BMC
Evolutionary Biology

A tensorial approach to the inversion of group-based phylogenetic models

Jeremy G Sumner[*], Peter D Jarvis and Barbara R Holland

Abstract

Background: Hadamard conjugation is part of the standard mathematical armoury in the analysis of molecular phylogenetic methods. For group-based models, the approach provides a one-to-one correspondence between the so-called "edge length" and "sequence" spectrum on a phylogenetic tree. The Hadamard conjugation has been used in diverse phylogenetic applications not only for inference but also as an important conceptual tool for thinking about molecular data leading to generalizations beyond strictly tree-like evolutionary modelling.

Results: For general group-based models of phylogenetic branching processes, we reformulate the problem of constructing a one-one correspondence between pattern probabilities and edge parameters. This takes a classic result previously shown through use of Fourier analysis and presents it in the language of tensors and group representation theory. This derivation makes it clear why the inversion is possible, because, under their usual definition, group-based models are defined for abelian groups only.

Conclusion: We provide an inversion of group-based phylogenetic models that can implemented using matrix multiplication between rectangular matrices indexed by ordered-partitions of varying sizes. Our approach provides additional context for the construction of phylogenetic probability distributions on network structures, and highlights the potential limitations of restricting to group-based models in this setting.

Keywords: Groups, Representation theory, Symmetry, Markov chains

Background

Fundamental to evolutionary biology is the development and implementation of molecular phylogenetic methods [1]. These methods provide the means to reconstruct the past evolutionary history of biological entities given present-day molecular data, such as DNA. Considering Kimura's neutral theory of molecular evolution, it is logical to apply a stochastic model at the level of DNA substitutions to construct probabilistic description of what molecular alignments are expected to be observed, given a proposed evolutionary history (tree topology and edge lengths). is commonly implemented assuming an IID (across sites in the alignment) and Markov process for DNA substitution, leading to a model that has a continuous-time Markov chain at its core (see Semple and Steel [2] for an introduction to the mathematics underlying modern phylogenetic methodology).

In a series of papers, Hendy and colleagues introduced the Hadamard conjugation as a novel tool for phylogenetic analyses [3-5]. They found an invertible relationship between a phylogenetic tree, as characterized by its edge length spectrum, and the probability distribution of site patterns (referred to as the sequence spectrum). Originally introduced only for the 2-state symmetric model, the Hadamard conjugation was later extended to the K3ST model [6-8] and further to any of the so-called "group-based" models [9]. Hadamard conjugation has been used as both a tool for simulation [10] and to look at statistical properties of methods, exploring the inconsistency of parsimony under a molecular clock [5,11]. For these sorts of applications, following the notation in Felsenstein [1], we can use the Hadamard transform H to start with an edge length spectrum γ and calculate the sequence spectrum $s = H^{-1}\log(H\gamma)$. The beauty of Hadamard conjugations is that one can also begin with an observed sequence spectrum $\hat{s}$ and perform the inverse of the conjugation to empirically obtain an edge length spectrum

*Correspondence: jsumner@utas.edu.au
School of Physical Sciences, University of Tasmania, Hobart TAS 7001, Australia

$\hat{\gamma} = H^{-1} \log(H\hat{s})$. Although it is not expected that the $\hat{\gamma}$ spectrum will precisely match a tree, Hendy [12] proposed using an optimisation criterion to map from $\hat{\gamma}$ to the "closest tree".

Several authors have commented that it is potentially a useful feature of Hadamard conjugation that data isn't forced onto a fixed tree. The conflicting information can be retained and interpreted in the form of a "lentoplot" [13] or a splits-graph [14], with both of these methods implemented in *Spectronet* [15]. Schliep [16] gives some more statistical justification for such an approach by making a link to modern statistical techniques such as the Lasso and Ridge regression.

von Haeseler and Churchill [17] seems to be the first paper that explicitly suggests using Hadamard conjugation to provide a likelihood framework for networks. The principle idea in this work was to start with an edge length spectrum that encodes a set of incompatible splits, use the Hadamard transformation to get site probabilities and use these to determine a likelihood. This idea was further explored by Bryant [18], and Bryant [19] followed this through defining the "*n*-taxon process" for group-based models. It should be noted that likelihoods calculated via Hadamard are not equivalent to likelihoods calculated by taking a mixture of trees. Indeed, Matsen and Steel [20], Matsen et al. [21] used Hadamard methods in combination with phylogenetic invariants to show that mixtures of trees with the same topology can exactly mimic another tree under the 2-state model. Considering biological applications, thinking in terms of mixtures of trees or partitions where the data can be thought of as arising on a set of trees [22-24] seems more reasonable than the Hadamard conjugation. Strimmer and Moulton [25] suggested using split networks as a spring board to likelihood-based analyses on DAGs, but later identified several problems with the approach [26]; most notably, in split-networks internal nodes do not have a biological interpretation as an ancestor.

In Sumner et al. [27], we gave some additional insight into the interpretation of applying the Hadamard conjugation in a network setting. We showed that permutation group structure inherent to the Hadamard transformation – as for any group-based model – restricts the resulting process from being capable of reproducing truly convergent processes. This is a serious limitation, as one of the biological motivations for explicit network models is the ability to model convergent processes. We also presented an alternative algebraic formalism for the general Markov model, analogous to the *n*-taxon process, but capable of reproducing convergent processes.

From the point of view of group representation theory, the inversion of group-based models relies on the fact that the *irreducible* representations of an abelian group are one-dimensional, and the model structure essentially reduces to analysing group characters – hence the standard presentation of a Fourier inversion. In this article, we make this connection concrete. For the general Markov model, it is then immediately apparent that an analogous inversion is not possible because the algebraic structure underlying the model is not abelian and hence the irreducible representations are not one-dimensional. In fact, to obtain one-dimensional representations for the general Markov model, it is necessary to apply higher-degree polynomial maps (beyond the degree 1, linear case), and define "Markov invariants" [28]. These invariants present one-dimensional representations but at the cost of the higher degree – degree 5 in the case of the general Markov model with four states on quartet trees [29,30]. This connection between Hadamard transformation and Markov invariants is an interesting one, but we do not discuss it further here.

In this paper we approach the inversion of group-based phylogenetic models by taking a representation-theoretic perspective and working explicitly with tensor indices. Our approach rests heavily on the formalism of "phylogenetic tensors", as presented in Bashford et al. [31], for the binary-symmetric and K3ST model, and Sumner et al. [27,28], for the general Markov model.

Although the main inversion results presented here are not more general than those in in Székely et al. [7], we think it is important to reformulate them using the language of tensors and representation theory. This viewpoint has already led to new approaches for modeling convergent evolution [27] and for studying non-group-based models [28]. However, in none of our previous work was the link to Hadamard conjugation explicitly discussed. By presenting an old technique (Hadamard conjugation) in a new light we hope to introduce other researchers to the viewpoint of tensor analysis and representation theory.

Methods
Group-based models
We consider the continuous-time formulation of Markov processes, and show how to implement the inversion of a group-based phylogenetic model based on *any* abelian group G. We note that such an inversion requires a map from tensor product space (where elements are indexed by ordered-*n*-partitions) to phylogenetic splits (where elements are indexed by bipartitions). We achieve this by finding canonical maps from bipartitions to ordered-*n*-partitions.

For a group G (not necessarily abelian) with order $|G| = d$, we write $G = \{\sigma_1, \sigma_2, \ldots, \sigma_d\}$, and, when necessary, write $\epsilon \in G$ to specify the identity element of G. We will discuss the "regular representation" of G shortly, but

skipping ahead we find that any rate matrix Q occurring in a group-based Markov model can be written in the form

$$Q = -\lambda \mathbf{1} + \sum_{\epsilon \neq \sigma \in G} \alpha^\sigma K_\sigma, \tag{1}$$

where each $0 \leq \alpha^\sigma \in \mathbb{R}$, $\lambda = \sum_{\epsilon \neq \sigma \in G} \alpha^\sigma$ and the K_σ are the permutation matrices corresponding to the (non-identity) group elements $\sigma \in G$.

For the reader interested in deriving this result, consider the d-dimensional vector space $\langle G \rangle_\mathbb{C} \equiv \langle \sigma_1, \sigma_2, \ldots, \sigma_d \rangle_\mathbb{C} = \{v = v_1 \sigma_1 + v_2 \sigma_2 + \ldots + v_d \sigma_d : v_i \in \mathbb{C}\}$, with scalar multiplication and vector addition defined via

$$
\begin{aligned}
v + \lambda v' &= (v_1 \sigma_1 + v_2 \sigma_2 + \ldots + v_d \sigma_d) \\
&\quad + \lambda(v'_1 \sigma_1 + v'_2 \sigma_2 + \ldots + v'_d \sigma_d) \\
&= (v_1 + \lambda v'_1)\sigma_1 + (v_2 + \lambda v'_2)\sigma_2 + \ldots \\
&\quad + (v_d + \lambda v'_d)\sigma_d,
\end{aligned}
$$

for all $v, v' \in \langle G \rangle_\mathbb{C}$ and $\lambda \in \mathbb{C}$. The *regular representation*, $\rho_{\text{reg}} : G \to GL(d, \mathbb{C})$, is then defined by setting the group action

$$\sigma : v \mapsto \sigma v = v_1(\sigma \sigma_1) + v_2(\sigma \sigma_2) + \ldots + v_d(\sigma \sigma_d),$$

for all $v \in \langle G \rangle_\mathbb{C}$ and $\sigma \in G$. If we fix $\{\sigma_1, \sigma_2, \ldots, \sigma_d\}$ as an ordered basis for $\langle G \rangle_\mathbb{C}$, it is then clear – via Cayley's theorem – that each group element σ gets mapped to a permutation matrix $K_\sigma := \rho_{\text{reg}}(\sigma)$, with $K_\sigma \sigma_i = \sum_j [K_\sigma]_i^j \sigma_j := \sigma \sigma_i$. Thus K_σ has matrix elements

$$[K_\sigma]_i^j = \begin{cases} 1, & \text{if } \sigma_j = \sigma \sigma_i, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Consider the unit column vectors

$$\xi_1 = (1, 0, 0, \ldots, 0)^T, \quad \xi_2 = (0, 1, 0, \ldots, 0)^T, \quad \ldots$$
$$\xi_d = (0, 0, \ldots, 0, 1)^T;$$

and identify each $\sigma_i \in \langle G \rangle_\mathbb{C}$ with $\xi_i \in \mathbb{C}^d$, so that the group action becomes $\sigma : \xi_i \mapsto K_\sigma \xi_i = \xi_j$ where $\sigma_j = \sigma \sigma_i$. Thus the matrix elements $[K_\sigma]_i^j$ have i as the column label and j as the row label.

A group-based Markov model is then obtained by taking a continuous-time Markov chain with state space $G = \{\sigma_1, \sigma_2, \ldots, \sigma_d\}$ and using the group multiplication in G to assign a rate α_σ to all substitutions $\sigma_1 \mapsto \sigma_2$ where $\sigma \sigma_1 = \sigma_2$. Following this through (as is done in detail in [32]) we are led to the formula (1) for rate matrices in any group-based model.

The regular representation is one example of the general concept of a *representation* of G on a vector space V, defined as a homomorphism $\rho : G \to GL(V)$ satisfying $\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$ for all $g_1, g_2 \in G$. A representation

is said to be *reducible* if there exists a proper subspace $U \subset V$ satisfying $\rho(g)U \subset U$, i.e. the set of matrices $\rho(G)$ send vectors in U back to U. In this case, U is called an *invariant subspace*. The representation ρ is then called *irreducible* if V does not contain any invariant subspaces.

The reader should note that the usual construction of a "group-based" model [2] stipulates that G be *abelian*. Although the construction just given using the regular representation allows for non-abelian G, we will nonetheless only consider the abelian case in this paper, because, as discussed in the introduction, it is only in the abelian case that a (linear) inversion of phylogenetic models is possible. In this case the irreducible representations of G are all one-dimensional [33], and hence the analysis reduces to computations with group characters, as is exploited in the previous approaches using Fourier analysis [9,34].

Phylogenetic tensors

We denote $[d] := \{1, 2, \ldots, d\}$ as the state space for a continuous-time Markov chain. Consider an n-taxa phylogenetic tree and a d-state phylogenetic pattern distribution $\{p_{i_1 i_2 \ldots i_n}\}_{i_1, i_2, \ldots, i_n \in [d]}$ with the interpretation that $p_{i_1 i_2 \ldots i_n}$ is the probability that the observed state at the kth leaf on the tree is i_k. As is shown in Sumner and Jarvis [35] and in more detail in Sumner et al. [27], such phylogenetic pattern distributions can be represented abstractly as tensors in the n-fold tensor product space $\otimes^n \mathbb{C}^d := \mathbb{C}^d \otimes \mathbb{C}^d \otimes \ldots \otimes \mathbb{C}^d$, as follows. If we choose $\{\xi_1, \xi_2, \ldots, \xi_d\}$ as an ordered basis for $\mathbb{C}^d$, and ordered basis $\{\xi_{i_1} \otimes \xi_{i_2} \otimes \ldots \otimes \xi_{i_d}\}_{i_1, i_2, \ldots, i_n \in [d]}$ for the tensor product space, a "phylogenetic tensor" $P \in \otimes^n \mathbb{C}^d$ is then defined as

$$P = \sum_{i_1, i_2, \ldots, i_n \in [d]} p_{i_1 i_2 \ldots i_n} \xi_{i_1} \otimes \xi_{i_2} \otimes \ldots \otimes \xi_{i_n}.$$

For readers who are unfamiliar with tensor products, it is possible to understand the general concept via the definition of the "Kronecker" product of a $n \times m$ matrix A and a $n' \times m'$ matrix B as the $nn' \times mm'$ matrix given by

$$A \otimes B := \begin{pmatrix} A_{11}B & A_{12}B & \ldots A_{1n} \\ A_{21}B & A_{22}B & \ldots A_{2n} \\ \vdots & & \vdots \\ A_{m1}B & A_{m2}B & \ldots A_{mn} \end{pmatrix}.$$

We can index the matrix $A \otimes B$ with row indicies $i_1 j_1 = 11, 12, \ldots, nn'$ and column indices $j_1 j_2 = 11, 12, \ldots, mm'$, i.e. generically $(A \otimes B)_{i_1 j_1, i_2 j_2} = A_{i_1 i_2} B_{j_1 j_2}$ and specifically $(A \otimes B)_{12,32} = A_{13}B_{22}$. This point of view is useful if one wants to write out specific matrix representations of tensors, however, in the development that follows will

focus heavily on the indexing of tensor components in the various cases discussed.

Suppose $\pi = \sum_{i \in [d]} \pi_i \xi_i \in \mathbb{C}^d$ represents the state distribution of a single taxa, i.e. π_i is the probability that a randomly chosen site in the sequence will be in state i. Now suppose a phylogenetic branching event occurs and the sequence is copied. The corresponding phylogenetic tensor $P = \sum_{i_1, i_2 \in [d]} p_{i_1 i_2} \xi_{i_1} \otimes \xi_{i_2}$ representing the joint distribution of the two-taxa just after the branching event then has the property that $p_{i_1 i_2} = \pi_{i_1}$ if $i_2 = i_1$ and is zero otherwise. Thinking in terms of tensor operations, we find that phylogenetic branching events can be generated by a linear operator $\delta : \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^d$ determined by $\delta(\pi) = P$ and defined in general using our chosen basis as

$$\delta(\xi_i) := \xi_i \otimes \xi_i, \qquad \delta(\pi) = \delta\left(\sum_i \pi_i \xi_i\right) = \sum_i \pi_i \delta(\xi_i)$$
$$= \sum_i \pi_i \xi_i \otimes \xi_i.$$

The remarkable fact for group-based models, central to the present article, is that the permutation matrices "intertwine" particularly simply with the branching operator:

$$\delta(K_\sigma \xi_i) = \delta(\xi_{\sigma(i)}) = \xi_{\sigma(i)} \otimes \xi_{\sigma(i)} = K_\sigma \otimes K_\sigma \cdot \delta(\xi_i).$$

Thus, for any rate matrix Q arising from a group-based model, we have (via the linearity of δ):

$$\delta \cdot Q = \left(-\lambda \mathbf{1} \otimes \mathbf{1} + \sum_{\epsilon \neq \sigma \in G} \alpha^\sigma K_\sigma \otimes K_\sigma\right) \cdot \delta. \quad (3)$$

We also note that, since Q can be expressed a linear combination of permutation matrices representing elements in a group G, the matrix powers $Q^2, Q^3, Q^4 \dots$ will also be expressible as linear combinations of the same permutation matrices (although precise expressions for the relevant coefficients may or may not be easily computable). Together with (3), this implies that, for any substitution matrix e^{Qt} arising from matrix exponentiation,

$$\delta \cdot e^{Qt} = e^{-\lambda} \exp\left(\sum_{\epsilon \neq \sigma \in G} \alpha^\sigma K_\sigma \otimes K_\sigma\right) \cdot \delta. \quad (4)$$

This relation shows that mathematically, and hence conceptually, "Markov evolution on a single followed by a branching event" can be replaced with "Branching event on a single taxon followed by (correlated) Markov evolution of two taxa." This equivalence is illustrated in Figure 1, and should be compared to the equivalent discussion of the "*n*-taxa process" given in [18] and [19].
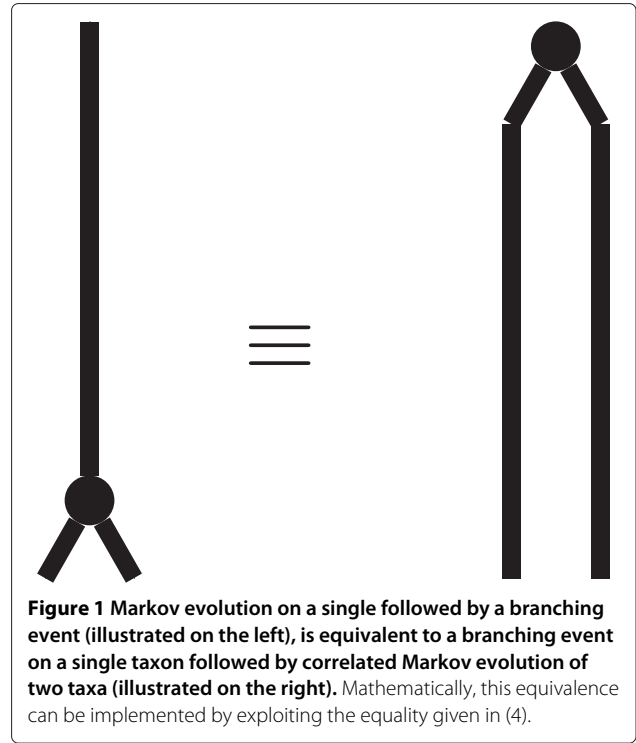

Figure 1 Markov evolution on a single followed by a branching event (illustrated on the left), is equivalent to a branching event on a single taxon followed by correlated Markov evolution of two taxa (illustrated on the right). Mathematically, this equivalence can be implemented by exploiting the equality given in (4).

In Sumner et al. [27] we showed how to generalise this intertwining action to the case of the general Markov model. Interestingly, for the general Markov model the appropriate intertwining has quite a different structure from what occurs in group-based models, and hence the simplicity of (4) is somewhat misleading in general. We refer the reader to Sumner et al. [27] for more discussion on this point.
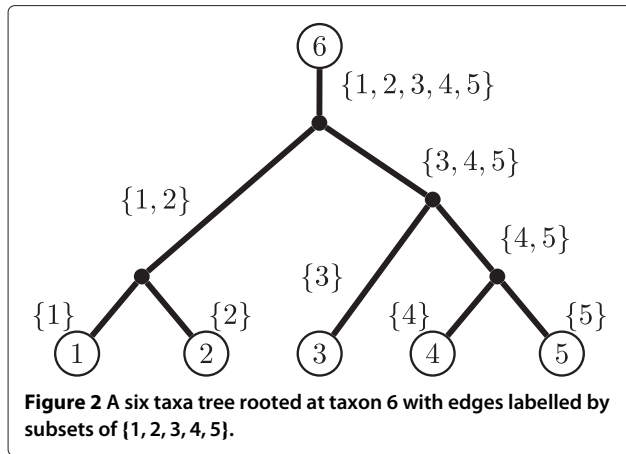
Returning to the case of group-based models, for each subset $A \subseteq [n]$, we define a linear map on $\otimes^n \mathbb{C}^d$ as the tensor product $K_\sigma^{(A)} := K_\sigma^{a_1} \otimes K_\sigma^{a_2} \otimes \dots \otimes K_\sigma^{a_n}$ where $a_i = 1$ if $i \in A$ and 0 otherwise. For example, if $n = 5$, we have

$$K_\sigma^{(\{1,2,4\})} = K_\sigma \otimes K_\sigma \otimes \mathbf{1} \otimes K_\sigma \otimes \mathbf{1}.$$

To develop a phylogenetic tensor on a tree, we root the phylogenetic tree at taxon n, and label edges by subsets $\emptyset \neq e \subseteq [n-1]$, where $i \in e$ if the path from taxon n to taxon i crosses the edge labelled by e. A five taxon tree with this labelling, is presented in Figure 2. To each edge labelled by $\emptyset \neq e \subseteq [n-1]$, we assign the rate matrix

$$Q_e := -\lambda_e \mathbf{1} + \sum_{\epsilon \neq \sigma \in G} \alpha_e^\sigma K_\sigma,$$

where each $\alpha_e^\sigma \geq 0$ is the rate of substitution for all states σ_1 to σ_2 satisfying $\sigma = \sigma_2 \sigma_1^{-1}$, and $\lambda_e = \sum_{\sigma \in G} \alpha_e^\sigma$. Each edge is then assigned substitution matrix $M_e = e^{Q_e}$, so

Figure 2 A six taxa tree rooted at taxon 6 with edges labelled by subsets of {1, 2, 3, 4, 5}.

that the time parameter for each edge is absorbed into the definition of Q_e.

Now iterating (4) multiple times, Bashford et al. [27,31] show that any phylogenetic tensor can be written as

$$P = e^{-\lambda} \exp\left(\sum_{\emptyset \neq e \subseteq [n-1], \sigma \in G} \alpha_e^\sigma K_\sigma^{(e)}\right) \cdot \delta^{n-1}\pi. \quad (5)$$

where $\lambda = \sum_{\emptyset \neq e \subseteq [n-1]} \lambda_e = \sum_{\emptyset \neq e \subseteq [n-1], \epsilon \neq \sigma \in G} \alpha_e^\sigma$, and $\delta^{n-1}\pi$ is the $d \times d \times \ldots \times d$ tensor that represents the "zero edge-length star tree" distribution on n taxa. It is this form of phylogenetic tensors that will do a lot of the heavy lifting in the discussion that follows. The reader should note that under this representation, there is no need for the edge parameters $\{\alpha_e^\sigma : \emptyset \neq e \subseteq [n-1], \sigma \in G\}$ to be chosen to be compatible with a particular tree, hence the possibilities for generalising to non-tree-like or network models, as discussed in the introduction.

The stationary distribution for group-based models is uniform (because the rate matrices are doubly stochastic). In this paper we always assume a stationary distribution, so that:

$$\pi = \tfrac{1}{d}(1, 1, \ldots, 1)^T,$$

and $\delta^{n-1}\pi$ has tensor components

$$\left[\delta^{n-1}\pi\right]_{i_1 i_2 \ldots i_n} = \begin{cases} \frac{1}{d}, & \text{if } i_1 = i_2 = \ldots = i_n, \\ 0, & \text{otherwise.} \end{cases}$$

This concludes our discussion of the tensor presentation of phylogenetic probability distributions under group-based models. It is important to note that everything discussed so far works for any group-based model, with no requirement that the underlying group G be abelian.

In what follows, we discuss the inversion of *abelian* group-based models. We present the simplest case with $G = \mathbb{Z}_2$; the $G = \mathbb{Z}_3$ case; the $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ case; the general $G = \mathbb{Z}_r$ case; and finally we discuss the case of any abelian group.

Results

The binary-symmetric case

We begin with the inversion of the so-called "binary-symmetric" model. Consider $\mathbb{C}^2$ with standard basis

$$\left\{\xi_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \xi_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}.$$

As a group-based model, the binary-symmetric model arises by taking the group

$$G := \mathbb{Z}_2 = \{0, 1\}_{+(\text{mod } 2)} \cong \langle \sigma | \sigma^2 = \epsilon \rangle,$$

with a generic rate matrix given by

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = -\mathbf{1} + K,$$

where $K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the permutation matrix representing σ in the standard basis.

Now $\rho_{\text{reg}} : \mathbb{Z}_2 \to \mathbb{M}_2(\mathbb{C})$, with $\sigma \mapsto K$, is the regular representation of $\mathbb{Z}_2$, and the character table of $\mathbb{Z}_2$ given in Table 1 is easily recognised to be the Hadamard matrix

$$h = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

As $\mathbb{Z}_2$ is an abelian group, the irreducible representations are one-dimensional.

The corresponding projection operators can be read off from the columns of the character table. That is, the operators

$$\Theta_{id} := \tfrac{1}{2}(\epsilon + \sigma), \qquad \Theta_{sgn} := \tfrac{1}{2}(\epsilon - \sigma);$$

project $\rho_{\text{reg}} = id \oplus sgn$ onto the *id* and *sgn* representations of $\mathbb{Z}_2$, respectively.

Table 1 The character table of $\mathbb{Z}_2$

	id	sgn
$[e]$	1	1
$[\sigma]$	1	-1

This observation prompts us to work in the alternative basis:

$$f_0 := \Theta_{id} \cdot \xi_0 = \Theta_{id} \cdot \xi_1 = h\xi_0 = \xi_0 + \xi_1,$$
$$f_1 := \Theta_{sgn} \cdot \xi_0 = -\Theta_{sgn} \cdot \xi_1 = h\xi_1 = \xi_0 - \xi_1.$$

In this basis the permutation matrix is diagonal:

$$\widehat{K} := hKh^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \widehat{Q} := -\mathbf{1} + \widehat{K} = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}.$$

The representation-theoretic perspective on $\widehat{K}$ is to observe that $id(\sigma) = 1$ and $sgn(\sigma) = -1$.

Referring to (5), we know that we can write a generic phylogenetic tensor as

$$P = e^{-\lambda} \exp\left(\sum_{\emptyset \neq e \subseteq [n-1]} \alpha_e K^{(e)} \right) \cdot \delta^{n-1}\pi,$$

where $\lambda = \sum_{\emptyset \neq e \subseteq [n-1]} \alpha_e$.

We index matrix and tensor indices by using $i, j, k = 0, 1 \in \mathbb{Z}_2$ and allow multiplication $\times$ in the ring of integers $\mathbb{Z}$. The Hadamard matrix then has matrix elements $[h]_i^j = (-1)^{i \times j}$ where j is the row index and i is the column index. Observe that in the diagonal basis, the permutation matrix has elements

$$\left[\widehat{K}\right]_i^j = \delta_{ij}(-1)^i.$$

Thus we have expressions such as

$$\left[\widehat{K}^{(\{2,3\})}\right]_{i_1 i_2 i_3}^{j_1 j_2 j_3} = \delta_{i_1 j_1}\delta_{i_2 j_2}\delta_{i_3 j_3}(-1)^{i_2 + i_3},$$

where $\widehat{K}^{(\{2,3\})} = \mathbf{1} \otimes \widehat{K} \otimes \widehat{K}$.

As we are dealing with tensors of arbitrary size, it is convenient to represent a string such as $i_1 i_2 \ldots i_n$ as an *ordered-bipartition* $\mu = \mu_0{:}\mu_1$ of the set $[n]$, where $\mu_0, \mu_1 \subseteq [n]$ with $j \in \mu_k$ if and only if $i_j = k$. For example we have the following equivalences:

$$00110 \equiv \{1, 2, 5\}{:}\{3, 4\}, \quad 01111 \equiv \{1\}{:}\{2, 3, 4, 5\},$$
$$10001 \equiv \{2, 3, 4\}{:}\{1, 5\}$$

and inequivalence:

$$01010 \equiv \{1, 3, 5\}{:}\{2, 4\} \neq \{2, 4\}{:}\{1, 3, 5\} \equiv 10101.$$

We then have

$$\left[\widehat{K}^{(e)}\right]_{i_1 i_2 \ldots i_n}^{j_1 j_2 \ldots j_n} = \left[\widehat{K}^{(e)}\right]_\mu^\nu = \left[\widehat{K}^{(e)}\right]_{\mu_0{:}\mu_1}^{\nu_0{:}\nu_1}$$
$$= \delta_{\mu_0 \nu_0}\delta_{\mu_1 \nu_1}(-1)^{|e \cap \mu_1|}.$$

Defining $h^{(n)} := h^{(n-1)} \otimes h$ where $h^{(1)} := h$, in the diagonal basis $\widehat{P} := h^{(n)} \cdot P$ and using our notation $h^{(n)}$ has tensor components

$$\left[h^{(n)}\right]_\mu^\nu = \left[h^{(n)}\right]_{\mu_0{:}\mu_1}^{\nu_0{:}\nu_1} = \left[h^{(n)}\right]_{i_1 i_2 \ldots i_n}^{j_1 j_2 \ldots j_n}$$
$$= (-1)^{i_1 \times j_1 + i_2 \times j_2 + \ldots + i_n \times j_n}$$
$$= (-1)^{|\mu_1 \cap \nu_1|}.$$

The zero edge-length star-tree initial distribution has tensor components

$$\left[\delta^{n-1}\pi\right]_{i_1 i_2 \ldots i_n} = \tfrac{1}{2}\delta_{i_1 i_2}\delta_{i_1 i_3} \ldots \delta_{i_1 i_n},$$

(where, although it seems we have given preference to taxon 1 in this expression, there are many ways that this distribution can be expressed using the δ_{ij}). In the diagonal basis with $\widehat{\delta^{n-1}\pi} := h^{(n)} \cdot \delta^{n-1}\pi$, we have components

$$\left[\widehat{\delta^{n-1}\pi}\right]_{i_1 i_2 \ldots i_n}$$
$$= \tfrac{1}{2}\sum_{j_1, j_2, \ldots, j_n}(-1)^{i_1 \times j_1 + i_2 \times j_2 + \ldots + i_n \times j_n}\delta_{j_1 j_2}\delta_{j_1 j_3} \ldots \delta_{j_1 j_n}$$
$$= \tfrac{1}{2}\sum_{j_1}(-1)^{(i_1 + i_2 + \ldots + i_n) \times j_1} = \tfrac{1}{2}\left(1 + (-1)^{i_1 + i_2 + \ldots + i_n}\right),$$

which is exactly the statement

$$\left[\widehat{\delta^{n-1}\pi}\right]_\mu = \left[\widehat{\delta^{n-1}\pi}\right]_{\mu_0{:}\mu_1} = \tfrac{1}{2}\left(1 + (-1)^{|\mu_1|}\right).$$

Since $\widehat{K}$ is diagonal in the transformed basis, we can conclude that

$$\left[\widehat{P}\right]_\mu = \left[\widehat{P}\right]_{\mu_0{:}\mu_1}$$
$$= e^{-\lambda} \exp\left(\sum_{\emptyset \neq e \subseteq [2,n]} \alpha_e \left[\widehat{K}^{(e)}\right]_{\mu_0{:}\mu_1}^{\mu_0{:}\mu_1} \right) \tfrac{1}{2}\left(1 + (-1)^{|\mu_1|}\right).$$

Of course many of these tensor components will be zero and we would like to ignore these.

Take $u = u_0{:}u_1$ as an ordered bipartition of the reduced set $[n-1]$, so that $u \equiv i_1 i_2 \ldots i_{n-1}$ where $j \in u_k$ if and only if $i_j = k$, and define

$$\gamma(u) = \begin{cases} 0, & \text{if } |u_1| \text{ is even}, \\ 1, & \text{if } |u_1| \text{ is odd}; \end{cases}$$
$$= 2 - (0|u_0| + 1|u_1|) \pmod 2,$$

and interpret $u \cdot \gamma(u)$ as a string: $u \cdot \gamma(u) = i_1 i_2 \ldots i_{n-1}\gamma(u)$.

If we make the definitions

$$\mathcal{P}_u := \left[\widehat{P}\right]_{u \cdot \gamma(u)}, \qquad \eta_u := \tfrac{1}{2}\sum_{\emptyset \neq e \subseteq [n-1]} \alpha_e \left[\widehat{K}^{(e)}\right]_{u \cdot \gamma(u)}^{u \cdot \gamma(u)},$$

then we can write the non-zero components as

$$\mathcal{P}_u = e^{-\lambda} \exp(\eta_u),$$

with inverses

$$\eta_u = \ln(\mathcal{P}_u) + \lambda. \tag{6}$$

This is the first part of the inversion.

We would like to go further and actually recover the individual edge weights α_e. To do this we define the (square) $2^{n-1} \times 2^{n-1}$ matrix F with components

$$[F]_u^e := \left[\widehat{K}^{(e)}\right]_{u\cdot\gamma(u)}^{u\cdot\gamma(u)} = (-1)^{|e\cap u|} = \left[h^{(n-1)}\right]_u^e,$$

with e a subset and u an ordered-bipartition of $[n-1]$. As $\left(h^{(n-1)}\right)^2 = \frac{1}{2^{n-1}}\mathbf{1}$, we see that F provides its own inverse F^{-1} with components

$$[F^{-1}]_e^u := \frac{1}{2^{n-1}}[F]_u^e.$$

Defining the column vectors $\vec{\alpha} = \{\alpha_e\}$ and $\vec{\eta} = \{\eta_u\}$, we can write the matrix equations

$$\vec{\eta} = F\vec{\alpha}, \qquad \vec{\alpha} = F^{-1}\vec{\eta}.$$

Together with the first part of the inversion (6), these equations give a one-one map between pattern probabilities and edge weights for the binary-symmetric model.

Inversion of the $\mathbb{Z}_3$ model

Taking confidence from the previous case we now discuss the inversion of the group-based phylogenetic model with $G = \mathbb{Z}_3$. We take

$$\mathbb{Z}_3 = \{0, 1, 2\}_{+(\text{mod } 3)} \cong \langle \sigma | \sigma^3 = \epsilon \rangle,$$

and, by analogy to the $\mathbb{Z}_2$ case, index tensors with indices $i, j = 0, 1, 2$ and allow multiplication $\times$ by extending $\mathbb{Z}_3$ to the ring $\mathbb{F}_3 = \{0, 1, 2\}_{+,\times(\text{mod } 3)}$.

In this case a generic rate matrix is given by

$$Q = \begin{pmatrix} -(\alpha + \beta) & \beta & \alpha \\ \alpha & -(\alpha + \beta) & \beta \\ \beta & \alpha & -(\alpha + \beta) \end{pmatrix}$$

$$= -(\alpha + \beta)\mathbf{1} + \alpha K_1 + \beta K_2,$$

where

$$K_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad K_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

are the matrices representing the permutations $\sigma \cong$ (123) and $\sigma^2 \cong$ (132) under the regular representation, respectively.

We define $\omega = e^{2\pi i/3}$, and present the character table of $\mathbb{Z}_3$ in Table 2. The decomposition of the regular representation is $\rho_{\text{reg}} = id \oplus \omega \oplus \omega^2$, and the columns of the character table give the projection operators onto the (one-dimensional) irreducible subspaces:

$$\Theta_{id} := \tfrac{1}{3}\left(\epsilon + \sigma + \sigma^2\right),$$
$$\Theta_{\omega} := \tfrac{1}{3}\left(\epsilon + \omega\sigma + \omega^2\sigma^2\right),$$
$$\Theta_{\omega^2} := \tfrac{1}{3}\left(\epsilon + \omega^2\sigma + \omega\sigma^2\right).$$

Therefore, the matrix

$$f = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix},$$

diagonalizes the generic rate matrix for this model:

$$\widehat{Q} = fQf^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha\omega + \beta\omega^2 & 0 \\ 0 & 0 & \alpha\omega^2 + \beta\omega \end{pmatrix},$$

or, equivalently,

$$\widehat{K}_1 = fK_1f^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix},$$

$$\widehat{K}_2 = fK_2f^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix}.$$

We recall our basic result (5) that for group-based models, a generic phylogenetic tensor can be expressed as

$$P = e^{-\lambda} \exp\left(\sum_{\emptyset \neq e \subseteq [n-1]} \left(\alpha_e K_1^{(e)} + \beta_e K_2^{(e)}\right)\right) \cdot \delta^{n-1}\pi,$$

where $\lambda = \sum_{\emptyset \neq e \subseteq [n-1]} (\alpha_e + \beta_e)$. We take the stationary distribution as initial distribution, so $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$.

The matrix elements of f can be expressed as $[f]_i^j = \omega^{i \times j}$, where we extend $i, j \in \mathbb{Z}_3$ to include multiplication $\times$ from the ring of integers $\mathbb{Z}$. Similarly,

$$\left[\widehat{K}_1\right]_i^j = \delta_{ij}\omega^i, \qquad \left[\widehat{K}_2\right]_i^j = \delta_{ij}(\omega^2)^i.$$

Table 2 The character table of $\mathbb{Z}_3$

	id	ω	ω^2
$[e]$	1	1	1
$[\sigma]$	1	ω	ω^2
$[\sigma^2]$	1	ω^2	ω

More generally, tensorial components can be expressed as

$$\left[\mathbf{1}\otimes\widehat{K_1}\otimes\widehat{K_1}\right]_{i_1i_2i_3}^{j_1j_2j_3}=\delta_{i_1j_1}\delta_{i_2j_2}\delta_{i_3j_3}\omega^{i_2+i_3}.$$

We represent a string $i_1i_2\ldots i_n$ as an *ordered-tripartition*, $i_1i_2\ldots i_n\equiv\mu=\mu_0{:}\mu_1{:}\mu_2$, of the set $[n]$, where $j\in\mu_k$ if and only if $i_j=k$. For example, if we take $n=5$, we have:

$$00000\equiv\{1,2,3,4,5\}{:}\emptyset{:}\emptyset,\quad 20120\equiv\{2,5\}{:}\{3\}{:}\{1,4\},$$

$$01122\equiv\{1\}{:}\{2,3\}{:}\{4,5\}.$$

Taking $n=3$, we have

$$\left[\widehat{K_1}^{(\{2,3\})}\right]_\mu^\nu=\left[\mathbf{1}\otimes\widehat{K_1}\otimes\widehat{K_1}\right]_\mu^\nu=\left[\mathbf{1}\otimes\widehat{K_1}\otimes\widehat{K_1}\right]_{\mu_0{:}\mu_1{:}\mu_2}^{\nu_0{:}\nu_1{:}\nu_2}$$
$$=\delta_{\mu\nu}\omega^{|\mu_1\cap\{2,3\}|+2|\mu_2\cap\{2,3\}|},$$

and in general:

$$\left[\widehat{K_1}^{(e)}\right]_\mu^\nu=\delta_{\mu\nu}\omega^{|e\cap\mu_1|+2|e\cap\mu_2|},$$
$$\left[\widehat{K_2}^{(e)}\right]_\mu^\nu=\delta_{\mu\nu}\omega^{|e\cap\mu_2|+2|e\cap\mu_1|}.$$

Taking the uniform distribution as initial distribution, the initial star-tree distribution can be written as

$$\left[\delta^{n-1}\pi\right]_{i_1i_2\ldots i_n}=\tfrac13\delta_{i_1i_2}\delta_{i_1i_3}\ldots\delta_{i_1i_n}.$$

Defining $f^{(n)}=f^{(n-1)}\otimes f$ where $f^{(1)}=f$, we have

$$\left[f^{(n)}\right]_\mu^\nu=\left[f^{(n)}\right]_{i_1i_2\ldots i_n}^{j_1j_2\ldots j_n}=\left[f\right]_{i_1}^{j_1}\left[f\right]_{i_2}^{j_2}\cdots\left[f\right]_{i_n}^{j_n}$$
$$=\omega^{i_1\times j_1+i_2\times j_2+\ldots+i_n\times j_n},$$

and in the transformed basis, where $\widehat{\delta^{n-1}\pi}:=f^{(n)}\cdot\delta^{n-1}\pi$, we have

$$\left[\widehat{\delta^{n-1}\pi}\right]_{i_1i_2\ldots i_n}=\tfrac13\sum_{j_1,j_2,\ldots,j_n}\omega^{i_1\times j_1+i_2\times j_2+\ldots+i_n\times j_n}$$
$$\times\,\delta_{j_1j_2}\delta_{j_1j_3}\ldots\delta_{j_1j_n}$$
$$=\tfrac13\sum_{j_1}\omega^{j_1\times(i_1+i_2+\ldots+i_n)}$$
$$=\tfrac13\left(1+\omega^{i_1+i_2+\ldots+i_n}+(\omega^2)^{i_1+i_2+\ldots+i_n}\right).$$

Indexing by ordered-tripartitions, we conclude that

$$\left[\widehat{\delta^{n-1}\pi}\right]_\mu=\tfrac13\left(1+\omega^{i_1+i_2+\ldots+i_n}+(\omega^2)^{i_1+i_2+\ldots+i_n}\right)$$
$$=\tfrac13\left(1+\omega^{|\mu_1|+2|\mu_2|}+(\omega^2)^{|\mu_1|+2|\mu_2|}\right).$$

Now suppose $|\mu_1|+2|\mu_2|=0\ (\mathrm{mod}\ 3)$, then

$$\left[\widehat{\delta^{n-1}\pi}\right]_\mu=\tfrac13\left(1+1+1\right)=1.$$

If $|\mu_1|+2|\mu_2|=1\ (\mathrm{mod}\ 3)$, then

$$\left[\widehat{\delta^{n-1}\pi}\right]_\mu=\tfrac13\left(1+\omega+\omega^2\right)=0,$$

and if $|\mu_1|+2|\mu_2|=2\ (\mathrm{mod}\ 3)$, then

$$\left[\widehat{\delta^{n-1}\pi}\right]_\mu=\tfrac13\left(1+\omega^2+\omega\right)=0.$$

Thus we have found a basis where all the elements of the initial star-tree tensor are zero *unless* the tripartion μ satisfies $|\mu_1|+2|\mu_2|=0\ (\mathrm{mod}\ 3)$. Crucially, this statement also holds for the phylogenetic tensor $\widehat{P}$ because in this basis the rate matrices of this model are diagonal:

$$\left[\widehat{P}\right]_\mu=\left[\widehat{P}\right]_{\mu_0{:}\mu_1{:}\mu_2}$$
$$=e^{-\lambda}\exp\left(\tfrac12\sum_{\emptyset\neq e\subseteq[n-1]}\left[\alpha_eK_1^{(e)}+\beta_eK_2^{(e)}\right]_{\mu_0{:}\mu_1{:}\mu_2}^{\mu_0{:}\mu_1{:}\mu_2}\right)$$
$$\times\tfrac13\left(1+\omega^{1|\mu_1|}+\omega^{2|\mu_2|}\right).$$

We deal with this condition on μ by taking $u=u_0{:}u_1{:}u_2$ as an ordered-tripartion of the reduced set $[n-1]$ and setting $\mu=u\cdot\gamma(u)$ (considered as the concatenation of strings) where

$$\gamma(u)=\begin{cases}0,&\text{if }|u_1|+2|u_2|=0\\1,&\text{if }|u_1|+2|u_2|=2\\2;&\text{if }|u_1|+2|u_2|=1\end{cases}$$
$$=3-(0|u_0|+1|u_1|+2|u_2|)\ (\mathrm{mod}\ 3).$$

If we make the definitions

$$\mathcal{P}_u:=\left[\widehat{P}\right]_{u\cdot\gamma(u)},$$
$$\eta_u:=\left[\sum_{\emptyset\neq e\subseteq[n-1]}\alpha_eK_1^{(e)}+\beta_eK_2^{(e)}\right]_{u\cdot\gamma(u)}^{u\cdot\gamma(u)},$$

we then have the first part of the inversion

$$\mathcal{P}_u=e^{-\lambda}\exp\left(\eta_u\right),\qquad\eta_u=\ln\left(\mathcal{P}_u\right)+\lambda.\qquad(7)$$

As in the $\mathbb{Z}_2$ case, we would like to use η_u to recover the rate parameters α_e,β_e for all $\emptyset\neq e\subseteq[n-1]$ and thus complete the full inversion for this model. Of course, it is little bit more difficult this time.

Recall that $\mu=\mu_0{:}\mu_1{:}\mu_2$ with $\mu_i\subseteq[n]$, whereas $u=u_0{:}u_1{:}u_2$ with $u_i\subseteq[n-1]$, and $\emptyset\neq e\subseteq[n-1]$. Considering

$$\left[K_1^{(e)}\right]_\mu^\mu=\omega^{|e\cap\mu_1|+2|e\cap\mu_2|},$$

it follows that

$$\left[K_1^{(e)}\right]_{u\cdot\gamma(u)}^{u\cdot\gamma(u)} = \omega^{|e\cap u_1|+2|e\cap u_2|},$$

and similarly

$$\left[K_2^{(e)}\right]_{u\cdot\gamma(u)}^{u\cdot\gamma(u)} = \omega^{|e\cap u_2|+2|e\cap u_1|}.$$

We make the observation that

$$[F_1]_u^e := \left[f^{(n-1)}\right]_{u_0:u_1:u_2}^{e^c:e:\emptyset} = \omega^{|u_1\cap e|+2|u_2\cap e|} = \left[K_\alpha^{(e)}\right]_{u\cdot\gamma(u)}^{u\cdot\gamma(u)},$$

and

$$[F_2]_u^e := \left[f^{(n-1)}\right]_{u_0:u_1:u_2}^{e^c:\emptyset:e} = \omega^{|u_2\cap e|+2|u_1\cap e|} = \left[K_\beta^{(e)}\right]_{u\cdot\gamma(u)}^{u\cdot\gamma(u)},$$

where F_1 and F_2 are $2^{n-1} \times 3^{n-1}$ matrices.

Thus we may write

$$\eta_u = \sum_{\emptyset \neq e \subseteq [n-1]} \alpha_e[F_1]_u^e + \beta_e[F_2]_u^e.$$

Defining the column vectors $\vec{\alpha} = \{\alpha_e\}, \vec{\beta} = \{\beta_e\}$ and $\vec{\eta} = \{\eta_u\}$, we can write

$$\vec{\eta} = F_1\vec{\alpha} + F_2\vec{\beta},$$

and define two $3^{n-1} \times 2^{n-1}$ matrices G_1 and G_2 as

$$[G_1]_e^u := \left[f^{-1(n-1)}\right]_{e^c:e:\emptyset}^u, \qquad [G_2]_e^u := \left[f^{-1(n-1)}\right]_{e^c:\emptyset:e}^u,$$

where

$$f^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix},$$

with $ff^{-1} = \mathbf{1}$.

Considering that

$$\sum_v \left[f^{-1(n-1)}\right]_u^v \left[f^{(n-1)}\right]_v^w = \delta_{uw},$$

for all ordered-triparitions u, w of $[n-1]$, we have the matrix products

$$G_1F_1 = \mathbf{1}, \; G_1F_2 = 0,$$
$$G_2F_2 = \mathbf{1}, \; G_2F_1 = 0.$$

Thus the second part of the inversion for this model is

$$\vec{\alpha} = G_1\vec{\eta}, \qquad \vec{\beta} = G_2\vec{\eta}.$$

Together with (7), these equations give a one-one map between pattern probabilities and edge weights for the group-based model with $G = \mathbb{Z}_3$.

Inversion of the K3ST model

We now consider the K3ST model [36] which occurs as the group-based model with

$$G = \mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0,0), (0,1), (1,0), (1,1)\}_{+(\text{mod } 2)}$$
$$\cong \langle (12)(34), (13)(24) \rangle.$$

In this model a generic rate matrix is given by

$$Q = -(\alpha + \beta + \gamma)\mathbf{1} + \alpha K_{01} + \beta K_{10} + \gamma K_{11},$$

where

$$K_{01} = \mathbf{1} \otimes K = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$K_{10} = K \otimes \mathbf{1} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad (8)$$

$$K_{11} = K \otimes K = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

We already know that the 2×2 Hadamard matrix h diagonalizes K, so we see immediately that $H = h \otimes h$ diagonalizes this model:

$$\widehat{K}_{01} := HK_{01}H^{-1} = \mathbf{1} \otimes hKh^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

$$\widehat{K}_{10} := HK_{10}H^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

$$\widehat{K}_{11} := HK_{11}H^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Of course H is the character table of $\mathbb{Z}_2 \times \mathbb{Z}_2$ and the permutation matrices (8), together with $K_{00} := \mathbf{1}$, give the regular representation $\rho_{\text{reg}} \cong id \otimes id \oplus id \otimes sgn \oplus sgn \otimes id \oplus sgn \otimes sgn$, where we recall the basic result that the tensor product of two irreducible representations of a group G gives an irreducible representation of $G \times G$.

Simplifying notation, for this model we index tensors with indices given as pairs: $i, j = 00, 01, 10, 11 \in \mathbb{Z}_2 \times \mathbb{Z}_2$; and we express the individual parts using lower case Roman characters. For example, we write $i := ab = 01$, with $a = 0$ and $b = 1$. This gives matrix elements:

$$\left[\widehat{K}_{01}\right]^{cd}_{ab} = \delta_{ac}\delta_{bd}(-1)^b, \quad \left[\widehat{K}_{10}\right]^{cd}_{ab} = \delta_{ac}\delta_{bd}(-1)^a,$$
$$\left[\widehat{K}_{11}\right]^{cd}_{ab} = \delta_{ac}\delta_{bd}(-1)^{a+b};$$

and more complicated tensor products such as

$$\left[\widehat{K}_{01} \otimes \widehat{K}_{01} \otimes \mathbf{1}\right]^{c_1 d_1 c_2 d_2 c_3 d_3}_{a_1 b_1 a_2 b_2 a_3 b_3}$$
$$= \delta_{a_1 c_1}\delta_{b_1 d_1}\delta_{a_2 c_2}\delta_{b_2 d_2}\delta_{a_3 c_3}\delta_{b_3 d_3}(-1)^{b_1 + b_2}.$$

Again we interpret strings such as $\mu \equiv a_1 a_2 \ldots a_n$ and $\nu \equiv b_1 b_2 \ldots b_n$ as ordered-bipartitions $\mu = \mu_0 : \mu_1$ and $\nu = \nu_0 : \nu_1$ of the set $[n]$. We can then write matrix elements of tensor products as

$$\left[\widehat{K}^{(e)}_{01}\right]^{\mu', \nu'}_{\mu, \nu} = \delta_{\mu\mu'}\delta_{\nu\nu'}(-1)^{|e \cap \nu_1|},$$
$$\left[\widehat{K}^{(e)}_{10}\right]^{\mu', \nu'}_{\mu, \nu} = \delta_{\mu\mu'}\delta_{\nu\nu'}(-1)^{|e \cap \mu_1|},$$
$$\left[\widehat{K}^{(e)}_{11}\right]^{\mu', \nu'}_{\mu, \nu} = \delta_{\mu\mu'}\delta_{\nu\nu'}(-1)^{|e \cap \mu_1| + |e \cap \nu_1|}.$$

Taking the stationary distribution $\pi = \frac{1}{4}(1, 1, 1, 1)^T$ as initial distribution, the zero edge-length star-tree distribution is given by

$$\left[\delta^{n-1}\pi\right]_{i_1 i_2 \ldots i_n} = \frac{1}{4}\delta_{i_1 i_2}\delta_{i_1 i_3}\ldots\delta_{i_1 i_n},$$

which in the finer index representation is

$$\left[\delta^{n-1}\pi\right]_{a_1 b_1 a_2 b_2 \ldots a_n b_n}$$
$$= \frac{1}{4}\delta_{a_1 a_2}\delta_{a_1 a_3}\ldots\delta_{a_1 a_n}\delta_{b_1 b_2}\delta_{b_1 b_3}\ldots\delta_{b_1 b_n}.$$

Recall that elements of the Hadamard matrix can be written as $[h]^a_b = (-1)^{a \times b}$, where $a, b \in \mathbb{Z}_2$ and we allow multiplication $\times$ by extending to the ring of integers $\mathbb{Z}$. In the transformed basis, we have

$$\left[\widehat{\delta^{n-1}\pi}\right]_{a_1 b_1 a_2 b_2 \ldots a_n b_n} = \left[\widehat{\delta^{n-1}\pi}\right]_{\mu, \nu}$$
$$= \frac{1}{4}\sum_{c_1, c_2, \ldots, c_n}^{d_1, d_2, \ldots, d_n} [h]^{a_1}_{c_1}[h]^{a_2}_{c_2}\ldots[h]^{a_n}_{c_n}[h]^{b_1}_{d_1}[h]^{b_2}_{d_2}\ldots$$
$$\times [h]^{b_n}_{d_n}\delta_{a_1 a_2}\delta_{a_1 a_3}\ldots\delta_{a_1 a_n}\delta_{b_1 b_2}\delta_{b_1 b_3}\ldots\delta_{b_1 b_n}$$
$$= \frac{1}{4}\sum_{c_1, d_1}(-1)^{(a_1 + a_2 + \ldots a_n) \times c_1 + (b_1 + b_2 + \ldots + b_n) \times d_1}$$
$$= \frac{1}{4}\left(1 + (-1)^{a_1 + \ldots + a_n} + (-1)^{b_1 + \ldots + b_n}\right.$$
$$\left. + (-1)^{a_1 + \ldots + a_n + b_1 + \ldots + b_n}\right)$$
$$= \begin{cases} 0, & \text{if either } |\mu_1| \text{ or } |\nu_1| \text{ is odd;} \\ 1, & \text{if } |\mu_1| \text{ and } |\nu_1| \text{ are both even.} \end{cases}$$

We recall (5), so under this model we can express a generic phylogenetic tensor as

$$P = e^{-\lambda}\exp\left(\sum_{\emptyset \neq e \subseteq [n-1]} \alpha_e K^{(e)}_{01} + \beta_e K^{(e)}_{10} + \gamma_e K^{(e)}_{11}\right) \cdot \delta^{n-1}\pi.$$

To exclude the vanishing components we define, for all ordered bipartitions $u = u_0 : u_1$ of the reduced set $[n-1]$,

$$\gamma(u) = \begin{cases} 0, & \text{if } |u_1| \text{ is even,} \\ 1, & \text{if } |u_1| \text{ is odd;} \end{cases}$$
$$= 2 - (0|u_0| + 1|u_1|)(\text{mod } 2),$$

and interpet $u \cdot \gamma(u)$ as the string $u \cdot \gamma(u) = a_1 a_2 \ldots a_{n-1}\gamma(u)$. Then, for each pair u, v of ordered-bipartitions of $[n-1]$, we define

$$\eta_{u,v} := \left[\sum_{\emptyset \neq e \subseteq [n-1]} \alpha_e K^{(e)}_{01} + \beta_e K^{(e)}_{10} + \gamma_e K^{(e)}_{11}\right]^{u \cdot \gamma(u), v \cdot \gamma(v)}_{u \cdot \gamma(u), v \cdot \gamma(v)},$$

and

$$\mathcal{P}_{u,v} := [P]_{u \cdot \gamma(u), v \cdot \gamma(v)},$$

This gives the inversion

$$\mathcal{P}_{u,v} = e^{-\lambda}\exp\left(\eta_{u,v}\right), \qquad \eta_{u,v} = \lambda + \ln\left(\mathcal{P}_{u,v}\right).$$

Consider the $2^n \times 2^{n-1}$ rectangular matrices F_{01}, F_{10} and F_{11} with components

$$[F_{01}]^e_{u,v} = \left[K^{(e)}_{01}\right]^{u,v}_{u,v} = (-1)^{|e \cap v_1|},$$
$$[F_{10}]^e_{u,v} = \left[K^{(e)}_{10}\right]^{u,v}_{u,v} = (-1)^{|e \cap u_1|},$$
$$[F_{11}]^e_{u,v} = \left[K^{(e)}_{11}\right]^{u,v}_{u,v} = (-1)^{|e \cap u_1| + |e \cap v_1|};$$

where $e \subseteq [n-1]$ and $u = u_0 : u_1$ and $v = v_0 : v_1$ are ordered-bipartitions of $[n-1]$. If we define the column vector $\vec{\eta} := \{\eta_{u,v}\}$ indexed by pairs of ordered-bipartitions and the column vectors $\vec{\alpha} := \{\alpha_e\}, \vec{\beta} := \{\alpha_e\}$ and $\vec{\gamma} := \{\alpha_e\}$ indexed by subsets of $[n-1]$, we then have the matrix equation

$$\vec{\eta} = F_{01}\vec{\alpha} + F_{10}\vec{\beta} + F_{11}\vec{\gamma}.$$

Writing $H^{(n)} = H^{(n-1)} \otimes H$ with $H^{(1)} = H$, we note that

$$[F_{01}]^e_{u,v} = \left[H^{(n-1)}\right]^{\emptyset, e}_{u,v}, \qquad [F_{10}]^e_{u,v} = \left[H^{(n-1)}\right]^{e, \emptyset}_{u,v},$$
$$[F_{11}]^e_{u,v} = \left[H^{(n-1)}\right]^{e, e}_{u,v};$$

and define the $2^{n-1} \times 2^n$ rectangular matrices G_{01}, G_{10} and G_{11} as

$$[G_{01}]_e^{u,v} = \left[H^{-1\,(n-1)}\right]_{\emptyset,e}^{u,v}, \qquad [G_{10}]_e^{u,v} = \left[H^{-1\,(n-1)}\right]_{e,\emptyset}^{u,v},$$

$$[G_{11}]_e^{u,v} = \left[H^{-1\,(n-1)}\right]_{e,e}^{u,v}.$$

Noting that

$$\sum_{w,x} \left[H^{-1\,(n-1)}\right]_{u,v}^{w,x}\left[H^{(n-1)}\right]_{w,x}^{y,z} = \delta_{u,y}\delta_{v,z},$$

for all u, v, y, z ordered-bipartitions of $[n-1]$, we then have the matrix identities

$$G_{01}F_{01} = \mathbf{1}, \qquad G_{10}F_{10} = \mathbf{1}, \qquad G_{11}F_{11} = \mathbf{1},$$

and

$$G_{01}F_{10} = 0 = G_{01}F_{11} = G_\beta F_{11} = G_{10}F_{01}$$
$$= G_{11}F_{01} = G_{11}F_{10}.$$

Writing

$$\vec{\alpha} = G_{01}\vec{\eta}, \qquad \vec{\beta} = G_{10}\vec{\eta}, \qquad \vec{\gamma} = G_{11}\vec{\eta},$$

completes the inversion for the K3ST model.

Inversion of the $\mathbb{Z}_r$ model

We now consider the group based model for $\mathbb{Z}_r = \{0, 1, 2, \ldots (r-1)\}_{+(\mathrm{mod}\ r)} \cong \langle \sigma : \sigma^r = e \rangle$. For this model the generic rate matrix has the form

$$Q = -\lambda\mathbf{1} + \sum_{i=1}^{r} \alpha^i K_{\sigma^i},$$

where $\lambda = \sum_{i=1}^{r} \alpha^i$ and

$$K_\sigma = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix},$$

so that $K_{\sigma^i} = K_\sigma^i$.

Defining $\omega = e^{2\pi i/r}$, we have $\omega^r = 1$ and $1 + \omega + \omega^2 + \ldots + \omega^{r-1} = 0$ and $[f]_i^j = \omega^{ij}$ where $i, j = 0, 1, 2, \ldots, r-1$. Of course, f is the character table of $\mathbb{Z}_r$ and $[f^{-1}]_j^i = \frac{1}{r}\omega^{-ij}$.

Lemma 1.

$$\sum_v [f \otimes f \otimes \ldots \otimes f]_\mu^v [f^{-1} \otimes f^{-1} \otimes \ldots \otimes f^{-1}]_v^{\mu'} = \delta_{\mu\mu'},$$

where μ, v, μ' are ordered-r-partitions of the set $[n]$ defined by the strings $i_1 i_2 \ldots i_n$, $j_1 j_2 \ldots j_n$ and $k_1 k_2 \ldots k_n$, respectively.

Proof. The result is obvious by the definition of tensor product. However, explicitly we have

$$\sum_v [f \otimes f \otimes \ldots \otimes f]_\mu^v [f^{-1} \otimes f^{-1} \otimes \ldots \otimes f^{-1}]_v^{\mu'}$$

$$= \frac{1}{r^n} \sum_{0 \leq j_1, j_2, \ldots, j_{r-1} \leq (r-1)}$$
$$\times \omega^{i_1 j_1 + i_2 j_2 + \ldots i_{r-1}j_{r-1}} \omega^{-(j_1 k_1 + j_2 k_2 + \ldots + j_n k_n)}$$

$$= \frac{1}{r^n} \sum_{0 \leq j_1, j_2, \ldots, j_{r-1} \leq (r-1)}$$
$$\times \omega^{j_1(i_1-k_1) + j_2(i_2-k_2) + \ldots + j_n(i_n-k_n)}$$

which clearly equals 1 if $i_\ell - k_\ell = 0$ for all ℓ, and, by repeatedly applying $1 + \omega + \omega^2 + \ldots + \omega^{r-1} = 0$, equals 0 otherwise. □

The regular representation contains exactly one copy of every irreducible representation and the irreducible representations of $\mathbb{Z}_r$ are given by the powers of ω:

$$\rho_i : \begin{array}{l} \mathbb{Z}_r \to \mathbb{C} \\ \sigma \mapsto \omega^i \end{array}.$$

Thus the change of basis $K_{\sigma^i} \mapsto \widehat{K}_{\sigma^i} = fK_{\sigma^i}f^{-1}$ will give diagonal matrices $\widehat{K}_{\sigma^i}$. Additionally,

Lemma 2. *In the diagonal basis, the matrices $\widehat{K}_{\sigma^i} := fK_{\sigma^i}f^{-1}$ have matrix elements given by $\left[\widehat{K}_{\sigma^s}\right]_i^j = \omega^{is}\delta_{ij}$.*

Proof. Consider the matrix elements $[K_{\sigma^s}]_i^j = \delta_{i\sigma^s(j)}$. Thus

$$\left[fK_{\sigma^s}f^{-1}\right]_j^i = \sum_{k,l} \omega^{ik}\delta_{k\sigma^s(l)}\omega^{-lj} = \sum_l \omega^{i\sigma^s(l)-lj}$$

$$= \sum_l \omega^{i(l+s)-lj}$$

$$= \omega^{is} \sum_l \omega^{l(i-j)} = \omega^{is}\delta_{ij},$$

where we have used $\omega^{\sigma^s(m)} = \omega^{m+s}$. □

Now

$$\left[\delta^{n-1}\pi\right]_{i_1 i_2 \ldots i_n} = \frac{1}{r}\delta_{i_1 i_2}\delta_{i_1 i_3}\ldots\delta_{i_1 i_n},$$

and

$$\left[\widehat{\delta^{n-1}\pi}\right]_{i_1 i_2 \ldots i_n} = \frac{1}{r}\sum_{j_1, j_2, \ldots, j_n}\omega^{i_1 j_1 + i_2 j_2 + \ldots + i_n j_n}$$
$$\times \delta_{j_1 j_2}\delta_{j_1 j_3}\ldots\delta_{j_1 j_n}$$

$$= \frac{1}{r}\sum_{j_1}\omega^{j_1(i_1 + i_2 + \ldots + i_n)}$$

$$= \begin{cases} 1 & \text{if } i_1 + i_2 + \ldots + i_n = 0(\mathrm{mod\ r}) \\ 0, & \text{otherwise.} \end{cases}$$

Translating this result using the ordered-r-partitions for indices, we have

Lemma 3. *In the diagonal basis, the uniform initial distribution on the star tree has components*

$$\left[\widehat{\delta^{n-1}\pi}\right]_\mu$$
$$= \begin{cases} 1 \ if\ 0|\mu_0|+1|\mu_1|+2|\mu_2|+\ldots+(r-1)|\mu_{r-1}| = 0(mod\ r) \\ 0,\ otherwise. \end{cases},$$

where $\mu = \mu_0{:}\mu_1{:}\mu_2{:}\ldots{:}\mu_{r-1}$ is an ordered-r-partition of the set $[n]$.

Again recall that for this model a generic phylogenetic tensor can be written as

$$P = e^{-\lambda}\exp\left(\sum_{\emptyset \neq e \subseteq [n-1], s \in [r-1]} \alpha_e^s K_{\sigma^s}^{(e)}\right)\delta^{n-1}\pi,$$

where $\pi = \frac{1}{r}(1,1,\ldots,1)^T$. In the diagonal basis $\widehat{P} := f^{(n)} \cdot P$ and as a consequence of Lemma 3 $\widehat{P}$ will have many vanishing components. To avoid these we take $u = u_0{:}u_1{:}u_2{:}\ldots{:}u_{r-1}$ as an *ordered-r-partition* of $[n-1]$ and set

$$\gamma(u) = r - (0|u_0| + 1|u_1| + 2|u_2| + \ldots$$
$$+(r-1)|u_{r-1}|) \pmod{r}.$$

If we define $\mathcal{P}_u := \left[\widehat{P}\right]_{u \cdot \gamma(u)}$ and

$$\eta_u := \left[\sum_{\emptyset \neq e \subseteq [n-1], s \in [r-1]} \alpha_e^s \widehat{K}_{\sigma^s}^{(e)}\right]_{u \cdot \gamma(u)}^{u \cdot \gamma(u)},$$

we then have the first part of the inversion for the $\mathbb{Z}_r$ model:

$$\mathcal{P}_u = e^{-\lambda}\exp(\eta_u),$$
$$\eta_u = \ln(\mathcal{P}_u) + \lambda.$$

For each $i \in [r-1]$, we define the column vectors $\vec{\alpha}_i := \left\{\alpha_e^i\right\}_{\emptyset \neq e \subseteq [n-1]}$, and, for each $\emptyset \neq e \subseteq [n-1]$ and u an ordered-$(r-1)$-partition of $[n-1]$, we define the rectangular $r^{n-1} \times 2^{n-1}$ matrices

$$[F_1]_u^e := \left[K_\sigma^{(e)}\right]_{u \cdot \gamma(u)}^{u \cdot \gamma(u)}, \quad [F_2]_u^e := \left[K_{\sigma^2}^{(e)}\right]_{u \cdot \gamma(u)}^{u \cdot \gamma(u)}, \ldots$$
$$[F_{r-1}]_u^e := \left[K_{\sigma^{r-1}}^{(e)}\right]_{u \cdot \gamma(u)}^{u \cdot \gamma(u)},$$

so we have the vector equation

$$\eta = F_1\vec{\alpha}_1 + F_2\vec{\alpha}_2 + \ldots + F_{r-1}\vec{\alpha}_{r-1}.$$

We claim that

Lemma 4.

$$[F_1]_u^e = \left[f^{(n-1)}\right]_u^{e^c{:}e{:}\emptyset{:}\emptyset{:}\ldots{:}\emptyset},$$
$$[F_2]_u^e = \left[f^{(n-1)}\right]_u^{e^c{:}\emptyset{:}e{:}\emptyset{:}\ldots{:}\emptyset}, \quad \cdots$$
$$\cdots \quad [F_{r-1}]_u^e = \left[f^{(n-1)}\right]_u^{e^c{:}\emptyset{:}\emptyset{:}\emptyset{:}\ldots{:}e}.$$

Proof. We recall that $\left[\widehat{K}_{\sigma^s}\right]_i^j = \omega^{is}\delta_{ij}$, so, for $\mu = \mu_0{:}\mu_1{:}\mu_2{:}\ldots{:}\mu_{r-1}$ an ordered-r-partition of $[n]$, and e a subset of $[n-1]$ we have

$$\left[\widehat{K}_{\sigma^s}^{(e)}\right]_\mu^\mu = \omega^{s(0|\mu_0 \cap e|+1|\mu_1 \cap e|+\ldots+(r-1)|\mu_{r-1}\cap e|)},$$

so

$$\left[\widehat{K}_{\sigma^s}^{(e)}\right]_{u \cdot \gamma(u)}^{u \cdot \gamma(u)} = \omega^{s(0|u_0 \cap e|+1|u_1 \cap e|+\ldots+(r-1)|u_{r-1}\cap e|)},$$

because $e \subseteq [n-1]$. On the other hand $\left[f\right]_i^j = \omega^{ij}$, so

$$\left[f^{(n-1)}\right]_u^{e^c{:}\emptyset{:}\ldots{:}\emptyset{:}e{:}\emptyset{:}\ldots{:}\emptyset} = \omega^{s(0|u_0 \cap e|+1|u_1 \cap e|+\ldots+(r-1)|u_{r-1}\cap e|)},$$

where e appears in the sth position. $\qquad\square$

Define, for $i \in [r-1]$, the rectangular $2^{n-1} \times r^{n-1}$ matrices

$$[G_1]_e^u := \left[f^{-1(n-1)}\right]_{u \cdot \gamma(u)}^{e^c \cdot \gamma(u){:}e{:}\emptyset{:}\emptyset{:}\ldots{:}\emptyset}$$
$$[G_2]_e^u := \left[f^{-1(n-1)}\right]_{u \cdot \gamma(u)}^{e^c \cdot \gamma(u){:}\emptyset{:}e{:}\emptyset{:}\ldots{:}\emptyset}$$
$$\vdots$$
$$[G_{r-1}]_e^u := \left[f^{-1(n-1)}\right]_{u \cdot \gamma(u)}^{e^c \cdot \gamma(u){:}\emptyset{:}\emptyset{:}\emptyset{:}\ldots{:}e}.$$

Of course $G_i F_j = \delta_{ij}\mathbf{1}$, so we now have the second part of the inversion:

$$\vec{\alpha}_i = G_i\eta.$$

Inversion of any abelian group-based model

Lemma 5. *Any (finitely generated) abelian group G is isomorphic to a direct product of cyclic groups of prime-power order, ie. $G \cong \mathbb{Z}_{r_1} \times \mathbb{Z}_{r_2} \times \ldots \times \mathbb{Z}_{r_q}$ where each $r_i = p_i^{n_i}$ where p_i is prime and n_i is a positive integer.*

Lemma 6. *The group-based model arising from the G is defined only up to group isomorphisms of G.*

Proof. A generic rate matrix for the group-based model arsing from G is given by

$$Q = -\lambda \mathbf{1} + \sum_{e \neq \sigma \in G} \alpha^\sigma K_\sigma.$$

Under a group isomorphism $\phi : G \rightarrow G'$, we have $\phi(\sigma_i \sigma_j) = \phi(\sigma_i)\phi(\sigma_j)$.

Recall (2), so that the matrix elements $[K_\sigma]_i^j$ is set via the action $\sigma_i \mapsto \sigma \sigma_i = \sigma_j$. If we consider the regular representation of G' we then have $[K_{\phi(\sigma)}]_i^j$ defined by $\phi(\sigma_i) \mapsto \phi(\sigma)\phi(\sigma_i)$. Now $\phi(\sigma)\phi(\sigma_i) = \phi(\sigma \sigma_i) = \phi(\sigma_j)$ and, because ϕ is a group isomorphism, this occurs if and only if $\sigma \sigma_i = \sigma_j$. Thus $[K_{\phi(\sigma)}]_i^j = [K_\sigma]_i^j$ for all i and j. □

This means that we can restrict attention to a single representitive in the isomorphism class of G. Of course, for this purpose we choose the representative guaranteed by Lemma 5.

Thus, for any abelian group G, with generators σ_1, $\sigma_2, \ldots, \sigma_q$ the corresponding group-based model has rate generators given by

$$L_\sigma = -\mathbf{1} + K_{\sigma_1}^{m_1} \otimes K_{\sigma_2}^{m_2} \otimes \ldots \otimes K_{\sigma_q}^{m_q},$$

for all $e \neq \sigma = \left(\sigma_1^{m_1}, \sigma_2^{m_2}, \ldots, \sigma_q^{m_q}\right) \in G$, where K_{σ_i} is the permutation matrix representing the generator $\sigma_i \in \mathbb{Z}_{r_i}$. The character table f of G is simply the tensor product of the individual character tables of the $\mathbb{Z}_{r_i}$:

$$f = f_1 \otimes f_2 \otimes \ldots \otimes f_q.$$

In the diagonal basis we have matrix elements

$$\left[f_k K_{\sigma_k^s} f_k^{-1}\right]_i^j = \left[\hat{K}_{\sigma_k^s}\right]_i^j = (\omega_k)^{is} \delta_{ij},$$

where ω_k is a kth root of unity. Thus

$$\left[\hat{K}_{\sigma_1}^{m_1} \otimes \hat{K}_{\sigma_2}^{m_2} \otimes \ldots \hat{K}_{\sigma_q}^{m_q}\right]_{i_1 i_2 \ldots i_q}^{j_1 j_2 \ldots j_q}$$
$$= \delta_{i_1 j_1} \delta_{i_2 j_2} \ldots \delta_{i_q j_q} (\omega_1)^{i_1 m_1} (\omega_2)^{i_2 m_2} \ldots (\omega_q)^{i_q m_q}.$$

We write phylogenetic tensors for this model in the form

$$P_{i_{11} i_{12} \ldots i_{1n}, i_{21} i_{22} \ldots i_{2n} \ldots \ldots i_{q1} i_{q2} \ldots i_{qn}},$$

where $0 \leq i_{sj} \leq r_s$ for all $0 \leq s \leq q$. We simplify notation by writing each group of indices as $\mu^{(s)} := i_{s1} i_{s2} \ldots i_{sn}$ where $\mu^{(s)}$ is an ordered-r_s-partition of $[n]$.

Lemma 7. *In the diagonal basis, the uniform initial distribution on the star tree has components*

$$\left[\widehat{\delta^{n-1}\pi}\right]_{\mu^{(1)}\mu^{(2)}\ldots\mu^{(q)}}$$
$$= \begin{cases} 1, & if\ 0|\mu_0^{(i)}| + 1|\mu_1^{(i)}| + \ldots + (r_i-1)|\mu_{r-1}^{(i)}| = 0 \forall i; \\ 0, & otherwise. \end{cases}$$

A generic phylogenetic tensor for this model can be expressed as

$$P = e^{-\lambda} \exp\left(\sum_{\emptyset \neq e \subseteq [n-1]}^{s_i \in [r_i-1]} \alpha_e^{s_1 s_2 \ldots s_q} K_{\sigma_1^{s_1}}^{(e)} \otimes K_{\sigma_2^{s_2}}^{(e)} \otimes \ldots \otimes K_{\sigma_q^{s_q}}^{(e)}\right) \cdot \delta^{n-1}\pi,$$

where π is the unifrom distribution on $\sum_{i=1}^q r_i$ states, i.e.

$$\pi = \left(\sum_{i=1}^q r_i\right)^{-1} (1, 1, \ldots, 1)^T.$$

In the diagonal basis $\widehat{P} = (f_1 \otimes f_2 \otimes \ldots \otimes f_q)^{(n)} \cdot P$, and, as a consequence of the previous lemma, P has many vanishing components. To avoid these, for each $i \in [q]$ we take $u^{(i)} = u_0^{(i)} : u_1^{(i)} : u_2^{(i)} : \ldots : u_{r_i-1}^{(i)}$ as an *ordered-r_i-partition* of $[n-1]$ and set

$$\gamma_i(u^{(i)}) = r_i - \left(0|u_0^{(i)}| + 1|u_1^{(i)}| + 2|u_2^{(i)}| + \ldots \right.$$
$$\left. + (r_i - 1)|u_{r-1}^{(i)}|\right) (\text{mod } r).$$

We then define

$$\mathcal{P}_{u^{(1)} u^{(2)} \ldots u^{(q)}} := \left[\widehat{P}\right]_{u^{(1)} \cdot \gamma_1(u^{(1)}) u^{(2)} \cdot \gamma_2(u^{(2)}) \ldots u^{(q)} \cdot \gamma_1(u^{(q)})},$$

and

$$\eta_{u^{(1)} \ldots u^{(q)}}$$
$$:= \left[\sum_{\emptyset \neq e \subseteq [n-1]}^{s_i \in [r_i-1]} \alpha_e^{s_1 \ldots s_q} \widehat{K}_{\sigma_1^{s_1}}^{(e)} \otimes \ldots \otimes \widehat{K}_{\sigma_q^{s_q}}^{(e)}\right]_{u^{(1)} \cdot \gamma_1(u^{(1)}) \ldots u^{(q)} \cdot \gamma_1(u^{(q)})}^{u^{(1)} \cdot \gamma_1(u^{(1)}) \ldots u^{(q)} \cdot \gamma_1(u^{(q)})},$$

so that we have the first part of the inversion

$$\mathcal{P}_{u^{(1)} u^{(2)} \ldots u^{(q)}} = e^{-\lambda} \exp\left(\eta_{u^{(1)} u^{(2)} \ldots u^{(q)}}\right),$$
$$\eta_{u^{(1)} u^{(2)} \ldots u^{(q)}} = \lambda + \ln\left(\mathcal{P}_{u^{(1)} u^{(2)} \ldots u^{(q)}}\right).$$

We define the column vectors $\vec{\alpha}^{s_1 s_2 \ldots s_q} := \left\{ \alpha_e^{s_1 s_2 \ldots s_q} \right\}_{\emptyset \neq e \subseteq [n-1]}$ and $\vec{\eta} := \{ \eta_{u^{(1)} u^{(2)} \ldots u^{(q)}} \}$ where u_i is an ordered-r_i-partition of $[n-1]$, and we define the $(r_1 r_2 \ldots r_q)^{n-1} \times 2^{n-1}$ matrices

$$\left[F_{s_1 s_2 \ldots s_q} \right]^e_{u_1 u_2 \ldots u_q}$$

$$:= \left[K_{\sigma_1^{s_1}}^{(e)} \right]^{u_1 \cdot \gamma(u_1)}_{u_1 \cdot \gamma(u_1)} \left[K_{\sigma_2^{s_2}}^{(e)} \right]^{u_2 \cdot \gamma(u_2)}_{u_2 \cdot \gamma(u_2)} \ldots \left[K_{\sigma_q^{s_q}}^{(e)} \right]^{u_q \cdot \gamma(u_q)}_{u_q \cdot \gamma(u_q)}$$

$$= \left[f_1^{(n-1)} \right]^{e^c:\emptyset:\ldots:\emptyset:e:\emptyset:\ldots:\emptyset}_{u_1} \left[f_2^{(n-1)} \right]^{e^c:\emptyset:\ldots:\emptyset:e:\emptyset:\ldots:\emptyset}_{u_2} \ldots$$

$$\ldots \left[f_q^{(n-1)} \right]^{e^c:\emptyset:\ldots:\emptyset:e:\emptyset:\ldots:\emptyset}_{u_q},$$

where in each term e appears in the s_i^{th} position and the equality follows from Lemma 4.

We can then write the vector equation

$$\vec{\eta} = \sum_{s_1 s_2 \ldots s_q : 1 \leq s_i \leq r_i - 1} F_{s_1 s_2 \ldots s_q} \vec{\alpha}^{s_1 s_2 \ldots s_q}.$$

If we define the $2^{n-1} \times (r_1 r_2 \ldots r_q)^{n-1}$ matrices

$$\left[G_{s_1 s_2 \ldots s_q} \right]^{u_1 u_2 \ldots u_q}_e$$

$$= \left[f_1^{-1 (n-1)} \right]^{e^c:\emptyset:\ldots:\emptyset:e:\emptyset:\ldots:\emptyset}_{u_1} \left[f_2^{-1 (n-1)} \right]^{e^c:\emptyset:\ldots:\emptyset:e:\emptyset:\ldots:\emptyset}_{u_2} \ldots$$

$$\ldots \left[f_q^{-1 (n-1)} \right]^{e^c:\emptyset:\ldots:\emptyset:e:\emptyset:\ldots:\emptyset}_{u_q},$$

where in each term e appears in the s_i^{th} position, we have the orthogonality relations

$$G_{s_1 s_2 \ldots s_q} F_{s_1' s_2' \ldots s_q'} = \delta_{s_1 s_1'} \delta_{s_2 s_2'} \ldots \delta_{s_q s_q'} \mathbf{1}.$$

This gives us the second part of the inversion of any group-based model:

$$\vec{\alpha}^{s_1 s_2 \ldots s_q} = G_{s_1 s_2 \ldots s_q} \vec{\eta}.$$

Conclusion

In this article we have given an alternative derivation of the inversion of group-based phylogenetic models. Primarily our method relies on the remarkable intertwining relation between branching events and Markov evolution (4), and the resulting simplified expression of phylogenetic tensors given in (5). From there we took a representation theoretic approach concentrating on the structure of tensor indices.

Authors' contributions
JGS conducted much of the theoretical work presented, and was primarily responsible for drafting of the paper. PDJ contributed to the theoretical work and assisted in editing of the paper. BRH contributed to the theoretical work and contributed substantially to the writing of the paper. All authors read and approved the final manuscript.

Authors' information
Jeremy G. Sumner is a ARC Research Fellow, Peter D. Jarvis is a Alexander von Humboldt Fellow and Barbara R. Holland is a ARC Future Fellow.

References
1. Felsenstein J: *Inferring Phylogenies*. Sunderland: Sinauer Associates; 2004.
2. Semple C, Steel M: *Phylogenetics*. Oxford: Oxford University Press; 2003.
3. Hendy MD, Penny D: **A framework for the quantitative study of evolutionary trees.** *Syst Zool* 1989, **38:**297–309.
4. Hendy MD: **The relationship between simple evolutionary tree models and observable sequence data.** *Syst Zool* 1989, **38:**310–321.
5. Hendy MD, Penny D: **Spectral analysis of phylogenetic data.** *J Class* 1993, **10:**1–20.
6. Steel M, Hendy M, Székely L, Erdös P: **Spectral analysis and a closest tree method for genetic sequences.** *Appl Math Lett* 1992, **5**(6):63–67.
7. Székely LA, Erdos P, Steel M, Penny D: **A fourier inversion formula for evolutionary trees.** *Appl Math Lett* 1993, **6**(2):13–16.
8. Hendy MD, Penny D, Steel M: **A discrete Fourier analysis for evolutionary trees.** *Proc Natl Acad Sci* 1994, **91:**3339–3343.
9. Székely LA, Steel MA, Erdős PL: **Fourier calculus on evolutionary trees.** *Adv Appl Math* 1993, **14:**200–216.
10. Hendy MD, Charleston MA: **Hadamard conjugation: a versatile tool for modelling nucleotide sequence evolution.** *New Zeal J Bot* 1993, **31:**231–237.
11. Holland BR, Penny D, Hendy MD: **Outgroup misplacement and phylogenetic inaccuracy under a molecular clock – a simulation study.** *Syst Biol* 2003, **52:**229–238.
12. Hendy MD: **A combinatorial description of the closest tree algorithm for finding evolutionary trees.** *Discrete Math* 1991, **96:**51–58.
13. Lento GM, Hickson RE, Chambers GK, Penny D: **Use of spectral analysis to test hypotheses on the origin of pinnipeds.** *Mol Biol Evol* 1995, **12:**28–52.
14. Huber KT, Watson EE, Hendy MD: **An algorithm for constructing local regions in a phylogenetic network.** *Mol Phylogenet Evol* 2001, **19:**1–8.
15. Huber KT, Langton M, Penny D, Moulton V, Hendy M: **Spectronet: a package for computing spectra and median networks.** *Appl Bioinform* 2002, **1:**2041–2059.
16. Schliep KP: **Some applications of statistical phylogenetics.** *Ph.D. thesis.* Massey University; 2009.
17. von Haeseler A, Churchill GA: **Network models for sequence evolution.** *J Mol Evol* 1993, **37:**77–85.
18. Bryant D: **Extending tree models to split networks.** In *Algebraic Statistics and Computational Biology*. Edited by Pachter L, Sturmfels B. Cambridge: Cambridge University Press; 2005:297–368.
19. Bryant D: **Hadamard phylogenetic methods and the *n*-taxon process.** *Bull Math Biol* 2009, **71:**297–309.
20. Matsen FA, Steel M: **Phylogenetic mixtures on a single tree can mimic a tree of another topology.** *Syst Biol* 2007, **56:**767–775.
21. Matsen FA, Mossel E, Steel M: **Mixed-up trees: the structure of phylogenetic mixtures.** *Bull Math Biol* 2008, **70:**1115–1139.
22. Griffiths RC, Majoram P: **Ancestral inference from samples of DNA sequences with recombination.** *J Comput Biol* 1996, **3:**479–502.
23. Griffiths RC, Marjoram P: **An ancestral recombination graph.** In *Progress in Population Genetics and Human Evolution, Volume 87 of IMA volumes in mathematics and its applications*. Berlin: Springer Verlag; 1997:257–270.
24. Jin G, Nakhleh L, Snir S, Tuller T: **Maximum likelihood of phylogenetic networks.** *Bioinformatics* 2006, **21:**2604–2611.
25. Strimmer K, Moulton V: **Likelihood analysis of phylogenetic networks using directed graphical models.** *Mol Biol Evol* 2000, **17:**875–881.

26. Strimmer K, Wiuf C, Moulton V: **Recombination analysis using directed graphical models.** *Mol Biol Evol* 2001, **18:**97–99.
27. Sumner JG, Holland BR, Jarvis PD: **The algebra of the general Markov model on trees and networks.** *Bull Math Biol* 2012, **74**(4):858–880.
28. Sumner JG, Charleston MA, Jermiin LS, Jarvis PD: **Markov invariants, plethysms, and phylogenetics.** *J Theor Biol* 2008, **253:**601–615.
29. Sumner JG, Jarvis PD: **Markov invariants and the isotropy subgroup of a quartet tree.** *J Theor Biol* 2009, **258:**302–310.
30. Holland BR, Jarvis PD, Sumner JG: **Low-parameter phylogenetic inference under the general markov model.** *Syst Biol* 2013, **62:**78–92.
31. Bashford JD, Jarvis PD, Sumner JG, Steel MA: $U(1) \times U(1) \times U(1)$ **symmetry of the Kimura 3ST model and phylogenetic branching processes.** *J Phys A Math Gen* 2004, **37:**L1–L9.
32. Sumner JG, Fernández-Sánchez J, Jarvis PD: **Lie Markov models.** *J Theor Biol* 2012, **298:**16–31.
33. Sagan BE: *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions.* Second Edition. New York: Springer; 2001. Graduate Texts in Mathematics.
34. Evans SN, Speed T: **Invariants of some probability models used in phylogenetic inference.** *Ann Stat* 1993, **21:**355–377.
35. Sumner JG, Jarvis PD: **Entanglement invariants and phylogenetic branching.** *J Math Biol* 2005, **51:**18–36.
36. Kimura M: **Estimation of evolutionary distances between homologous nucleotide sequences.** *Proc Natl Acad Sci* 1981, **78:**1454–1458.