rangle\right)$$
$$= \left(1 - \tfrac{1}{2}q\right)\tfrac{1}{2}\left(|000\rangle + |111\rangle\right) + \tfrac{1}{2}q\tfrac{1}{2}\left(|011\rangle + |100\rangle\right),$$

which is consistent with the probability distribution that would arise under a tree where taxa 1 has diverged from 2 and 3, but there has been zero divergence of taxa 2 and 3 themselves. This behaviour is of the course the reasoning behind the way we have chosen to draw our diagram Figure 6 (b).

The K-representation cannot achieve this type of convergence, with its limiting state being

$$(1 - q)\tfrac{1}{4}\left(|000\rangle + |011\rangle + |111\rangle + |100\rangle\right) + q\tfrac{1}{4}\left(|001\rangle + |010\rangle + |110\rangle + |101\rangle\right).$$

7 Conclusion

In this paper we have shown how to express the two state continuous-time general Markov model on trees in such a way that allows extension to arbitrary split systems and even more general network models. By reviewing the lemmas in Section 2 it is clear that the results extend easily to the general Markov model with more character states. However, we defer confirmation of this observation to future work.

In Section 5 we showed that our discussion gives rise to phylogenetic models for the general Markov model that are identical to previous approaches only on *trees*, and in Section 6 we give a simple example that shows how our approach allows for convergent evolution of previously divergent lineages (a structural property that was previously unobtainable).

Besides its theoretical interest, we expect that the ability to model convergent evolution in this way will have a significant application where it is known that particular datasets exhibit non-treelike behaviour due to population genetic properties such as incomplete lineage sorting and other effects that confound strictly treelike models (as discussed in Section 1). We suspect that exploration of the relation between the network models that arise in our discussion and simple models of population genetics is likely to yield significant additional insight.

Finally, comparison of our network models to the distribution space generated by the so-called "mixture models" (Pagel & Meade, 2004) is also in need of investigation. For example, comparison of a mixture of the trees 12|3 and 1|23 to our network example given in Section 6 should yield significant theoretical insight into the biological meaning and plausibility of these differing model classes. Careful scientific thought is required to tease out what biological processes are explicitly (or implicitly) being modelled by either of these approaches.

Of course these exciting possibilities must be tempered by analysis of the *identifiability* (or otherwise) of the models that arise by taking more general networks. Establishing identifiability is not only essential from a statistical inference point of view, but, in this case, may lead to natural restrictions of the types of network that can be realistically used for phylogenetic inference.

References

BANDELT, H.-J. (1994). Phylogenetic networks. In: *Verhandlungen des Naturwissenschaftlichen Vereins Hamburg*, vol. 34.

BANDELT, H. J. & DRESS, A. W. M. (1992). Split Decomposition: A new and useful approach to phylogenetic analysis of distance data. *Mol. Phylogenet. Evol.* **1**, 242–252.

BASHFORD, J. D., JARVIS, P. D., SUMNER, J. G. & STEEL, M. A. (2004). $U(1) \times U(1) \times U(1)$ symmetry of the Kimura 3ST model and phylogenetic branching processes. *J. Phys. A Math. Gen.* **37**, L1–L9.

BRYANT, D. (2005a). Extending tree models to split networks. In: *Algebraic Statistics and Computational Biology* (PACHTER, L. & STURMFELS, B., eds.). Cambridge University Press, pp. 297–368.

BRYANT, D. (2005b). On the uniqueness of the selection criterion in Neighbor-Joining. *J. Class.* **22**, 3–15.

BRYANT, D. (2009). Hadamard phylogenetic methods and the *n*-taxon process. *Bull. Math. Biol.* **71**, 297–309.

BRYANT, D. & MOULTON, V. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265.

HENDY, M. D. & PENNY, D. (1989). A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**, 297–309.

HOLLAND, B. & MOULTON, V. (2004). *Algorithms in Bioinformatics*, chap. Consensus networks: A method for visualising incompatibilities in collections of trees. Springer, pp. 165–176.

HOLLAND, B. R., JERMIIN, L. S. & MOULTON, V. (2006). Improved consensus network techniques for genome-scale phylogeny. *Mol. Biol. Evol.* **23**, 848–855.

JARVIS, P. D., BASHFORD, J. D. & SUMNER, J. G. (2005). Path integral formulation and Feynman rules for phylogenetic branching models. *J. Phys. A Math. Gen.* **38**, 9621–9647.

JARVIS, P. D. & SUMNER, J. G. (2010). Markov invariants for phylogenetic rate matrices derived from embedded submodels. *ArXiv e-prints* **1008.1121**.

JERMIIN, L. S., HO, S. Y. W., ABABNEH, F., ROBINSON, J. & LARKUM, A. W. D. (2004). The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.* **53**, 638–643.

JIN, G., NAKHLEH, L., SNIR, S. & TULLER, T. (2006). Maximum likelihood of phylogenetic networks. *Bioinformatics* **21**, 26042611.

JOHNSON, J. E. (1985). Markov-type Lie groups in $GL(n, R)$. *J. Math. Phys.* **26**, 252–257.

PAGEL, M. & MEADE, A. (2004). A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* **53**, 571–581.

PENNY, D. (2005). Relativity for molecular clocks. *Nature* **436**, 183–184.

PROCESI, C. (2007). *Lie Groups: An Approach through Invariants and Representations*. Springer.

SEMPLE, C. & STEEL, M. (2003). *Phylogenetics*. Oxford Press.

STRIMMER, K. & MOULTON, V. (2000). Likelihood analysis of phylogenetic networks using directed graphical models. *Mol. Biol. Evol.* **17**, 875881.

Sumner, J. G., Charleston, M. A., Jermiin, L. S. & Jarvis, P. D. (2008). Markov invariants, plethysms, and phylogenetics. *J. Theor. Biol.* **253**, 601–615.

Sumner, J. G. & Jarvis, P. D. (2005). Entanglement invariants and phylogenetic branching. *J. Math. Biol.* **51**, 18–36.

von Haeseler, A. & Churchill, G. A. (1993). Network models for sequence evolution. *J. Mol. Evol.* **37**, 77–85.