

Lie Markov models with purine/pyrimidine symmetry

Jesús Fernández-Sánchez · Jeremy G. Sumner ·
Peter D. Jarvis · Michael D. Woodhams

Received: 14 December 2012 / Revised: 13 February 2014 / Published online: 11 April 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Continuous-time Markov chains are a standard tool in phylogenetic inference. If homogeneity is assumed, the chain is formulated by specifying time-independent rates of substitutions between states in the chain. In applications, there are usually extra constraints on the rates, depending on the situation. If a model is formulated in this way, it is possible to generalise it and allow for an inhomogeneous process, with time-dependent rates satisfying the same constraints. It is then useful to require that, under some time restrictions, there exists a homogeneous average of this inhomogeneous process within the same model. This leads to the definition of “Lie Markov models” which, as we will show, are *precisely* the class of models where such an average exists. These models form Lie algebras and hence concepts from Lie group theory are central to their derivation. In this paper, we concentrate on applications to phylogenetics and nucleotide evolution, and derive the complete hierarchy of Lie Markov models that respect the grouping of nucleotides into purines and pyrimidines—that is, models with purine/pyrimidine symmetry. We also discuss how to handle the subtleties of applying Lie group methods, most naturally defined

Electronic supplementary material The online version of this article (doi:[10.1007/s00285-014-0773-z](https://doi.org/10.1007/s00285-014-0773-z)) contains supplementary material, which is available to authorized users.

J. Fernández-Sánchez (✉)
Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Barcelona, Spain
e-mail: jesus.fernandez.sanchez@upc.edu

J. G. Sumner · P. D. Jarvis · M. D. Woodhams
School of Mathematics and Physics, University of Tasmania, Tasmania, Australia
e-mail: jsumner@utas.edu.au

P. D. Jarvis
e-mail: peter.jarvis@utas.edu.au

M. D. Woodhams
e-mail: michael.woodhams@utas.edu.au

over the complex field, to the stochastic case of a Markov process, where parameter values are restricted to be real and positive. In particular, we explore the geometric embedding of the cone of stochastic rate matrices within the ambient space of the associated complex Lie algebra.

Keywords Phylogenetics · Markov model · Representation theory · Permutation groups · Lie algebras

Mathematics Subject Classification 15A30 · 22E60 · 52B99 · 62P10

1 Introduction

Most of the commonly implemented models in molecular phylogenetics are based on the continuous-time Markov assumption. For these models, molecular substitution events (along an edge of a phylogenetic tree) are ruled by substitution rates. For DNA models—where the state space consists of the four nucleotides adenine, cytosine, guanine and thymine—12 substitution rates must be specified for each edge of the evolutionary tree, and the precise characteristics of the process are fixed by constraints on these rates. These constraints define a space of parameter values where each point corresponds to unknown evolutionary quantities such as base composition and mutation rate. As in all applied statistics, there is a trade-off between more complex, realistic models, and simpler, tractable models: complex models can provide very close fits to the observed data, but are more vulnerable to random error. A standard assumption in molecular phylogenetics is to work with homogeneous Markov chains, where the substitution rates are assumed to be constant in time.

The motivation behind our previous work (Sumner et al. 2012a) was to consider the consequences of allowing for some change in individual substitution rates which may well have occurred independently across the evolutionary history. With this perspective, the evolutionary process can still be modelled as a continuous-time Markov chain, but we must allow the process to be inhomogeneous, where the rates are allowed to vary as a function of time throughout the evolutionary history. This leads to considering evolutionary model classes that are “locally multiplicatively closed”, that is, models where the product of substitution matrices is still in the model as long as they are sufficiently close to the identity.¹ For such models, it is possible to interpret the time average of their inhomogeneous behaviour as a homogeneous process within the same model class. Many oft-used models, such as the general time-reversible model (Tavaré 1986; Posada and Crandall 1998), are not multiplicatively closed do not satisfy this property and this deficiency poses a problem for phylogenetic analysis in both flexibility of interpretation, and as a potential source of model-misspecification (Sumner et al. 2012). For a locally multiplicatively closed model, under some time restrictions it

¹ The reader may notice that we have changed the terminology of Sumner et al. (2012a) and we refer to the desired property as “locally multiplicative closure” instead of “multiplicative closure”. The problem of *global* multiplicative closure for a continuous-time Markov model is a deep problem related to the convergence of the Baker–Campbell–Hausdorff formula (see Blanes and Casas 2004). Notice that this is not a serious drawback as the nature of the problem is local.

is possible to model evolutionary processes homogeneously, by interpreting the fitted substitution rates as an “average” of the true inhomogeneous process occurring on each branch of the tree. Sumner et al. (2012a) presented sufficient conditions for local multiplicative closure of continuous-time Markov chains, and this lead directly to the concept of Lie Markov models. These models arise when we demand the set of rate matrices of the model form a Lie algebra. This is a technical condition guaranteeing the corresponding set of substitution probability matrices will be locally multiplicatively closed, as desired. Moreover, we will show that this condition is actually sufficient (see Theorem 1). Mathematically, Lie Markov models can be regarded as a generalisation of other model classes, such as *equivariant* models introduced by Draisma and Kuttler (2008) or *group-based* models (Semple and Steel 2003; Michałek 2011; Donten-Bury and Michałek 2012).

Sumner et al. (2012a) discussed the symmetry properties of DNA models to nucleotide permutations, and noted the statistical relevance of these symmetries to likelihood calculations. The main result of that paper was a procedure to generate multiplicatively closed Markov models with a prescribed symmetry, which has desirable properties in terms of model selection. For instance, a biologist may wish that candidate models do not provide any natural groupings of nucleotides, and hence \mathfrak{S}_4 symmetry—i.e. the symmetry group of all possible nucleotide permutations—is appropriate. It is then a matter of choosing how many free parameters are appropriate for the given data set. The complete hierarchy of Lie Markov models with \mathfrak{S}_4 symmetry was derived by Sumner et al. (2012a).

In this paper, we deal with the case of locally multiplicatively closed Markov models whose symmetry is consistent with the grouping of nucleotides into purines and pyrimidines, i.e. $AG \mid CT = \{\{A, G\}, \{C, T\}\}$. As will be discussed, this motivates us to produce and examine the Lie Markov models with symmetry governed by the permutation subgroup of \mathfrak{S}_4 that preserves the purine/pyrimidine grouping:²

$$\mathcal{G} := \{e, (AG), (CT), (AG)(CT), (AC)(GT), (AT)(CG), (ACGT), (ATGC)\},$$

where e is the identity, or “do nothing”, permutation.

We will also go further than Sumner et al. (2012a) by exploring the definition of these models and investigate the geometrical properties that arise naturally when we deal with the tension between the algebraic formalism of Lie groups, where one works over the complex field, and the stochastic constraints of Markov models, where parameter values are constrained to be real and positive. In particular, we discuss the geometric embedding of the stochastic rate matrices within the vector space of complex rate matrices. These considerations motivate our definition of the *stochastic cone* of a Lie Markov model. Besides its geometrical interest, the stochastic cone is the set of stochastic rate matrices of the model and in a practical context is actually the main object of interest. We plan to discuss implementation and performance of the models we present here in a future publication.

² Note this group is isomorphic to the dihedral group \mathbf{D}_4 , which describes the symmetries of a square. However, it also admits a more natural description in our setting as $\mathfrak{S}_2 \wr \mathfrak{S}_2$, the wreath product of \mathfrak{S}_2 with itself (see Rotman 1995, Chapter VII).

Although our presentation focuses on the purine/pyrimidine grouping $AG|CT$, given the appropriate nucleotide permutation, exactly the same hierarchy of models would arise if we were to consider the grouping $AC|GT$, or the grouping $AT|GC$. The reader should note that choosing the grouping $AT|GC$ would give the classification of all Lie Markov models that preserve *complementation* $A \leftrightarrow T$, $C \leftrightarrow G$ (see [Yap and Pachter 2004](#)). In particular, the “strand-symmetric” model defined by [Casanellas and Sullivant \(2005\)](#) arises in this way from our Model 6.6 (see [Table 1](#)). The conversion of our hierarchy of models from the $AG|CT$ grouping to the $AT|GC$ grouping would follow by simultaneously permuting the G and T rows and columns of the rate matrices in each model.

In [Sect. 2](#) we recall some of the basic definitions and tools introduced by [Sumner et al. \(2012a\)](#). We revisit the definition of Lie Markov models, and introduce the concept of the stochastic cone of a Lie Markov model. We also recall the basic results on group theory and representation theory necessary for the development of our results. In [Sect. 3](#) we recall the idea of Lie Markov model with prescribed symmetry given by a permutation group G . We introduce the ray-orbits of the corresponding stochastic cone, which are the orbits under the action of G of the rays of the stochastic cone. In [Sect. 4](#), we take $G = \mathcal{G}$, decompose the space of rate matrices as a \mathcal{G} -module and provide a basis consistent with this decomposition. We also determine the isomorphism classes of possible \mathcal{G} -orbits and the decomposition of their (abstract) span into irreducible modules. In [Sect. 5](#), we give the whole list of Lie Markov models with purine/pyrimidine symmetry. Each model is given by exhibiting a basis of the corresponding space of matrices as well as the ray-orbits of its stochastic cone. Among these models, we obtain the Jukes–Cantor model, the Kimura models with two or three parameters, the general Markov model and a number of new models that may have special interest for the biologists. Some of these are shown in [Table 1](#) as an appetizer before the whole list, which itself can be found in explicit form as [Supplementary Material online](#). Finally, in the conclusions we discuss implications and possibilities for future research.

2 Preliminaries

Throughout this section, we will recall some definitions and basic facts from [Sumner et al. \(2012a\)](#), which we also refer to for some proofs. We keep the assumptions and the notation already introduced there. In particular, we work over the complex field \mathbb{C} , and for simplicity refer to a matrix as “Markov” if the entries in each column sum to one. Later we will discuss how to specialise to the stochastic case where the entries must be real numbers in the range $[0, 1]$. This will lead us to considering the stochastic cone of the Lie Markov model, which will be the set of real rate matrices with non-negative entries outside the diagonal.

We define the *general Markov* model \mathfrak{M}_{GM} as the set of $n \times n$ matrices whose columns sum to one:

$$\mathfrak{M}_{GM} := \left\{ M \in \mathbb{M}_n(\mathbb{C}) : \theta^T M = \theta^T \right\},$$

Table 1 Some Lie Markov models with purine/pyrimidine symmetry that may have special interest for biologists

Model	Rate matrix	Description
3.3b	$\begin{pmatrix} * & \alpha & \beta & \gamma \\ \alpha & * & \gamma & \beta \\ \gamma & \beta & * & \alpha \\ \beta & \gamma & \alpha & * \end{pmatrix}$	A 3-dimensional model, with parameter α for transitions $A \leftrightarrow G, C \leftrightarrow T$, and two different parameters for transversions: β for $A \mapsto C \mapsto G \mapsto T \mapsto A$ and γ for $A \mapsto T \mapsto G \mapsto C \mapsto A$.
3.3c	$\begin{pmatrix} * & \alpha & \beta & \beta \\ \alpha & * & \beta & \beta \\ \beta & \beta & * & \gamma \\ \beta & \beta & \gamma & * \end{pmatrix}$	A 3-dimensional model, with two parameters for transitions: α for $A \leftrightarrow G$ and γ for $C \leftrightarrow T$, and one parameter for transversions: β for $A \leftrightarrow C \leftrightarrow G \leftrightarrow T \leftrightarrow A$. Note the similarity of both of these models with model by Kimura (1981), which also belongs to our hierarchy as Model 3.3a.
4.4a	$\begin{pmatrix} * & \alpha & \alpha & \alpha \\ \beta & * & \beta & \beta \\ \gamma & \gamma & * & \gamma \\ \delta & \delta & \delta & * \end{pmatrix}$	The model by Felsenstein (1981): a 4-dimensional model, where mutations share the same rate when they change to the same nucleotide.
4.4b	$\begin{pmatrix} * & \alpha & \beta & \beta \\ \alpha & * & \beta & \beta \\ \gamma & \gamma & * & \delta \\ \gamma & \gamma & \delta & * \end{pmatrix}$	A 4-dimensional model, with two parameters for transitions: α for $A \leftrightarrow G$ and γ for $C \leftrightarrow T$, and the other two parameters for transversions: β for $A \mapsto C \mapsto G \mapsto T \mapsto A$ and δ for $A \mapsto T \mapsto G \mapsto C \mapsto A$.
5.6b	$\begin{pmatrix} * & a+x & b+x & b+x \\ a+y & * & b+y & b+y \\ b+z & b+z & * & a+z \\ b+t & b+t & a+t & * \end{pmatrix}$	A 5-dimensional model, where rates depend on two families of parameters $\{a, b\}$ and $\{x, y, z, t\}$: transitions and transversions have parameters a and b respectively, while they are affected by some other parameters according to the nucleotide they change to: x for mutations to A , y for mutations to G , z for mutations to C and t for mutations to T . Notice the resemblance of this model with the model by Hasegawa et al. (1988); see Remark 11.
6.6	$\begin{pmatrix} * & \alpha & \beta & \gamma \\ \alpha & * & \gamma & \beta \\ \delta & \varepsilon & * & \zeta \\ \varepsilon & \delta & \zeta & * \end{pmatrix}$	This 6-dimensional model has two parameters for different transitions: α for $A \leftrightarrow G$ and ζ for $C \leftrightarrow T$, and 4 for transversions: $\beta : A \mapsto C, G \mapsto T, \gamma : A \mapsto T, G \mapsto C, \delta : C \mapsto A, G \mapsto T, \varepsilon : A \mapsto T, G \mapsto C$. By permuting rows and columns according to (GT) (or (AC)), we obtain the <i>strand symmetric</i> model introduced by Casanellas and Sullivan (2005).
6.7a	$\begin{pmatrix} * & a+x & b+x & c+x \\ a+y & * & c+y & b+y \\ b+z & c+z & * & a+z \\ c+t & b+t & a+t & * \end{pmatrix}$	This 6-dimensional model is similar to 5.6b, but it has two parameters for different transversions: b for $A \mapsto C \mapsto G \mapsto T \mapsto A$ and c for $A \mapsto T \mapsto G \mapsto C \mapsto A$.

where θ is the column n -vector with all its entries equal to 1, i.e. $\theta^T = (1, 1, \dots, 1)$. Recall that, in a homogeneous continuous-time Markov chain, the corresponding Markov matrices occur as exponentials $M = e^{Qt}$, where Q is a “rate matrix” and t is time elapsed. We write

$$\mathcal{L}_{GM} := \left\{ Q \in M_n(\mathbb{C}) : \theta^T Q = \mathbf{0}^T \right\}$$

to indicate the set of all (complex) rate matrices. We refer to a Markov matrix $M \in \mathfrak{M}_{GM}$, or a rate matrix $Q \in \mathfrak{L}_{GM}$, as “stochastic” if its off-diagonal elements are real and positive.

Under matrix multiplication, the set

$$GL_1(n, \mathbb{C}) := \left\{ M \in \mathbb{M}_n(\mathbb{C}) : \theta^T M = \theta^T, \det(M) \neq 0 \right\},$$

forms a subgroup of the general linear group of invertible $n \times n$ matrices with complex entries, i.e. $GL_1(n, \mathbb{C}) < GL(n, \mathbb{C})$. It contains the matrix exponential of any rate matrix, that is,

$$e^{\mathfrak{L}_{GM}} := \left\{ e^Q : Q \in \mathfrak{L}_{GM} \right\} \subset GL_1(n, \mathbb{C}).$$

We refer to $e^{\mathfrak{L}_{GM}}$ as the *general rate matrix model*.

A *Markov model* \mathfrak{M} is some subset $\mathfrak{M} \subseteq \mathfrak{M}_{GM}$ of the general Markov model containing the identity matrix $\mathbf{1}$. A Markov model \mathfrak{M} is *locally multiplicatively closed* if for all $M_1, M_2 \in \mathfrak{M}$ in a neighbourhood of $\mathbf{1}$ we also have the matrix product $M_1 M_2 \in \mathfrak{M}$. Similarly, given a subset $\mathfrak{L} \subseteq \mathfrak{L}_{GM}$ of rate matrices containing the null matrix, we refer to $e^{\mathfrak{L}}$ as a *rate matrix model*. It is clear that all rate matrix models are Markov models, and we simplify terminology and also refer to \mathfrak{L} as a “model”.

We are primarily interested in rate matrix models $\mathfrak{M} = e^{\mathfrak{L}}$ which that are locally multiplicatively closed. For such a model \mathfrak{M} , suppose that it is a smooth manifold around the identity matrix $\mathbf{1}$, so there exist differentiable paths $A(t) \in \mathfrak{M}$ with $A(0) = \mathbf{1}$. Then, we can define the tangent space at the identity: $T_{\mathbf{1}}(\mathfrak{M}) = \{A'(0) : A(t) \in \mathfrak{M}, A(0) = \mathbf{1}\}$. The following theorem provides a characterization for these models. Recall that a $\mathfrak{L} \subseteq \mathfrak{L}_{GM}$ is a *Lie algebra* if for all $Q_1, Q_2 \in \mathfrak{L}$ and $\lambda \in \mathbb{C}$:

1. $Q_1 + \lambda Q_2 \in \mathfrak{L}$,
2. $[Q_1, Q_2] := Q_1 Q_2 - Q_2 Q_1 \in \mathfrak{L}$.

The first condition states that \mathfrak{L} is a vector space, and the second states that \mathfrak{L} is closed under “Lie brackets”.

Theorem 1 (cf. [Birkhoff 1938](#)) *A model $\mathfrak{M} = e^{\mathfrak{L}}$ is locally multiplicatively closed if and only if $T_{\mathbf{1}}(\mathfrak{M})$ is a Lie subalgebra of \mathfrak{L}_{GM} .*

Proof The sufficient condition is a consequence of the Baker–Campbell–Hausdorff formula ([Campbell 1897](#)): if \mathfrak{L} forms a Lie algebra, there is a small ball $B = B_\varepsilon$ around 0 in \mathfrak{L}_{GM} such that if $X, Y \in B$, then the product given by the Baker–Campbell–Hausdorff expansion

$$X * Y := X + Y + \frac{1}{2}[X, Y] + \dots$$

is absolutely convergent and associative. Then, for any $X, Y \in B$, we can write $e^X e^Y = e^{X * Y}$. Define $U = \exp(B \cap \mathfrak{L})$, which is a neighbourhood of $\mathbf{1}$ in \mathfrak{M} , and conclude that if $M_1, M_2 \in U$, then $M_1 M_2 \in \mathfrak{M}$.

Conversely, assume that $\mathfrak{M} = e^{\mathcal{L}}$ is locally multiplicatively closed, so there is a neighbourhood U of $\mathbf{1}$ in \mathfrak{M} such that if $M_1, M_2 \in U$, then $M_1 M_2 \in \mathfrak{M}$. Given X, Y in the tangent space $T_1(\mathfrak{M})$, define $A(t) = e^{tX}$, $B(t) = e^{tY}$, and $C(s, t) = A(t)B(s)A(t)^{-1}$. There exists some $\varepsilon > 0$ such that if $0 < t, s < \varepsilon$, then $A(t), B(s), A(t)^{-1} \in U$ and also $A(t)B(s) \in U$. Because of the assumption, we infer that $C(s, t) \in \mathfrak{M}$. By taking derivatives, we conclude that

$$\frac{d}{ds} \left(\left. \frac{dC(s, t)}{dt} \right|_{t=0} \right) \Big|_{s=0} = [X, Y] \in T_1(\mathfrak{M}).$$

It follows that $T_1(\mathfrak{M})$ is a Lie subalgebra. □

Presently, we recall the Lie algebra structure of the general Markov model (Johnson 1985; Sumner et al. 2012a). To this aim, consider the set of “elementary” rate matrices $\{L_{ij} : 1 \leq i \neq j \leq n\}$, where L_{ij} is the $n \times n$ matrix with 1 in the ij entry, -1 in the ji entry and 0 everywhere else. The matrices $\{L_{ij}\}_{i \neq j}$ form a \mathbb{C} -basis for the tangent space of $GL_1(n, \mathbb{C})$ and, in particular, we can express any rate matrix Q as a linear combination:

$$Q = \sum_{i \neq j} \alpha_{ij} L_{ij}. \tag{1}$$

This is a convenient basis for \mathcal{L}_{GM} because the stochastic condition on Q is simply that the coefficients α_{ij} are real and non-negative. Moreover, if δ_{ij} denotes the Kronecker delta ($\delta_{ii} = 1$ and $\delta_{ij} = 0$ when $i \neq j$), the equalities

$$[L_{ij}, L_{kl}] = (L_{ij} - L_{jl})(\delta_{jk} - \delta_{jl}) - (L_{kj} - L_{lj})(\delta_{il} - \delta_{jl})$$

exhibit the Lie algebra structure of \mathcal{L}_{GM} .

Given a vector subspace $\mathcal{L} \subset \mathcal{L}_{GM}$, a *stochastic generating set* for \mathcal{L} is a generating set $B_{\mathcal{L}} = \{L_1, L_2, \dots, L_d\}$ of \mathcal{L} such that each L_k is a non-negative linear combination of the L_{ij} , i.e. $L_k = \sum_{i \neq j} \alpha_{ij} L_{ij}$ where $\alpha_{ij} \geq 0$. A *stochastic basis* of \mathcal{L} is a stochastic generating set where the vectors are linearly independent.

Definition 1 (cf. Sumner et al. 2012a) A *Lie Markov model* is a Lie subalgebra \mathcal{L} of \mathcal{L}_{GM} for which there exists a stochastic basis.

Leaving the technical aspects aside, a Lie Markov model is a model for which the product of two substitution matrices is still in the model. The motivation of such models is given by the fact a non-homogeneous evolutionary processes can be described in a homogeneous fashion. In more concrete terms, if $M_1, M_2 \in e^{\mathcal{L}}$, then $M_1 M_2 \in e^{\mathcal{L}}$, i.e. for any inhomogeneous process on an edge where the rate matrices always lie within \mathcal{L} , there is an equivalent homogeneous process on that edge, whose rate matrix also lies within \mathcal{L} . This is not the case for the general time reversible model (the reader is referred to the paper by Sumner et al. (2012a) for a detailed proof of the non-closure of GTR).

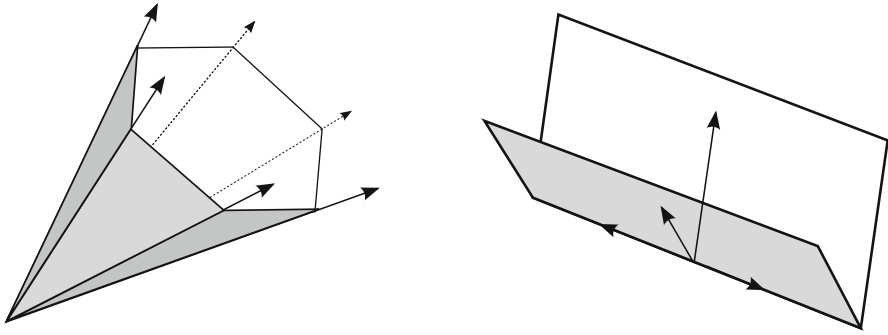


Fig. 1 A strongly convex polyhedral cone of dimension 3 with six rays (represented by arrows) and a convex polyhedral cone which is not strongly convex

Remark 1 By an elementary result in linear algebra, any generating set for a vector space can be reduced to a basis by removing elements, and hence Definition 1 would remain unchanged if “stochastic basis” were replaced with “stochastic generating set”. \diamond

We are especially interested in the study of the set of stochastic rate matrices of the model. The condition of Definition 1 ensures \mathcal{L} contains *enough* stochastic rate matrices (see Theorem 2 forthcoming), and it is useful to give some geometrical interpretation of this condition. To this aim, we need to recall some basic definitions on convex polyhedral cones.

Following the book by Alexandrov (2005), a *convex polyhedral cone* in \mathbb{R}^n is defined as a set

$$C = \{\lambda_1 v_1 + \dots + \lambda_r v_r : \lambda_i \geq 0\}$$

generated by some finite set of vectors v_1, \dots, v_r in \mathbb{R}^n . Such vectors are called *generators* of the cone C . The reader may note that, with this definition, every linear subspace of \mathbb{R}^n is a convex polyhedral cone. When a convex polyhedral cone contains no nonzero linear subspaces, it is said to be *strongly convex*. In this case, which has special interest for us, any minimal system of generators of the cone is unique up to multiplication with positive scalars (Alexandrov 2005). The *rays* of the cone are the non-negative spans of each vector in a minimal system of generators, and they correspond to the 1-dimensional faces of the cone (see Fig. 1 for an illustration). Farkas’s theorem ensures the polyhedral cones can be equivalently defined as the intersection of finitely many halfspaces. It follows that the intersection of any two convex polyhedral cones in \mathbb{R}^n is again a convex polyhedral cone.

Note 1 Consider a collection of vectors $X = \{X_1, \dots, X_r\}$. In what follows we will use the notation $\mathbb{F}X$ or $\langle X_1, \dots, X_r \rangle_{\mathbb{F}}$ to indicate the linear span of X over the field \mathbb{F} , where $\mathbb{F} = \mathbb{R}$ or \mathbb{C} . That is,

$$\mathbb{F}X = \langle X_1, \dots, X_r \rangle_{\mathbb{F}} := \{\lambda_1 X_1 + \dots + \lambda_r X_r : \lambda_i \in \mathbb{F}\}.$$

Of course, $\mathbb{F}X$ is a vector space. In particular, we can consider $V := \mathbb{C}X$ as a complex vector space with dimension r , or as a real vector space $V = \mathbb{R}X + \mathbb{R}(iX)$ with dimension $2r$. To distinguish these dimensions, we use the notation $\dim_{\mathbb{C}}(V) = r$ and $\dim_{\mathbb{R}}(V) = 2r$. ◊

The *dimension* of the cone C is defined as the dimension of the linear space $\mathbb{R}C = C + (-C)$ spanned by C , i.e. $\dim(C) := \dim(\mathbb{R}C)$. Of course, since a set of generators of a cone C is also a system of generators of the linear space $\mathbb{R}C$, we conclude that the number of rays of a cone is at least its dimension.

Returning to our setting, we consider the *real* vector space $\mathcal{L}_{GM}^{\mathbb{R}}$ of dimension $n(n - 1)$ spanned by the $n \times n$ elementary rate matrices $L_{ij}, i \neq j$ defined above. We denote

$$\mathcal{L}_{GM}^+ := \left\{ Q = \sum_{i \neq j} \alpha_{ij} L_{ij} : \alpha_{ij} \geq 0 \right\},$$

which is clearly a convex polyhedral cone in $\mathcal{L}_{GM}^{\mathbb{R}}$. Given a (complex) vector subspace \mathcal{L} in \mathcal{L}_{GM} , we consider

$$\mathcal{L}^+ := \mathcal{L} \cap \mathcal{L}_{GM}^+.$$

Notice that all the entries of each matrix in \mathcal{L}^+ are real and non-negative.

Theorem 2 *For any (complex) vector subspace \mathcal{L} , $\mathcal{L}^+ = \mathcal{L} \cap \mathcal{L}_{GM}^+$ is a strongly convex polyhedral cone in $\mathcal{L}_{GM}^{\mathbb{R}}$. The dimension of \mathcal{L}^+ as a cone is less than or equal to the complex dimension of \mathcal{L} , and equality holds if and only if \mathcal{L} has a stochastic generating set.*

Proof The set \mathcal{L}^+ is the intersection of two convex polyhedral cones, so it is also a convex polyhedral cone. Moreover, being contained in \mathcal{L}_{GM}^+ it is clear it contains no linear subspaces, so it is strongly convex, as required. Now, to show that the dimension of \mathcal{L}^+ is less than or equal to the complex dimension of \mathcal{L} , consider the vector space $\mathbb{C}\mathcal{L}^+$ and observe it is a subspace: $\mathbb{C}\mathcal{L}^+ \subset \mathcal{L}$. This implies $\dim_{\mathbb{C}}(\mathbb{C}\mathcal{L}^+) \leq \dim_{\mathbb{C}}(\mathcal{L})$, and since \mathcal{L}^+ contains only real vectors, we have $\dim_{\mathbb{C}}(\mathbb{C}\mathcal{L}^+) = \dim_{\mathbb{R}}(\mathbb{R}\mathcal{L}^+) = \dim(\mathcal{L}^+) \leq \dim_{\mathbb{C}}(\mathcal{L})$, as required. Now, assume \mathcal{L} has a stochastic generating set $B_{\mathcal{L}}$ so $B_{\mathcal{L}} \subset \mathcal{L}^+$ and $\mathbb{C}B_{\mathcal{L}} = \mathcal{L}$. As $B_{\mathcal{L}}$ contains only real vectors, we have $\dim_{\mathbb{R}}(\mathbb{R}B_{\mathcal{L}}) = \dim_{\mathbb{C}}(\mathbb{C}B_{\mathcal{L}}) = \dim_{\mathbb{C}}(\mathcal{L})$; and because \mathcal{L}^+ contains only real vectors and $\mathcal{L}^+ \subset \mathcal{L}$, we have $B_{\mathcal{L}} \subset \mathcal{L}^+ \subset \mathbb{R}B_{\mathcal{L}}$, so $\mathbb{R}B_{\mathcal{L}} = \mathbb{R}\mathcal{L}^+$. Together this implies $\dim_{\mathbb{R}}(\mathbb{R}B_{\mathcal{L}}) = \dim_{\mathbb{R}}(\mathbb{R}\mathcal{L}^+) = \dim(\mathcal{L}^+) = \dim_{\mathbb{C}}(\mathcal{L})$. Conversely, suppose $\dim_{\mathbb{C}}(\mathcal{L}) = \dim(\mathcal{L}^+)$. Take a generating set for $\mathbb{R}\mathcal{L}^+$ composed of vectors in \mathcal{L}^+ ; by removing vectors in this generating set, we can always assume they actually form a basis $B \subset \mathcal{L}^+$ of $\mathbb{R}\mathcal{L}^+$. Now consider the vector subspace $\mathbb{C}B \subset \mathcal{L}$ and observe $\dim_{\mathbb{C}}(\mathbb{C}B) = \dim_{\mathbb{R}}(\mathbb{R}B) = \dim_{\mathbb{R}}(\mathbb{R}\mathcal{L}^+) = \dim(\mathcal{L}^+) = \dim_{\mathbb{C}}(\mathcal{L})$. Thus $\mathbb{C}B = \mathcal{L}$, as required. □

Remark 2 The previous result implies that the vector subspace \mathcal{L} has a stochastic basis if and only if there is no drop in dimension when we restrict to the intersection with

the positive orthant \mathfrak{L}_{GM}^+ . From a practical perspective, it is the cone \mathfrak{L}^+ that contains the relevant part of the model, as only real matrices with non-negative off-diagonal entries make some sense as rate matrices of a continuous-time Markov chain. This observation justifies the requirement of stochastic basis in Definition 1. \diamond

Definition 2 The *dimension* of a Lie Markov model is the dimension of \mathfrak{L} as a complex vector space (which by virtue of Theorem 2 equals the dimension of \mathfrak{L}^+ as a cone). The *stochastic cone* of \mathfrak{L} is the convex polyhedral cone \mathfrak{L}^+ and the *rays of the model* are the rays of \mathfrak{L}^+ .

Remark 3 It is important to note that not every stochastic generating set of \mathfrak{L} is a set of generators of the cone \mathfrak{L}^+ . If this is the case and the set of generators is *minimal*, the positive linear span of each generator is a ray of the cone. \diamond

Background on group representation theory

In what follows we recall basic results from the representation theory of permutation groups $G \leq \mathfrak{S}_n$. We recommend the books by Sagan (2001) and James and Liebeck (2001) as an excellent introductions to the required material.

A (linear) *representation* of a group G is a group homomorphism $\rho : G \rightarrow GL(V) \cong GL(m, \mathbb{C})$, where V is a \mathbb{C} -vector space of dimension m . In this situation, ρ provides an *action* of G on V , and we say that V forms a *G-module*. A representation is said to be *irreducible* if it does not contain any proper G -submodules.

Let $G \leq \mathfrak{S}_n$ be a permutation group on n elements. Write $\{V_i\}_{i=1,\dots,l}$ for the irreducible G -modules and $\rho_i : G \rightarrow GL(V_i)$ for the corresponding group homomorphism. Since G is finite, any representation $\rho : G \rightarrow GL(V)$ is completely reducible and there is a decomposition of the corresponding module V into irreducible parts called *isotypic components*, so we can write (Maschke’s theorem):

$$V \cong \bigoplus_{i=1}^{\ell} c_i V_i, \tag{2}$$

where the c_i are non-negative integers specifying the number of copies of the module V_i in the decomposition of V .

Example 1 The irreducible representations of \mathfrak{S}_n are indexed by the integer partitions of n (Sagan 2001). The *defining* representation of \mathfrak{S}_n is defined on the vector space $\mathbb{C}^n = \{e_i\}_{1 \leq i \leq n} \mathbb{C}$ by $\sigma : e_i \mapsto e_{\sigma(i)}$. It decomposes as $\{n\} \oplus \{n - 1, 1\}$, where $\{n\}$ is the (one-dimensional) trivial representation and $\{n - 1, 1\}$ has dimension $n - 1$. \diamond

Example 2 After identifying the nucleotides A, G, C, T with the integers $1, 2, 3, 4$, consider \mathcal{G} as a subgroup of \mathfrak{S}_4 :

$$\mathcal{G} := \{e, (12), (34), (12)(34), (13)(24), (14)(23), (1324), (1423)\}.$$

The group \mathcal{G} has five conjugacy classes:

$$\begin{aligned}
 [e] &= \{e\}, \\
 [(12)] &= \{(12), (34)\}, \\
 [(12)(34)] &= \{(12)(34)\}, \\
 [(13)(24)] &= \{(13)(24), (14)(23)\}, \\
 [(1324)] &= \{(1324), (1423)\}.
 \end{aligned}$$

Recall the number of irreducible representations of a finite group is equal to the number of its conjugacy classes, and the sum of the dimension of each irreducible representation squared is equal to the order of the group (see Sagan 2001 for example). We conclude there are five irreducible representations of \mathcal{G} , which we denote as id , sgn , ζ_1 , ζ_2 , and ξ ; with the corresponding character table presented as Table 2. Notice the first row in the character table gives the dimension of each representation. Notice also there are four one-dimensional representations, namely id (the trivial representation), sgn (each permutation σ is mapped to $\text{sgn}(\sigma)$), ζ_1 and ζ_2 . Besides these, the representation ξ is two-dimensional. The rows of Table 2 represent the conjugacy classes of \mathcal{G} . ◊

Every irreducible module V_i of G has a *projection operator* associated to it:

$$\Theta_i(\sigma) := \frac{1}{|G|} \sum_{\sigma \in G} \overline{\chi^i(\sigma)} \rho(\sigma), \tag{3}$$

where $\chi^i : G \rightarrow \mathbb{C}$ is the *character* of the irreducible representation ρ_i defined as $\chi^i(\sigma) := \text{tr}(\rho_i(\sigma))$, i.e. the trace of the representing matrix $\rho_i(\sigma)$. These operators project a given G -module V onto its isotypic components, i.e. $\Theta_i(V) = c_i V_i$, so they can be used to compute the c_i as well as to identify generators of the components.

Of course, we can restrict ρ to any subgroup $H \leq G$, inducing an H -module structure on making a H -module of V . By virtue of Maschke’s theorem, we can also decompose V into the irreducible H -modules. Recall an irreducible representation of G does not necessarily stay irreducible when restricted to a subgroup H of G . The *branching rule* $G \downarrow H$ applies to describe the decomposition of the irreducible representations of G in terms of the irreducible representations of H (see Sagan 2001, Chap. 2.8). By applying orthogonality in the character tables of \mathfrak{S}_4 and \mathcal{G} (see Table 2), and concentrating on the conjugacy class $[(12)(34)]$ in \mathfrak{S}_4 compared to the same class in \mathcal{G} , it is straightforward to derive the group branching rules shown in Table 3.

Background on discrete group actions

Whenever a group G acts on a finite set $B = \{b_1, \dots, b_t\}$, there is a group homomorphism

$$\rho : G \rightarrow \mathfrak{S}_t. \tag{4}$$

A G -orbit in B is a subset $\mathcal{B} = \{b_{i_1}, b_{i_2}, \dots, b_{i_l}\} \subset B$ which is invariant under G and is minimal. That is

Table 2 The character tables of \mathfrak{S}_4 and $\mathcal{G} = \{e, (12), (34), (12)(34), (13)(24), (14)(23), (1324), (1423)\}$

\mathfrak{S}_4	{4}	{31}	{2 ² }	{21 ² }	{1 ⁴ }	\mathcal{G}	id	sgn	ζ_1	ζ_2	ξ
e	1	3	2	3	1	e	1	1	1	1	2
[(12)]	1	1	0	-1	-1	[(12)]	1	-1	-1	1	0
[(123)]	1	0	-1	0	1	[(12)(34)]	1	1	1	1	-2
[(12)(34)]	1	-1	2	-1	1	[(13)(24)]	1	1	-1	-1	0
[(1234)]	1	-1	0	1	-1	[(1324)]	1	-1	1	-1	0

The rows are labelled by the conjugacy classes and the columns are labelled by the irreducible characters. The character table of a group plays a key role to obtain the decomposition of any representation of that group into its irreducible representations (see (2))

Table 3 The branching rule of $\mathfrak{S}_4 \downarrow \mathcal{G}$ describes the decomposition of the irreducible representations of \mathfrak{S}_4 when restricted to the subgroup \mathcal{G}

$$\begin{array}{l}
 \mathfrak{S}_4 \downarrow \mathcal{G} : \\
 \{4\} \mapsto \text{id} \\
 \{1^4\} \mapsto \text{sgn} \\
 \{31\} \mapsto \xi + \zeta_2 \\
 \{2^2\} \mapsto \text{id} + \text{sgn} \\
 \{21^2\} \mapsto \xi + \zeta_1
 \end{array}$$

For example, $\{2^2\} \mapsto \text{id} + \text{sgn}$ means that, when restricted to \mathcal{G} , the irreducible representation $\{2^2\}$ of \mathfrak{S}_4 decomposes as one copy of the *identity* representation of \mathcal{G} , and one copy of the *sign* representation of \mathcal{G}

$$\sigma \mathcal{B} := \{b_{i_{\rho(\sigma)(1)}}, b_{i_{\rho(\sigma)(2)}}, \dots, b_{i_{\rho(\sigma)(l)}}\} = \mathcal{B}, \quad \text{for all } \sigma \in G,$$

and \mathcal{B} contains no smaller subsets with this property. From this, we can decompose B as a disjoint union of G -orbits: $B = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_k$.

If we focus on each orbit, the *orbit stabilizer theorem* (see [Bogopolski 2008](#)) states that, up to bijective correspondence, every G -orbit has the form of the quotient

$$G/H = \{[\sigma_1], \dots, [\sigma_q]\}, \quad [\sigma_i] := \sigma_i H,$$

where H is a subgroup of G and the $\sigma_i \in G$ are chosen so that the coset $\sigma_j H \neq \sigma_i H$ if $i \neq j$. The group operation of G induces an action in the finite set G/H by $\sigma : \sigma_i H \mapsto (\sigma \sigma_i) H$. Actually, H is the *stabilizer* of some element $x \in \mathcal{B}$, $G_x := \{g \in G : gx = x\}$. As $G_x \leq G$, and there are only finitely many subgroups of G , it is thus possible to give a complete list of G -orbits (up to isomorphism) by simply listing all quotients G/H with $H \leq G$. We recall we can turn the quotient G/H into a G -module by considering the vector space generated by the cosets of G/H :

$$\langle G/H \rangle_{\mathbb{C}} = \langle [e], [\sigma_2], \dots, [\sigma_q] \rangle_{\mathbb{C}} = \{v = c_1[e] + c_2[\sigma_2] + \dots + c_q[\sigma_q] : c_i \in \mathbb{C}\},$$

with the action $\sigma : v = \sum c_i [\sigma_i] \mapsto v' = \sum c_i [\sigma \sigma_i]$.

Back to the general case, the action of G on the set \mathcal{B} induces a representation of G in the vector space $\mathbb{C}\mathcal{B}$. We will refer to these as *permutation representations* and they will play a key role throughout the paper. Notice they decompose as

$$\mathbb{C}\mathcal{B} \cong \bigoplus_{i=1}^k \langle G/H_i \rangle_{\mathbb{C}}$$

where $H_i \leq G$ is the stabilizer of some element in the orbit \mathcal{B}_i . The reader should note this decomposition is not the decomposition into irreducible representations of (2). In fact it is possible to show that any nontrivial permutation representation is actually reducible.

3 Lie Markov models with prescribed symmetry

Sumner et al. (2012a) showed the search for Lie Markov models is significantly simplified by demanding the models to have some non-trivial symmetry since this reduces a potential infinity of models to just a number of special cases. The idea is to rely on imposing symmetry to assist in the search for Lie Markov models. An alternative strategy would be to enumerate all possible Lie Markov models and toss out those without the desired symmetry. However, unless the number of states is equal to 2 or 3, this approach is computationally infeasible. Thus, we are led to deal with the technicalities of this section in order to refine our search of the Lie Markov models with some prescribed symmetry. Of course, it is expected that the larger the symmetry we demand, the easier the analysis will be.

To this aim, recall the symmetric group \mathfrak{S}_n has an action on \mathcal{L}_{GM} defined on the elementary rate matrices as $\rho(\sigma) \cdot L_{ij} := L_{\sigma(i)\sigma(j)}$ for all $\sigma \in \mathfrak{S}_n$, and extended to all of \mathcal{L}_{GM} by linearity. Equivalently, the action can be defined by

$$Q = \sum_{i \neq j} \alpha_{ij} L_{ij} \mapsto \sigma \cdot Q := K_{\sigma} Q K_{\sigma}^{-1} = \sum_{i \neq j} \alpha_{ij} L_{\sigma(i)\sigma(j)}, \quad (5)$$

where K_{σ} is the *permutation matrix* associated to σ .

Definition 3 (cf. Sumner et al. 2012a) We say a Lie Markov model \mathcal{L} has the *symmetry* of the group $G \leq \mathfrak{S}_n$ if there is a basis $B_{\mathcal{L}}$ of \mathcal{L} invariant under the action of G induced by (5), that is, a basis $B_{\mathcal{L}} = \{L_1, L_2, \dots, L_d\}$ such that

$$\sigma \cdot B_{\mathcal{L}} := \left\{ K_{\sigma} L_1 K_{\sigma}^{-1}, K_{\sigma} L_2 K_{\sigma}^{-1}, \dots, K_{\sigma} L_d K_{\sigma}^{-1} \right\} = B_{\mathcal{L}}, \quad \forall \sigma \in G.$$

In this case, we will say that $B_{\mathcal{L}}$ is a *permutation basis* of \mathcal{L} .

Notice a Lie Markov model \mathcal{L} has the symmetry of G if and only if there is a permutation representation of G on \mathcal{L} , so we have a decomposition $\mathcal{L} \cong \bigoplus_{i=1}^k \langle G/H_i \rangle_{\mathbb{C}}$. A permutation basis for \mathcal{L} is then obtained by collecting a permutation basis B_i for each $\langle G/H_i \rangle_{\mathbb{C}}$ and putting them together.

Remark 4 Notice if \mathfrak{L} has the symmetry of a permutation group G , then it also has the symmetry of any subgroup $H \leq G$. \diamond

The reader is referred to [Sumner et al. \(2012a\)](#) for the statistical motivations for this definition. In a nutshell, parameter estimation under such a model is invariant under nucleotide permutations belonging to G . In particular, we have a group homomorphism

$$\rho : G \rightarrow \mathfrak{S}_d, \tag{6}$$

where the image of any permutation $\sigma \in G$ is determined by the equality $K_\sigma L_i K_\sigma^{-1} = L_{\rho(\sigma)(i)}$. Thus, for any rate matrix $Q = \sum_{i=1}^d \alpha_i L_i \in \mathfrak{L}$, we have

$$\sigma : Q = \sum_{i=1}^d \alpha_i L_i \mapsto \sum_{i=1}^d \alpha_i L_{\rho(\sigma)(i)} = \sum_{i=1}^d \alpha_{\rho(\sigma^{-1}(i))} L_i,$$

so G acts by permuting the model parameters, ie. $\alpha_i \mapsto \alpha_{\rho(\sigma^{-1}(i))}$, and hence leaves maximum likelihood estimates invariant.

Example 3 ([Sumner et al. 2012a](#)) The list of 4-state Lie Markov models with \mathfrak{S}_4 symmetry is:

1. the 1-dimensional model by [Jukes and Cantor \(1969\)](#);
2. the 3-dimensional model by [Kimura \(1981\)](#);
3. the 4-dimensional model by [Felsenstein \(1981\)](#);
4. the Kimura+Felsenstein model or ‘‘K3ST+F81’’, with dimension 6; see [Example 5](#) below ([Sumner et al. 2012](#));
5. the General Markov model, with dimension 12.

Presently, we recall the general procedure to obtain Lie Markov models with prescribed symmetry. Suppose we have a Lie Markov algebra \mathfrak{L} with dimension d and a permutation group $G \leq \mathfrak{S}_n$. We demand that \mathfrak{L} satisfies the conditions of [Definition 3](#) for the permutation group G . Then, \mathfrak{L} is provided with a basis $B_\mathfrak{L}$ which is invariant under G . As explained above, we have a decomposition of $B_\mathfrak{L}$ into G -orbits. We can then compare the irreducible G -modules that occur in the decomposition of \mathfrak{L}_{GM} to those that occur in the decomposition of $\langle G/H \rangle_{\mathbb{C}}$ for each $H \leq G$. Finally, we can attempt to construct subalgebras $\mathfrak{L} \subset \mathfrak{L}_{GM}$ with a basis $B_\mathfrak{L}$ such that $B_\mathfrak{L} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_r$ is a plausible union of orbits \mathcal{B}_i consistent with the linear decomposition of \mathfrak{L}_{GM} induced by the action of G .

The general procedure is:

1. Decompose the Lie algebra of the GM model into irreducible modules of G :

$$\mathfrak{L}_{GM} = \bigoplus_k f_k V_k, \tag{7}$$

where k labels the irreducible G -module V_k and the f_k are non-negative integers specifying the number of copies of each irreducible module in the decomposition.

2. Apply the orbit stabilizer theorem and construct the list of G -orbits, G/H , by working through the subgroups $H \leq G$. For each subgroup H , extend the orbits linearly over \mathbb{C} to the G -module $\langle G/H \rangle_{\mathbb{C}}$ and decompose this space into irreducible G -modules:

$$\langle G/H \rangle_{\mathbb{C}} \cong \oplus_k b_k^H V_k,$$

where again the b_k^H are non-negative integers.

3. Working up in dimension d , consider all unions of G -orbits $S = \bigcup_{i=1}^q (G/H_i)$ such that $|S| = \sum_{1 \leq i \leq q} |G/H_i| = d$ (where $|\cdot|$ stands for cardinality). For each S , consider its linear decomposition into irreducible G -modules

$$\langle S \rangle_{\mathbb{C}} \cong \oplus_k a_k V_k$$

where $a_k := b_k^{H_1} + b_k^{H_2} + \dots + b_k^{H_q}$, and, in order to exclude unions of G -orbits that do not occur in the linear decomposition of \mathfrak{L}_{GM} as a G -module, check $a_k \leq f_k$, for each k .

4. For each case thus identified, consider the vector space $\mathfrak{L} = \oplus_k a_k V_k$ and use explicit computation to check whether \mathfrak{L} forms a Lie algebra. If so, attempt to show it has a stochastic basis.

This procedure is guaranteed to produce all Lie Markov models with symmetry G . In Sect. 5, we will give a complete presentation of the 4-state models with purine/pyrimidine symmetry.

Remark 5 In our procedure we first look for all possible decompositions into irreducible modules for a permutation representations and we investigate how these decompositions are realised into Lie subalgebras of \mathfrak{L}_{GM} . A different approach would be to deal first with possible Lie subalgebras of \mathfrak{L}_{GM} (up to isomorphism) and then, for each isomorphism class, look for possible subalgebras which are permutation representations of G . Our experience tells us this second part is rather unfeasible and, in the last section, we adopt the procedure just explained. \diamond

Remark 6 *Equivariant* models were first introduced by Draisma and Kuttler (2008) and have also been studied by Casanellas and Fernández-Sánchez (2010), and Casanellas et al. (2012). Sumner et al. (2012a) modified slightly the definition of equivariant models to adapt it to the continuous-time Markov model setting. Under this definition, equivariant models appear as a particular case of Lie Markov models. Actually, the G -equivariant model is the Lie Markov model with G symmetry obtained when we take \mathfrak{L} to be the isotypic component of \mathfrak{L}_{GM} associated to the *trivial* or *identity* representation id (which maps each permutation to the identity map): $\mathfrak{L} = f_{\text{id}} V_{\text{id}}$ (see (7)). For example, the \mathfrak{S}_4 -equivariant model is the Lie Markov model with symmetry \mathfrak{S}_4 and decomposition $\mathfrak{L} \cong \text{id}$: it is the model by Jukes and Cantor (1969). In a similar way, we will recover the Kimura model with two parameters (Kimura 1980) as the Lie Markov model with symmetry \mathcal{G} and decomposition $\mathfrak{L} \cong 2\text{id}$ (see Model 2.2b in Sect. 5). \diamond

The stochastic cone of a Lie Markov model

We want to explore the geometry of the stochastic cone associated to a Lie Markov model with symmetry given by some permutation group $G \leq \mathfrak{S}_n$. Since the action of G on \mathfrak{L}_{GM} is as given in (5), we infer that the space \mathfrak{L}_{GM}^+ is invariant under this action, i.e. $G\mathfrak{L}_{GM}^+ = \mathfrak{L}_{GM}^+$. From this, we conclude that if $\mathfrak{L} \subset \mathfrak{L}_{GM}$ is a vector subspace which is invariant under the action of G , then the stochastic cone $\mathfrak{L}^+ = \mathfrak{L} \cap \mathfrak{L}_{GM}^+$ is invariant under G as well.

Because each permutation in G induces a linear automorphism in \mathfrak{L}_{GM} and the cone \mathfrak{L}^+ is invariant, the set of rays of the cone must also be invariant under the action of G . We infer that, after giving an ordering to the set of rays, there is a group homomorphism

$$G \rightarrow \mathfrak{S}_r, \tag{8}$$

where r is the number of rays of \mathfrak{L}^+ . From this, we can decompose the set of rays of \mathfrak{L}^+ into G -orbits, which we will refer to as *ray-orbits*. Notice in general, the above homomorphism is different from the homomorphism arising from a permutation basis, as described in (6).

Example 4 The number of rays of \mathfrak{L}_{GM}^+ is $n(n - 1)$. These rays are exactly the positive span of the elementary rate matrices L_{ij} . The group homomorphism $G \rightarrow \mathfrak{S}_r$ of (8) corresponds to the action described in (5). \diamond

Example 5 We know there is only one six-dimensional Lie Markov model with \mathfrak{S}_4 symmetry (Sumner et al. 2012a, Result 17). The Lie algebra \mathfrak{L} is the vector space sum of the model by Kimura (1981) and the model by Felsenstein (1981). It is generated by

$$W_{ij} := L_{s(ij)} + (R_i + R_j), \quad i < j, \quad i, j \in \{1, 2, 3, 4\},$$

where $R_i = \sum_{j \neq i} L_{ij}$ and $L_{s(ij)} = L_{ij} + L_{ji} + L_{kl} + L_{lk}$ with i, j, k, l all different. The reader may notice, although the six vectors W_{ij} do form a permutation basis of \mathfrak{L} , by taking the convex cone generated by them, $\{\sum \lambda_{ij} W_{ij} : \lambda_{ij} \geq 0\}$, we are not considering all the stochastic rate matrices in the model. For example, the vector R_1 is in the stochastic cone \mathfrak{L}^+ but we cannot obtain it as a positive linear combination of the vectors W_{ij} .

The reader may argue this situation occurs because of our particular choice of a permutation basis, but this will be the case no matter the permutation basis of \mathfrak{L} we consider. Actually, the stochastic cone \mathfrak{L}^+ has seven rays $\{L_\alpha, L_\beta, L_\gamma, R_1, R_2, R_3, R_4\}$ (with the notation used there: $L_\alpha = L_{s(12)}, L_\beta = L_{s(13)}, L_\gamma = L_{s(12)}$). We will find this model again in Sect. 5 of this paper as Model 6.7a. \diamond

4 Decomposition of \mathfrak{L}_{GM} as a \mathcal{G} -module

As we are especially interested in nucleotide evolution, we fix $n = 4$ and deal with the group of permutations that preserves the partitioning of nucleotides into purines and pyrimidines: $AG|CT := \{\{A, G\}, \{C, T\}\}$.

By identifying nucleotides $\{A, G, C, T\}$ with numbers $\{1, 2, 3, 4\}$, this leads to consider the subgroup \mathcal{G} of \mathfrak{S}_4 presented in Example 2:

$$\mathcal{G} := \{e, (12), (34), (12)(34), (13)(24), (14)(23), (1324), (1423)\}.$$

Of course, we expect to recover the Kimura model with three parameters in the list of Lie Markov models with this symmetry: it is Model 3.3a. However, as already noted in Remark 3, this model has a wider symmetry and in fact, it is \mathfrak{S}_4 -symmetric (see Sumner et al. 2012a).

Presently, we use the projection operators to decompose the Lie algebra of the general Markov model into the irreducible representations of \mathcal{G} .

Remark 7 The reader may notice the irreducible characters of the group \mathcal{G} in Table 2 take only real values. As a consequence, irreducible real representations remain irreducible over the complex field and all the representation theory for \mathcal{G} can be dealt over the real field. However, we prefer to keep our study over the complex as this is the field where the general theory is developed. For instance, it is important to work over the complex field when computing the full list of \mathcal{G} -submodules of \mathcal{L}_{GM} isomorphic to $\bigoplus_k a_k V_k$ (see the step 4 of the procedure of Sect. 3): to this aim, we apply that the only \mathcal{G} -endomorphisms of an irreducible module are of the form $\lambda \mathbf{1}$ (Schur’s lemma), which is known to be false if the field is not algebraically closed. \diamond

From now on, we will consider the restriction of the action of \mathfrak{S}_4 described in (5) to the group \mathcal{G} . We will denote this action by $\rho_{\mathcal{G}}$:

$$\rho_{\mathcal{G}}(\sigma) : Q \mapsto K_{\sigma} Q K_{\sigma}^{-1}. \tag{9}$$

Sumner et al. (2012a, Result 8) showed the decomposition of the \mathcal{L}_{GM} into the irreducible representations of \mathfrak{S}_4 (expressed using integer partitions of 4) is $\mathcal{L}_{GM} \cong \{4\} \oplus 2\{31\} \oplus \{2^2\} \oplus \{21^2\}$. By applying the branching rule of \mathfrak{S}_4 to \mathcal{G} (see Table 3) we obtain:

Theorem 3 *The decomposition of the 4-state general rate matrix model \mathcal{L}_{GM} into irreducible representations of \mathcal{G} is given by*

$$\mathcal{L}_{GM} \cong 2 \text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus 2 \zeta_2 \oplus 3 \xi, \tag{10}$$

where the decomposition of the dimension is given by $12 = 2 \times 1 + 1 + 1 + 2 \times 1 + 3 \times 2$.

4.1 Decomposition of the orbits of \mathcal{G} in \mathcal{L}_{GM}

Following the general scheme described in Sect. 3, our task now is to identify the Lie Markov models occurring as subalgebras of \mathcal{L}_{GM} and with symmetry \mathcal{G} . In Table 4 we present the decomposition of the orbits of \mathcal{G} . These are computed by using the orbit stabilizer theorem and projecting $\langle \mathcal{G}/H \rangle_{\mathbb{C}}$ onto the irreducible modules V_i of \mathcal{G} using the projection operators Θ_i defined in (3).

Example 6 Here we develop the case of $\mathcal{H} = \{e, (12)(34)\}$ as an illustrative example. We have $\mathcal{G} / \mathcal{H} = \{[e], [(12)], [(13)(24)], [(1324)]\}$, where $[\sigma]$ represents the coset in $\mathcal{G} / \mathcal{H}$ containing the element σ . Namely, $[e] = \{e, (12)(34)\}$, $[(12)] = \{(12), (34)\}$, $[(13)(24)] = \{(13)(24), (14)(23)\}$ and $[(1324)] = \{(1324), (1423)\}$. These cosets inherit an action of \mathcal{G} by taking $\sigma : [\sigma'] \mapsto [\sigma\sigma']$, which can be extended linearly to a linear representation of \mathcal{G} by taking the module $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}} \cong \mathbb{C}^4$. Next, we decompose $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}}$ into irreducible modules of \mathcal{G} by applying the projection operators: Θ_{id} , Θ_{sgn} , Θ_{ζ_1} , Θ_{ζ_2} and Θ_{ξ} . For example:

$$\Theta_{\text{id}}[e] = \frac{1}{8} \sum_{\sigma \in \mathcal{G}} \sigma \cdot [e] = \frac{1}{4} ([e] + [(12)] + [(13)(24)] + [(1324)]).$$

As this projection is non-zero, we conclude $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}}$ contains the trivial representation id . We can check that the image by Θ_{id} of the other coset elements gives the same projection, so $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}}$ contains id only *once*. Similarly, referring to the character table of \mathfrak{S}_4 (see Table 2), we have

$$\Theta_{\text{sgn}}[e] = \frac{1}{8} \sum_{\sigma \in \mathcal{G}} \chi^{\text{sgn}}(\sigma) \sigma \cdot [e] = \frac{1}{4} ([e] - [(12)] + [(13)(24)] - [(1324)]),$$

and we check that $\Theta_{\text{sgn}}[e] = \Theta_{\text{sgn}}[(12)] = \Theta_{\text{sgn}}[(13)(24)] = \Theta_{\text{sgn}}[(1324)]$ to learn that $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}}$ does contain a copy of the sgn representation. Similarly, we check $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}}$ contains a copy of ζ_1 and ζ_2 representations. On the other hand, we see

$$\Theta_{\xi}[e] = \frac{1}{8} \sum_{\sigma \in \mathcal{G}} \chi^{\xi}(\sigma) \cdot [e] = \frac{1}{4} ([e] - [(12)(34)]) = 0,$$

and we check $\Theta_{\xi}[(12)] = \Theta_{\xi}[(13)(24)] = \Theta_{\xi}[(1324)] = 0$ to learn $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}}$ does not contain a copy of the ξ representation. Putting this together and counting dimensions, we infer that $\langle \mathcal{G} / \mathcal{H} \rangle_{\mathbb{C}} \cong \text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2$. ◊

Proceeding as in this example, we have produced the results summarised in Table 4. It gives the decomposition of $\langle \mathcal{G} / H \rangle_{\mathbb{C}}$ into irreducible representations for each subgroup $H \leq \mathcal{G}$. The first column shows how many copies of each H occur as a subgroup in \mathcal{G} , with automorphism classes accounted for with distinct decomposition in the fourth column. For example, there are three automorphism classes of \mathbb{Z}_2 in $\mathcal{G} : \{e, (12)\} \cong \{e, (34)\}$, $\{e, (12)(34)\}$ and $\{e, (13)(24)\} \cong \{e, (14)(23)\}$, and the corresponding spaces $\langle \mathcal{G} / H \rangle_{\mathbb{C}}$ have different decomposition into irreducible modules, as shown in Table 4. Similarly, there are two “types” of $\mathfrak{S}_2 \times \mathfrak{S}_2 : \{e, (12), (34), (12)(34)\}$ and $\{e, (12)(34), (13)(24), (14)(23)\}$. Again, these two types have differing decomposition into irreducible subspaces.

Finally, Table 5 shows all possible decompositions for a \mathcal{G} -invariant subspace of \mathfrak{L}_{GM} allowed by the decomposition of \mathfrak{L}_{GM} of Theorem 3. The list is obtained by adding decompositions of \mathcal{G} -orbits (see Table 4) as long as they are allowed by the decomposition of \mathfrak{L}_{GM} as a \mathcal{G} -module (see Theorem 3). Note the decomposition (10) of \mathfrak{L}_{GM} has two copies of the trivial representation while the decomposition of each \mathcal{G} / H has only one copy.

Referring to Table 5, we conclude:

Table 4 Decomposition of the orbits of \mathcal{G} into irreducible modules

Automorphism classes of $H \leq \mathcal{G}$	$\frac{ \mathcal{G} }{ H }$	Decomp. of $\langle \mathcal{G}/H \rangle_{\mathbb{C}}$
$\{e\}$	8	(1) : $\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$
$\mathfrak{S}_2 \cong \{e, (12)\} \cong \{e, (34)\}$	4	(2) : $\text{id} \oplus \zeta_2 \oplus \xi$
$\mathfrak{S}_2 \cong \{e, (14)(23)\} \cong \{e, (13)(24)\}$	4	(3) : $\text{id} \oplus \text{sgn} \oplus \xi$
$\mathfrak{S}_2 \cong \{e, (12)(34)\}$	4	(4) : $\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2$
$\mathbb{Z}_4 \cong \{e, (1324), (12)(34), (1423)\}$	2	(5) : $\text{id} \oplus \zeta_1$
$\mathfrak{S}_2 \times \mathfrak{S}_2 \cong \{e, (12), (34), (12)(34)\}$	2	(6) : $\text{id} \oplus \zeta_2$
$\mathfrak{S}_2 \times \mathfrak{S}_2 \cong \{e, (12)(34), (13)(24), (14)(23)\}$	2	(7) : $\text{id} \oplus \text{sgn}$
\mathcal{G}	1	(8) : id

Theorem 4 *There are no Lie Markov models with purine/pyrimidine symmetry of dimension seven or eleven.*

Remark 8 Being a G -orbit, we can consider the abstract vector space generated by any ray-orbit $B = \{Q_1, \dots, Q_r\}$, that is, $\{\sum_{i=1}^r a_i [Q_i] : a_i \in \mathbb{C}\}$, where the notation $[Q_i]$ is used to emphasise the fact that we are avoiding any reference to matrix addition between the elements of the ray-orbit. The dimension of this vector space equals the number of elements in the orbit, and as a permutation representation, the decomposition into irreducible representations will be one of the decompositions shown in Table 4. On the other hand, we can also consider the vector subspace of \mathcal{L}_{GM} spanned by the matrices Q_1, \dots, Q_r . Notice that these matrices may not be linearly independent as vectors of \mathcal{L}_{GM} and the dimension of this vector subspace will be smaller than the number of them. In this case, this vector space is not a permutation representation and its decomposition into irreducible representations does not appear in Table 4. For an example of this, the reader is referred to ray-orbits $(4, \frac{1}{3}, \frac{2}{3})d, (4, \frac{1}{3}, \frac{2}{3})e, (4, \frac{1}{3}, \frac{2}{3})f$ presented in Table 8. \diamond

4.2 A convenient basis

In this section we derive a basis for the vector space of 4×4 rate matrices \mathcal{L}_{GM} where the matrices comprising the basis are organised naturally into subsets that span each of the irreducible components of the decomposition of \mathcal{L}_{GM} with respect to \mathcal{G} (as given in Theorem 3). This basis is presented in Theorem 5 below. The reader should note these basis vectors play the role of the L_{ij} when models with \mathfrak{S}_4 symmetry were considered (Sumner et al. 2012a).

Permutation vectors For each $\sigma \in \mathcal{G}, \sigma \neq e$, a permutation vector is defined as

$$L_{\sigma} := -\mathbf{1} + K_{\sigma} = \sum_{1 \leq j \leq 4} L_{j\sigma(j)}.$$

Notice that each L_{σ} is a rate matrix in \mathcal{L}_{GM} . The linear span of these vectors has dimension 5 because of the linear dependencies $L_{(12)} + L_{(34)} = L_{(12)(34)}$, and $L_{(1324)} +$

Table 5 Decompositions into irreducible modules of all possible \mathcal{G} -permutation subrepresentations of $\mathfrak{L}_{GM} \cong 2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus 2\zeta_2 \oplus 3\xi$ (see Theorem 3)

Dim.	Orbits	Decomp. into irreps.
1	(8)	id
2	2(8)	2id
	(7)	id \oplus sgn
	(5)	id \oplus ζ_1
	(6)	id \oplus ζ_2
3	(7)+(8)	2id \oplus sgn
	(5)+(8)	2id \oplus ζ_1
	(6)+(8)	2id \oplus ζ_2
4	(2)	id \oplus $\zeta_2 \oplus \xi$
	(5)+(7)	2id \oplus sgn \oplus ζ_1
	(4)	id \oplus sgn \oplus $\zeta_1 \oplus \zeta_2$
	(3)	id \oplus sgn $\oplus \xi$
	2(6)	2id \oplus 2 ζ_2
	(6)+(7)	2id \oplus sgn \oplus ζ_2
	(5)+(6)	2id \oplus $\zeta_1 \oplus \zeta_2$
5	(3)+(8)	2id \oplus sgn $\oplus \xi$
	(4)+(8)	2id \oplus sgn \oplus $\zeta_1 \oplus \zeta_2$
	(2)+(8)	2id \oplus $\zeta_2 \oplus \xi$
6	(5)+(3)	2id \oplus sgn \oplus $\zeta_1 \oplus \xi$
	(4)+(6)	2id \oplus sgn \oplus $\zeta_1 \oplus 2\zeta_2$
	(2)+(7)	2id \oplus sgn \oplus $\zeta_2 \oplus \xi$
	(2)+(5)	2id \oplus $\zeta_1 \oplus \zeta_2 \oplus \xi$
	(2)+(6)	2id \oplus 2 $\zeta_2 \oplus \xi$
8	(1)	id \oplus sgn \oplus $\zeta_1 \oplus \zeta_2 \oplus 2\xi$
	2(2)	2id \oplus 2 $\zeta_2 \oplus 2\xi$
	(2)+(4)	2id \oplus sgn \oplus $\zeta_1 \oplus 2\zeta_2 \oplus \xi$
	(2)+(3)	2id \oplus sgn \oplus $\zeta_2 \oplus 2\xi$
9	(1)+(8)	2id \oplus sgn \oplus $\zeta_1 \oplus \zeta_2 \oplus 2\xi$
10	(1)+(6)	2id \oplus sgn \oplus $\zeta_1 \oplus 2\zeta_2 \oplus 2\xi$
12	(1)+(2)	2id \oplus sgn \oplus $\zeta_1 \oplus 2\zeta_2 \oplus 3\xi$

$L_{(1423)} = L_{(13)(24)} + L_{(14)(23)}$. Moreover, the permutation vectors span a Lie algebra (Sumner et al. 2012a, Proposition 4.12):

$$[L_\sigma, L_{\sigma'}] = [-\mathbf{1} + K_\sigma, -\mathbf{1} + K_{\sigma'}] = [K_\sigma, K_{\sigma'}] = K_{\sigma\sigma'} - K_{\sigma'\sigma} = L_{\sigma\sigma'} - L_{\sigma'\sigma}.$$

The permutation vectors are useful because they provide simple expressions of generators of \mathfrak{L}_{GM} consistent with the decomposition of Theorem 3. The action $\rho_{\mathcal{G}}$ of \mathcal{G} on these permutation vectors is given by $\tau : L_\sigma \mapsto K_\tau L_\sigma K_\tau^{-1} = L_{\tau\sigma\tau^{-1}}$. Notice this action maps each matrix L_σ to $L_{\sigma'}$, where σ' is some permutation in the conjugacy class of σ . It follows that the vectors $\{L_\sigma : \sigma' \in [\sigma]\}$ span a \mathcal{G} -module, and by applying character theory we can obtain the decomposition of these \mathcal{G} -modules into

isotypic components. Moreover, a basis for these \mathcal{G} -modules consistent with these decompositions can be described with the assistance of the projection operators. The following example illustrates this procedure.

Example 7 Consider the 2-dimensional subspace $S = \langle L_{(12)}, L_{(34)} \rangle_{\mathbb{C}}$ corresponding to the conjugacy class $[(12)] = \{(12), (34)\}$, and the representation $\rho_{\mathcal{G}} : \mathcal{G} \rightarrow GL(S)$ induced by the action just defined. It is straightforward to check $\tau(12)\tau^{-1} = (12)$, $\tau(34)\tau^{-1} = (34)$ if $\tau \in \{e, (12), (12)(34)\}$, while $\tau(12)\tau^{-1} = (34)$, $\tau(34)\tau^{-1} = (12)$ if $\tau \in \{(13)(24), (1324)\}$. Adopting matrix notation, we obtain

$$\begin{aligned} \rho_{\mathcal{G}}(e) &= \rho_{\mathcal{G}}((12)) = \rho_{\mathcal{G}}((12)(34)) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and} \\ \rho_{\mathcal{G}}((13)(24)) &= \rho_{\mathcal{G}}((1324)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

If χ denotes the character associated with $\rho_{\mathcal{G}}$, we infer

$$\chi(e) = \chi((12)) = \chi((12)(34)) = 2, \quad \text{and} \quad \chi((13)(24)) = \chi((1324)) = 0.$$

By virtue of the character table of \mathcal{G} (see Table 2), we infer $S \cong \text{id} \oplus \xi_2$, and applying the projection operators (see (3)):

$$\begin{aligned} \Theta_{\text{id}}(L_{(12)}) &= \Theta_{\text{id}}(L_{(34)}) = \frac{1}{2}(L_{(12)} + L_{(34)}); \\ \Theta_{\xi_2}(L_{(12)}) &= \Theta_{\xi_2}(L_{(34)}) = \frac{1}{2}(L_{(12)} - L_{(34)}). \end{aligned}$$

◇

Proceeding in this way for each conjugacy class of \mathcal{G} (excluding the trivial class), we identify the following \mathcal{G} -modules and decompositions:

$$\begin{aligned} \langle L_{(12)}, L_{(34)} \rangle_{\mathbb{C}} &\cong \text{id} \oplus \xi_2, & \langle L_{(12)(34)} \rangle_{\mathbb{C}} &\cong \text{id}, \\ \langle L_{(13)(24)}, L_{(14)(23)} \rangle_{\mathbb{C}} &\cong \text{id} \oplus \text{sgn}, & \langle L_{(1324)}, L_{(1423)} \rangle_{\mathbb{C}} &\cong \text{id} \oplus \xi_1. \end{aligned}$$

For future convenience, we keep the vectors obtained by applying the projection operators to these decompositions. From now on, we will use the following notation

$$\begin{aligned} B_1^{\text{id}} &:= L_{(12)(34)}, & B_2^{\text{id}} &:= L_{(13)(24)} + L_{(14)(23)}, \\ B^{\text{sgn}} &:= L_{(13)(24)} - L_{(14)(23)}, & B^{\xi_1} &:= L_{(1423)} - L_{(1324)}, \\ B_1^{\xi_2} &:= L_{(12)} - L_{(34)}; \end{aligned}$$

where the superscript indicates which irreducible \mathcal{G} -module each vector belongs to.

Cherry vectors Referring to Table 4 and the permutation representation spanned by the “cherries” $\{1, 2\}$ and $\{3, 4\}$, we introduce the matrices

$$\begin{aligned} \text{Ch}_{12} &:= L_{13} + L_{14} + L_{23} + L_{24}, \\ \text{Ch}_{34} &:= L_{31} + L_{32} + L_{41} + L_{42}, \end{aligned}$$

and obtain $\langle \text{Ch}_{12}, \text{Ch}_{34} \rangle_{\mathbb{C}} \cong \text{id} \oplus \zeta_2$. The action of \mathcal{G} on each of these vectors is given by $\tau : \text{Ch}_{ij} \mapsto \text{Ch}_{\tau(i)\tau(j)}$. Notice that $\text{Ch}_{12} + \text{Ch}_{34} = B_2^{\text{id}}$. By applying the projection operator Θ_{ζ_2} , we see that $B_2^{\zeta_2} := \text{Ch}_{12} - \text{Ch}_{34}$ accounts for the second copy of ζ_2 .

Row-sum and twisted vectors Keeping the notation used by Sumner et al. (2012a), define the row-sum vectors

$$R_i := \sum_{j:1 \leq i \neq j \leq 4} L_{ij}.$$

The action $\rho_{\mathcal{G}}$ of \mathcal{G} on each of these is $\sigma : R_i \mapsto R_{\sigma(i)}$, and it is isomorphic to the restriction of the defining representation of \mathfrak{S}_4 to \mathcal{G} . Therefore, the (invariant) subspace generated by the row-sum vectors is isomorphic to $\text{id} \oplus \{3, 1\}$, restricted to the subgroup \mathcal{G} . By applying the branching rule $\mathfrak{S}_4 \downarrow \mathcal{G}$ given in Table 3, we obtain

$$\langle R_1, R_2, R_3, R_4 \rangle_{\mathbb{C}} \cong \text{id} \oplus \zeta_2 \oplus \xi.$$

Obviously, we have $R_1 + R_2 + R_3 + R_4 = B_1^{\text{id}} + B_2^{\text{id}}$ and $(R_1 + R_2) - (R_3 + R_4) = B_1^{\zeta_2} + B_2^{\zeta_2}$. By applying the projection operator Θ_{ξ} we find that $\langle R_1 - R_2, R_3 - R_4 \rangle_{\mathbb{C}}$ accounts for a copy of the ξ representation. We define

$$B_1^{\xi} := R_1 - R_2, \quad B_2^{\xi} := R_3 - R_4.$$

Next, define the *twisted vectors* as

$$\begin{aligned} H_i &:= L_{ik} + L_{il} + L_{ji}, \\ V_i &:= L_{ki} + L_{li} + L_{ij}, \end{aligned}$$

where $\{\{i, j\}, \{k, l\}\} = \{\{1, 2\}, \{3, 4\}\}$. For example, $V_2 = L_{21} + L_{32} + L_{42}$ and $H_3 = L_{31} + L_{32} + L_{43}$. The action $\rho_{\mathcal{G}}$ of \mathcal{G} on these vectors is given by $\sigma : V_i \mapsto V_{\sigma(i)}$ and $\sigma : H_i \mapsto H_{\sigma(i)}$, again we have with we are dealing with the restriction of the defining representation of \mathfrak{S}_4 to \mathcal{G} . As above,

$$\langle V_1, V_2, V_3, V_4 \rangle_{\mathbb{C}} \cong \langle H_1, H_2, H_3, H_4 \rangle_{\mathbb{C}} \cong \text{id} \oplus \zeta_2 \oplus \xi.$$

Notice that $\sum_i H_i = \sum_i V_i = B_1^{\text{id}} + B_2^{\text{id}}$, $(H_1 + H_2) - (H_3 + H_4) = B_1^{\zeta_2} + B_2^{\zeta_2}$ and $(V_1 + V_2) - (V_3 + V_4) = B_1^{\zeta_2} - B_2^{\zeta_2}$. By applying the projection operator Θ_{ξ}

in $\langle V_1, V_2, V_3, V_4 \rangle_{\mathbb{C}}$ and $\langle H_1, H_2, H_3, H_4 \rangle_{\mathbb{C}}$, we find that $\langle H_1 - H_2, H_3 - H_4 \rangle_{\mathbb{C}}$ and $\langle V_1 - V_2, V_3 - V_4 \rangle_{\mathbb{C}}$ account for the two other copies of ξ , so we define

$$\begin{aligned} B_3^{\xi} &:= H_1 - H_2, & B_5^{\xi} &:= V_1 - V_2, \\ B_4^{\xi} &:= H_3 - H_4, & B_6^{\xi} &:= V_3 - V_4. \end{aligned}$$

Putting all of these results together:

Theorem 5 *The Lie algebra \mathfrak{L}_{GM} can be expressed as*

$$\begin{aligned} \mathfrak{L}_{GM} &= \langle \{L_{ij}\}_{1 \leq i \neq j \leq 4} \rangle_{\mathbb{C}} \\ &= \langle \{L_{\sigma}\}_{\sigma \in \mathcal{G}, \sigma \neq e} \cup \{\text{Ch}_{12}, \text{Ch}_{34}\} \cup \{R_i\}_{1 \leq i \leq 4} \cup \{H_j\}_{1 \leq j \leq 4} \cup \{V_k\}_{1 \leq k \leq 4} \rangle_{\mathbb{C}}, \end{aligned}$$

with linear dependencies

$$\begin{aligned} L_{(12)} + L_{(34)} &= L_{(12)(34)}, \\ L_{(13)(24)} + L_{(14)(23)} &= L_{(1324)} + L_{(1423)} = \text{Ch}_{12} + \text{Ch}_{34}, \\ H_1 + H_2 &= R_1 + R_2 = \text{Ch}_{12} + L_{(12)}, \\ H_3 + H_4 &= R_3 + R_4 = \text{Ch}_{34} + L_{(34)}, \\ V_1 + V_2 &= \text{Ch}_{34} + L_{(12)}, \\ V_3 + V_4 &= \text{Ch}_{12} + L_{(34)}. \end{aligned}$$

A basis for \mathfrak{L}_{GM} consistent with the decomposition of Theorem 3 is given by

$$\begin{aligned} B_1^{\text{id}} &= L_{(12)(34)}, & B_1^{\xi} &= R_1 - R_2, \\ B_2^{\text{id}} &= L_{(13)(24)} + L_{(14)(23)}, & B_2^{\xi} &= R_3 - R_4, \\ B^{\text{sgn}} &= L_{(13)(24)} - L_{(14)(23)}, & B_3^{\xi} &= H_1 - H_2, \\ B_1^{\zeta_1} &= L_{(1324)} - L_{(1423)}, & B_4^{\xi} &= H_3 - H_4, \\ B_1^{\zeta_2} &= L_{(12)} - L_{(34)}, & B_5^{\xi} &= V_1 - V_2, \\ B_2^{\zeta_2} &= \text{Ch}_{12} - \text{Ch}_{34}, & B_6^{\xi} &= V_3 - V_4; \end{aligned}$$

where $\langle B_1^{\xi}, B_2^{\xi} \rangle_{\mathbb{C}}$, $\langle B_3^{\xi}, B_4^{\xi} \rangle_{\mathbb{C}}$ and $\langle B_5^{\xi}, B_6^{\xi} \rangle_{\mathbb{C}}$ are the three copies of ξ in \mathfrak{L}_{GM} . With respect to this basis, the Lie algebra structure of \mathfrak{L}_{GM} is summarised in Table 7.

5 The list of Lie Markov models with purine/pyrimidine symmetry

We proceed to give the list of Lie Markov models with purine/pyrimidine symmetry, working up in dimension $d \leq 12$. For each d , Table 5 lists all the possible decompositions allowed by Theorem 3. For each decomposition, all possible complex Lie subalgebras \mathfrak{L} of \mathfrak{L}_{GM} are obtained by direct computation using code written by the authors and implemented in the open-source mathematical software SAGE (Stein et al. 2012) (this code is available online at the website Fernández-Sánchez 2013). For each Lie algebra, we then impose that it has a stochastic basis (see Definition 1). Since

a matrix $B_{a,b}^{id} = aB_1^{id} + bB_2^{id}$, with $a, b > 0$, has all its non-diagonal entries positive, the reader can notice that the above condition is guaranteed if \mathcal{L} contains such a matrix for, in this case, if $\{B_1, \dots, B_t\}$ is a basis for \mathcal{L} , then a stochastic basis for \mathcal{L} is given by $\{B_1 + \lambda B_{a,b}^{id}, \dots, B_t + \lambda B_{a,b}^{id}\}$ as long as $\lambda > 0$ is large enough.

For each model in the list, we describe a basis for the Lie algebra in terms of the vectors introduced in the Sect. 4 and the rays of the stochastic cone arranged in orbits (see Table 8). Both data are required to completely describe the model. The general form of the stochastic rate matrix, as well as a permutation basis (a basis invariant under the action of \mathcal{G}), is also shown when it is not too complicated. In particular, stochastic rate matrices are presented as linear combinations of the rays with non-negative coefficients. Since the rays are the generators of the stochastic cone, every stochastic rate matrix in the model can be expressed in this way (the reader should notice that in general, we cannot write down all the stochastic rate matrices of a model in terms of the same permutation basis if we require the coefficients to be non-negative). The name of each model has the form “ $d.r$ ”, where d is the dimension of the model and r is the number of rays of the corresponding stochastic cone (in particular, $d \leq r$). In case there is more than one model with a given dimension and number of rays, we will differentiate them by using letters: for example, $5.7a$, $5.7b$ and so on.

Note 2 Throughout the following list, we adopt the notation $X_{ij}^+ = X_i + X_j$ and $X_{ij}^- = X_i - X_j$, for $i, j \in \{1, 2, 3, 4\}$ and $X \in \{R, H, V\}$. ◇

Ray-orbits The rays of the stochastic cones of the forthcoming models appear in orbits of cardinality 1, 2, 4 and 8 (as is demanded by the orbit-stabilizer theorem) that we call *ray-orbits*. A system of generators for any of these ray-orbits is obtained as the \mathcal{G} -orbit of a rate matrix Q in any of the rays of the family. Notice incidentally the action of \mathcal{G} preserves the total sums of transition rates and of transversions rates of the rate matrices within the \mathcal{G} -orbit.

For the Lie Markov models with symmetry \mathcal{G} , we explicitly describe the rays of the corresponding stochastic cone arranged in ray-orbits. In order to denote and compare these ray-orbits in a convenient fashion, we first *normalize* the generators of the rays, and take rate matrices whose trace is equal to -1 (recall the absolute value of the trace of the rate matrix can be understood as the expected number of changes in one unit of time under the Markov process). Then, taking into account that the sum of transition rates and of transversion rates is constant, each ray-orbit is referred to as “ $(r, \frac{s}{s+v}, \frac{v}{s+v})$ ”, where

- r is the number of rays in the orbit: 1, 2, 4 or 8;
- s is the sum of the transition rates in (any matrix of) the orbit;
- v is the sum of the transversion rates in (any matrix of) the orbit.

The reader is referred to Table 8 for the whole list of ray-orbits arising in Lie Markov models with \mathcal{G} -symmetry.

Note 3 From now on, we write $B^{id} := B_1^{id} + B_2^{id}$, $B^{\zeta_2} := B_1^{\zeta_2} + B_2^{\zeta_2}$. ◇

DIMENSION ONE

From Table 4 we see that there is only one abstract orbit of \mathcal{G} with cardinality one, and it has decomposition id .

(id): The general Markov model contains two copies of the trivial representation, so we can consider the subspace generated by any linear combination $aB_1^{\text{id}} + bB_2^{\text{id}}$. Moreover, since $[B_1^{\text{id}}, B_2^{\text{id}}] = 0$, we see the subspace generated by any $aB_1^{\text{id}} + bB_2^{\text{id}}$, is a Lie algebra for any fixed choice $a, b \in \mathbb{C}$. When we request these spaces to have a stochastic basis, we have to restrict to the condition $a, b \geq 0$. Therefore, we conclude:

Theorem 6 *In the 4-state case, there is a continuum of one-dimensional Lie Markov models with \mathcal{G} symmetry and decomposition id . Each model in the family has the form*

$$\mathcal{L} = \langle B_{a,b}^{\text{id}} \rangle_{\mathbb{C}},$$

where $a + b = 1, a, b \geq 0$ and

$$B_{a,b}^{\text{id}} := aB_1^{\text{id}} + bB_2^{\text{id}} = \begin{pmatrix} * & a & b & b \\ a & * & b & b \\ b & b & * & a \\ b & b & a & * \end{pmatrix}.$$

where we use $*$ to indicate the diagonal entry needed for the column to sum to zero.

Remark 9 This result is not completely satisfactory as all these models will appear as 1-dimensional Lie subalgebras of the 2-dimensional Lie Markov model $\langle B_1^{\text{id}}, B_2^{\text{id}} \rangle_{\mathbb{C}}$. This situation is quite general and we will avoid the consequent redundancy in the present list by considering families of Lie Markov models depending on some parameters as submodels of a Lie Markov models with larger dimension. Then, the family of models in Theorem 6 should be regarded as a Lie Markov model with decomposition 2id .

On the other hand, notice that if we expand the symmetry and request the models in the family of Theorem 6 to have the symmetry of \mathcal{S}_4 , we are lead to the constraint $a = b$, which corresponds to the model by [Jukes and Cantor \(1969\)](#). Of course, this model already appeared as a Lie Markov model with symmetry \mathcal{S}_4 ([Sumner et al. 2012a](#)). \diamond

Model 1.1 Take $\mathcal{L} = \langle B^{\text{id}} \rangle_{\mathbb{C}}$. The stochastic cone has only one ray, spanned by B^{id} . Therefore, in this case, we only have a ray-orbit. We refer to it by the ray-orbit $(1, \frac{1}{3}, \frac{2}{3})$ (see Table 8). The generic stochastic rate matrix is

$$\begin{pmatrix} * & 1 & 1 & 1 \\ 1 & * & 1 & 1 \\ 1 & 1 & * & 1 \\ 1 & 1 & 1 & * \end{pmatrix}.$$

DIMENSION TWO

($\text{id} \oplus \text{sgn}$): We have $[\mathbf{B}_1^{\text{id}}, \mathbf{B}^{\text{sgn}}] = [\mathbf{B}_2^{\text{id}}, \mathbf{B}^{\text{sgn}}] = 0$, so for any fixed $a, b \geq 0$ with $a + b = 1$ and $b \neq 0$, there is a well-defined Lie Markov model:

$$\mathcal{L} = \langle \mathbf{B}_{a,b}^{\text{id}}, \mathbf{B}^{\text{sgn}} \rangle_{\mathbb{C}} \cong \text{id} \oplus \text{sgn}.$$

The condition $b \neq 0$ is needed to ensure that the dimension of the stochastic cone is equal to the dimension of the Lie algebra. As in Remark 9, these models are considered as submodels of the model with decomposition $2\text{id} \oplus \text{sgn}$ (see Model 3.3a).

($\text{id} \oplus \zeta_1$): Since $[\mathbf{B}_1^{\text{id}}, \mathbf{B}^{\zeta_1}] = [\mathbf{B}_2^{\text{id}}, \mathbf{B}^{\zeta_1}] = 0$, we find that, for any choice of $a, b \geq 0$ with $a + b = 1$ and $b \neq 0$,

$$\mathcal{L} = \langle \mathbf{B}_{a,b}^{\text{id}}, \mathbf{B}^{\zeta_1} \rangle_{\mathbb{C}} \cong \text{id} \oplus \zeta_1,$$

provides a 2-dimensional Lie Markov model. These are submodels of the 3-dimension model with decomposition $2\text{id} + \zeta_1$ (see Model 3.3b).

($\text{id} \oplus \zeta_2$): We find the same situation for this decomposition. The following Lie algebras will appear as submodels of the 3-dimensional model indicated:

$$\mathcal{L} = \langle \mathbf{B}_{a,b}^{\text{id}}, \mathbf{B}^{\zeta_2} \rangle_{\mathbb{C}} \text{ is a submodel of Model 3.4;}$$

$$\mathcal{L} = \langle \mathbf{B}_{a,b}^{\text{id}}, \mathbf{B}_1^{\zeta_2} \rangle_{\mathbb{C}} \text{ is a submodel of Model 3.3c.}$$

As a special case of the last family of models, when we take $a = 1$ and $b = 0$, we obtain the following model:

Model 2.2a $\mathcal{L} = \langle \mathbf{B}_1^{\text{id}}, \mathbf{B}_1^{\zeta_2} \rangle_{\mathbb{C}}$. The stochastic cone has two rays generated by $L_{(12)}$ and $L_{(34)}$, which form the ray-orbit $(2, 1, 0)$ of Table 8. The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha & 0 & 0 \\ \alpha & * & 0 & 0 \\ 0 & 0 & * & \beta \\ 0 & 0 & \beta & * \end{pmatrix}, \quad \alpha, \beta \geq 0.$$

This model gives a reducible Markov chain, that is, it is not possible to get to some states from some other states. We see that the purine states A and G communicate with each other, and the same for the pyrimidine states C and T (transitions) while no replacement between purines are pyrimidines (transversions) is allowed.

Apart from these models, our analysis of 1-dimensional Lie Markov models produces another 2-dimensional model with decomposition 2id .

(2id): Of course, there is only one possible model with this decomposition. Namely,

Model 2.2b $\mathcal{L} = \langle \mathbf{B}_1^{\text{id}}, \mathbf{B}_2^{\text{id}} \rangle_{\mathbb{C}}$. If we focus on the stochastic rate matrices, we find a cone with 2 rays, corresponding to the (see Table 8): ray-orbit $(1, 1, 0) = \{L_{(12)(34)}\}$,

and ray-orbit $(1, 0, 1) = \{L_{(13)(24)} + L_{(14)(23)}\}$. The Lie algebra is abelian and the stochastic rate matrices for this model are given by

$$Q = \begin{pmatrix} * & \alpha & \beta & \beta \\ \alpha & * & \beta & \beta \\ \beta & \beta & * & \alpha \\ \beta & \beta & \alpha & * \end{pmatrix}, \quad \alpha, \beta \geq 0.$$

Permutation basis: $L_{(12)(34)}, L_{(13)(24)} + L_{(14)(23)}$.

This model corresponds to the Kimura model with 2 parameters (Kimura 1980).

DIMENSION THREE

$(2\text{id} \oplus \text{sgn})$: There is only one model with this decomposition:

Model 3.3a $\mathcal{L} = \langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}} \rangle_{\mathbb{C}}$ is an abelian Lie Markov model. The stochastic cone has 3 rays in 2 ray-orbits: ray-orbit $(1, 1, 0) = \{L_{(12)(34)}\}$, and ray-orbit $(2, 0, 1)a = \{L_{(13)(24)}, L_{(14)(23)}\}$ (see Table 8). The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha & \beta & \gamma \\ \alpha & * & \gamma & \beta \\ \beta & \gamma & * & \alpha \\ \gamma & \beta & \alpha & * \end{pmatrix}, \quad \alpha, \beta, \gamma \geq 0.$$

Permutation basis: $L_{(12)(34)}, L_{(13)(24)}, L_{(14)(23)}$.

Of course, this is the Kimura model with 3 parameters (Kimura 1981). Note this is the group-based model corresponding to $\mathbb{Z}_2 \times \mathbb{Z}_2 \cong \{e, (12)(34), (13)(24), (14)(23)\}$.

$(2\text{id} \oplus \zeta_1)$: There is only one model with this decomposition:

Model 3.3b $\mathcal{L} = \langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_1} \rangle_{\mathbb{C}}$ is a 3-dimensional abelian Lie Markov model. The stochastic cone has 3 rays, in 2 ray-orbits: ray-orbit $(1, 1, 0) = \{L_{(12)(34)}\}$, and ray-orbit $(2, 0, 1)b = \{L_{(1324)}, L_{(1423)}\}$. The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha & \beta & \gamma \\ \alpha & * & \gamma & \beta \\ \gamma & \beta & * & \alpha \\ \beta & \gamma & \alpha & * \end{pmatrix}, \quad \alpha, \beta, \gamma \geq 0.$$

Permutation basis: $L_{(12)(34)}, L_{(1324)}, L_{(1423)}$.

Note this is the group-based model corresponding to $\mathbb{Z}_4 \cong \{e, (1324), (12)(34), (1423)\}$. This new model may be regarded as a “twisted” version of the Kimura model with three parameters.

$(2\text{id} \oplus \zeta_2)$: There are two models with this decomposition:

Model 3.3c $\mathcal{L} = \langle B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2} \rangle_{\mathbb{C}}$ is a 3-dimensional abelian Lie algebra. The stochastic cone has 3 rays, in 2 ray-orbits: ray-orbit $(1, 0, 1) = \{L_{(13)(24)} + L_{(14)(23)}\}$,

and ray-orbit $(2, 1, 0) = \{L_{(12)}, L_{(34)}\}$. The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha & \beta & \beta \\ \alpha & * & \beta & \beta \\ \beta & \beta & * & \gamma \\ \beta & \beta & \gamma & * \end{pmatrix}, \quad \alpha, \beta, \gamma \geq 0.$$

Permutation basis: $L_{(12)}, L_{(34)}, L_{(1324)} + L_{(1423)}$.

Model 3.4 $\mathcal{L} = \langle B_1^{id}, B_2^{id}, B^{\zeta_2} \rangle_{\mathbb{C}}$ is a 3-dimensional Lie algebra. The stochastic cone has 4 rays, in 3 ray-orbits: ray-orbit $(1, 0, 1) = \{L_{(13)(24)} + L_{(14)(23)}\}$, ray-orbit $(1, 1, 0) = \{L_{(12)(34)}\}$, and ray-orbit $(2, 1/3, 2/3) = \{R_{12}^+, R_{34}^+\}$. This is the first model with \mathcal{G} symmetry where the number of rays is larger than the dimension of the model. It is also the first case where the Lie algebra \mathcal{L} is not abelian: the Lie algebra structure is given by

$$\begin{aligned} [L_{(13)(24)} + L_{(14)(23)}, R_{ij}^+] &= R_{kl}^+ - R_{ij}^+, & [L_{(12)(34)}, R_{ij}^+] &= 0, \\ [L_{(13)(24)} + L_{(14)(23)}, L_{(12)(34)}] &= 0, & [R_{ij}^+, R_{kl}^+] &= R_{ij}^+ - R_{kl}^+, \end{aligned}$$

for $\{ij, kl\} = \{12, 34\}$. The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha + \gamma & \beta + \gamma & \beta + \gamma \\ \alpha + \gamma & * & \beta + \gamma & \beta + \gamma \\ \beta + \delta & \beta + \delta & * & \alpha + \delta \\ \beta + \delta & \beta + \delta & \alpha + \delta & * \end{pmatrix}, \quad \alpha, \beta, \gamma, \delta \geq 0.$$

Permutation basis: $L_{(12)(34)}, R_{12}^+, R_{34}^+$.

DIMENSION FOUR

Lie Markov models with this dimension appear as special submodels of forthcoming models 5.7a, 5.7b and 5.7c with decomposition $2id \oplus sgn \oplus \xi$, when we restrict the identity component of their Lie algebra \mathcal{L} to a subspace $\langle B_{a,b}^{id} \rangle_{\mathbb{C}}$ with $a, b \geq 0$. The reader can check that depending on the values of a and b the number of rays of the cones of these models may vary.

$(id \oplus sgn \oplus \zeta_1 \oplus \zeta_2)$: The models with this decomposition appear as special cases of Model 5.6a with decomposition $2id \oplus sgn \oplus \zeta_1 \oplus \zeta_2$ (see Remark 9).

$(id \oplus \zeta_2 \oplus \xi)$: Similarly, these models are special cases of the models 5.6b, 5.11a, 5.11b, 5.11c and 5.16 with decomposition $2id \oplus \zeta_2 \oplus \xi$. As a particular case, if we request these models to have \mathcal{S}_4 symmetry, we obtain the restriction $a = b$, leading to the model by Felsenstein (1981):

Model 4.4a $\mathcal{L} = \langle B^{id}, B^{\zeta_2}, B_1^{\xi}, B_2^{\xi} \rangle_{\mathbb{C}}$ is a 4-dimensional Lie algebra. The stochastic cone has 4 rays in one single ray-orbit: $(4, \frac{1}{3}, \frac{2}{3})a = \{R_1, R_2, R_3, R_4\}$, and the Lie algebra structure is given by $[R_i, R_j] = R_i - R_j$. The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha & \alpha & \alpha \\ \beta & * & \beta & \beta \\ \gamma & \gamma & * & \gamma \\ \delta & \delta & \delta & * \end{pmatrix}, \quad \alpha, \beta, \gamma, \delta \geq 0.$$

Permutation basis: R_1, R_2, R_3, R_4 .

$(2\text{id} \oplus 2\zeta_2)$: There is only one model with this decomposition.

Model 4.4b Take $\mathfrak{L} = \langle B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_2^{\zeta_2} \rangle_{\mathbb{C}}$. The stochastic cone has 4 rays, in 2 ray-orbits: ray-orbit $(2, 0, 1)c = \{\text{Ch}_{12}, \text{Ch}_{34}\}$, and ray-orbit $(2, 1, 0) = \{L_{(12)}, L_{(34)}\}$. The Lie algebra is given by

$$\begin{aligned} [L_{(12)}, L_{(34)}] &= 0, \\ [L_{(12)}, \text{Ch}_{ij}] &= [L_{(34)}, \text{Ch}_{ij}] = 0, \quad ij \in \{12, 34\} \\ [\text{Ch}_{12}, \text{Ch}_{34}] &= 2(\text{Ch}_{34} - \text{Ch}_{12}) + 2(L_{(34)} - L_{(12)}). \end{aligned}$$

The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha & \beta & \beta \\ \alpha & * & \beta & \beta \\ \gamma & \gamma & * & \delta \\ \gamma & \gamma & \delta & * \end{pmatrix}, \quad \alpha, \beta, \gamma, \delta \geq 0$$

Permutation basis: $L_{(12)}, L_{(34)}, \text{Ch}_{12}, \text{Ch}_{34}$.

$(2\text{id} \oplus \text{sgn} \oplus \zeta_2)$: There is only one model with this decomposition.

Model 4.5a $\mathfrak{L} = \langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_2} \rangle_{\mathbb{C}}$ is a 4-dimensional Lie algebra. The stochastic cone has five rays spanned, in three ray-orbits: $(1, 1, 0)$, $(2, 0, 1)a$ and $(2, \frac{1}{3}, \frac{2}{3})$. The general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha + \delta & \beta + \delta & \gamma + \delta \\ \alpha + \delta & * & \gamma + \delta & \beta + \delta \\ \beta + \varepsilon & \gamma + \varepsilon & * & \alpha + \varepsilon \\ \gamma + \varepsilon & \beta + \varepsilon & \alpha + \varepsilon & * \end{pmatrix}, \quad \alpha, \beta, \gamma, \delta, \varepsilon \geq 0.$$

Permutation basis: $R_{12}^+, R_{34}^+, L_{(13)(24)}, L_{(14)(23)}$.

$(2\text{id} \oplus \zeta_1 \oplus \zeta_2)$: There is only one model with this decomposition.

Model 4.5b $\mathfrak{L} = \langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_1}, B^{\zeta_2} \rangle_{\mathbb{C}}$ is a 4-dimensional Lie algebra. The stochastic cone has five rays, in three ray-orbits: $(1, 1, 0)$, $(2, 0, 1)b$ and $(2, \frac{1}{3}, \frac{2}{3})$. Then, the general stochastic rate matrix is

$$\begin{pmatrix} * & \alpha + \delta & \beta + \delta & \gamma + \delta \\ \alpha + \delta & * & \gamma + \delta & \beta + \delta \\ \gamma + \varepsilon & \beta + \varepsilon & * & \alpha + \varepsilon \\ \beta + \varepsilon & \gamma + \varepsilon & \alpha + \varepsilon & * \end{pmatrix}, \quad \alpha, \beta, \gamma, \delta, \varepsilon \geq 0.$$

Permutation basis: $R_{12}^+, R_{34}^+, L_{(1324)}, L_{(1423)}$.

The remaining models, with dimensions 5–12, are presented in the Table 6, with a complete list in explicit form available as Supplementary Material online. The first column of the table gives the name of the model and the second column gives a basis for the corresponding Lie subalgebra. The third column gives the ray-orbits for the stochastic cone of that Lie subalgebra.

5.1 Some remarks

We conclude with some remarks and comments about the previous models.

Remark 10 – Model 5.6b can be regarded as the vector sum of the model by Kimura (1980) and the model by Felsenstein (1981).

- Model 6.7a already appeared as a model with \mathfrak{S}_4 symmetry under the name $K3ST + F81$ (see Example 3). A permutation basis consistent with the symmetry \mathfrak{S}_4 is given by the vectors W_{ij} of Example 5 above (see Sumner et al. 2012a). Another permutation basis for this model is $\{L_{(13)(24)}, L_{(14)(23)}, R_1, R_2, R_3, R_4\}$, which is not invariant under the action of \mathfrak{S}_4 .
- Model 9.20b is actually invariant under the action of the whole symmetric group \mathfrak{S}_4 , with Schur decomposition $\{4\} \oplus \{2^2\} \oplus \{31\} \oplus \{21^2\}$. However, it did not appeared in the list of Lie Markov models with \mathfrak{S}_4 symmetry given in Sumner et al. (2012a) as it does not have a permutation basis for that group (see Definition 3). This model is obtained by taking the set of all doubly stochastic (but otherwise unrestricted) rate matrices.
- Of course, Model 12.12 is the general Markov model and we include it in the list for completion.

◇

Remark 11 The reader may notice the resemblance of Model 5.6b with the model by Hasegawa et al. (1988):

$$Q_{5.6b} = \begin{pmatrix} * & a+x & b+x & b+x \\ a+y & * & b+y & b+y \\ b+z & b+z & * & a+z \\ b+t & b+t & a+t & * \end{pmatrix} \quad Q_{HKY} = \begin{pmatrix} * & \pi_A\alpha & \pi_A\beta & \pi_A\beta \\ \pi_G\alpha & * & \pi_G\beta & \pi_G\beta \\ \pi_C\beta & \pi_C\beta & * & \pi_C\alpha \\ \pi_T\beta & \pi_T\beta & \pi_T\alpha & * \end{pmatrix},$$

where $\pi_A + \pi_C + \pi_G + \pi_T = 1$, and all these parameters are non-negative. Although the rates of these models depend on the parameters in a different way, the rate-matrices Q_{HKY} and $Q_{5.6b}$ have the same structure. It is interesting to notice that the form of the non-diagonal entries of $Q_{5.6b}$ arises from the corresponding entries in Q_{HKY} just by applying minus the logarithm, producing the following correspondence between the parameters of both models

$$\begin{aligned} x &= -\log(\pi_A), & y &= -\log(\pi_G), & z &= -\log(\pi_C), & t &= -\log(\pi_T), \\ a &= -\log(\alpha), & b &= -\log(\beta). \end{aligned}$$

Table 6 List of Lie Markov models with purine/pyrimidine symmetry for dimension 5, 6, 8, 9, 10 and 12 (the Lie Markov models with dimension 1 to 4 are described within the text)

$2\text{id} \oplus \text{sgn} \oplus \xi$		
5.7a	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B_1^\xi, B_2^\xi)_{\mathbb{C}}$	$(1, 1, 0), (2, 0, 1)a, (4, \frac{1}{3}, \frac{2}{3})d$
5.7b	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B_3^\xi, B_4^\xi)_{\mathbb{C}}$	$(1, 1, 0), (2, 0, 1)a, (4, \frac{1}{3}, \frac{2}{3})e$
5.7c	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(1, 1, 0), (2, 0, 1)a, (4, \frac{1}{3}, \frac{2}{3})f$
$2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2$		
5.6a	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2})_{\mathbb{C}}$	$(2, 0, 1)a, (2, 0, 1)b, (2, 1, 0)$
$2\text{id} \oplus \zeta_2 \oplus \xi$		
5.6b	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_2}, B_1^\xi, B_2^\xi)_{\mathbb{C}}$	$(1, 0, 1), (1, 1, 0), (4, \frac{1}{3}, \frac{2}{3})a$
5.16	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_2}, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(1, 0, 1), (1, 1, 0), (2, \frac{1}{3}, \frac{2}{3}), (4, \frac{1}{7}, \frac{6}{7}),$ $(4, \frac{1}{3}, \frac{2}{3})f, (4, \frac{3}{5}, \frac{2}{5})$
5.11a	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_1^\xi, B_2^\xi)_{\mathbb{C}}$	$(1, 0, 1), (2, 1, 0), (4, \frac{1}{3}, \frac{2}{3})d, (4, \frac{1}{5}, \frac{4}{5})a$
5.11b	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_3^\xi, B_4^\xi)_{\mathbb{C}}$	$(1, 0, 1), (2, 1, 0), (4, \frac{1}{3}, \frac{2}{3})e, (4, \frac{1}{5}, \frac{4}{5})b$
5.11c	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(1, 0, 1), (2, 1, 0), (4, \frac{1}{3}, \frac{2}{3})f, (4, \frac{1}{5}, \frac{4}{5})c$
$2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus 2\zeta_2$		
6.6	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2}, B_2^{\zeta_2})_{\mathbb{C}}$	$(2, 1, 0), (4, 0, 1)e$
$2\text{id} \oplus \text{sgn} \oplus \zeta_2 \oplus \xi$		
6.7a	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_2}, B_1^\xi, B_2^\xi)_{\mathbb{C}}$	$(1, 1, 0), (2, 0, 1)a, (4, \frac{1}{3}, \frac{2}{3})a$
6.17a	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_2}, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(1, 1, 0), (2, 0, 1)a, (2, \frac{1}{3}, \frac{2}{3}), (4, \frac{1}{7}, \frac{6}{7}),$ $(4, \frac{1}{3}, \frac{2}{3})f, (4, \frac{3}{5}, \frac{2}{5})$
$2\text{id} \oplus \zeta_1 \oplus \zeta_2 \oplus \xi$		
6.7b	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_1}, B^{\zeta_2}, B_1^\xi, B_2^\xi)_{\mathbb{C}}$	$(1, 0, 1), (2, 0, 1)b, (4, \frac{1}{3}, \frac{2}{3})a$
6.17b	$(B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_1}, B^{\zeta_2}, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(1, 1, 0), (2, 0, 1)b, (2, \frac{1}{3}, \frac{2}{3}),$ $(4, \frac{1}{7}, \frac{6}{7}), (4, \frac{1}{3}, \frac{2}{3})f, (4, \frac{3}{5}, \frac{2}{5})$
$2\text{id} \oplus 2\zeta_2 \oplus \xi$		
6.8a	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_1^\xi, B_2^\xi)_{\mathbb{C}}$	$(2, 0, 1)c, (2, 1, 0), (4, \frac{1}{3}, \frac{2}{3})a$
6.8b	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(2, 0, 1)c, (2, 1, 0), (4, \frac{1}{3}, \frac{2}{3})b$
$2\text{id} \oplus 2\zeta_2 \oplus 2\xi$		
8.8	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_1^\xi, B_2^\xi, B_3^\xi, B_4^\xi)_{\mathbb{C}}$	$(4, 0, 1)c, (4, 1, 0)a$
8.16	$(B_1^{\text{id}}, B_2^{\text{id}}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_1^\xi, B_2^\xi, B_5^\xi, B_6^\xi)_{\mathbb{C}}$	$(2, 0, 1)c, (2, 1, 0), (4, 0, 1)a,$ $(4, \frac{1}{3}, \frac{2}{3})a, (4, \frac{1}{3}, \frac{2}{3})b$

Actually, this map induces a bijection between the Lie algebra $\mathcal{L}_{5.6b} = (B_1^{\text{id}}, B_2^{\text{id}}, B^{\zeta_2})_{\mathbb{C}}, B_1^\xi, B_2^\xi$ and the set of (not necessarily stochastic) rate matrices of HKY model. The inverse is given by

$$Q = \sum_{i \neq j} q_{ij} L_{ij} \mapsto \sum_{i \neq j} e^{-q_{ij}} L_{ij}.$$

However, these two models have different essential properties. For instance, while Model 5.6b is given by the linear variety $\mathcal{L}_{5.6b}$, it can be seen that the set of rate

Table 6 continued

$2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus 2\zeta_2 \oplus \xi$		
8.10a	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_1^\xi, B_2^\xi \rangle_{\mathbb{C}}$	$(2, 1, 0), (4, 0, 1)e, (4, \frac{1}{3}, \frac{2}{3})a$
8.10b	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_5^\xi, B_6^\xi \rangle_{\mathbb{C}}$	$(2, 1, 0), (4, 0, 1)e, (4, \frac{1}{3}, \frac{2}{3})b$
$2\text{id} \oplus \text{sgn} \oplus \zeta_2 \oplus 2\xi$		
8.17	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_2}, B_1^\xi \rangle_{\mathbb{C}}, B_2^\xi, B_3^\xi, B_4^\xi$	$(1, 1, 0), (4, 0, 1)d, (4, \frac{1}{3}, \frac{2}{3})a,$ $(4, \frac{1}{2}, \frac{1}{2})a, (4, \frac{3}{5}, \frac{2}{5})$
8.18	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_2}, B_1^\xi, B_2^\xi, B_3^\xi, B_4^\xi \rangle_{\mathbb{C}}$	$(2, 0, 1)a, (4, 0, 1)b,$ $(4, \frac{1}{3}, \frac{2}{3})a, (4, \frac{1}{3}, \frac{2}{3})b,$ $(4, 1, 0)b$
$2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$		
9.20a	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2}, B_1^\xi, B_2^\xi, B_3^\xi, B_4^\xi \rangle_{\mathbb{C}}$	$(4, 1, 0)a, (2, 0, 1)a, (2, 0, 1)b,$ $(4, 0, 1)b, (8, 0, 1)b$
9.20b	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_2^{\zeta_2}, B_3^\xi, B_4^\xi, B_5^\xi, B_6^\xi \rangle_{\mathbb{C}}$	$(2, 0, 1)b, (2, 1, 0), (4, 0, 1)f,$ $(4, \frac{1}{2}, \frac{1}{2})b, (8, \frac{1}{3}, \frac{2}{3})b$
$2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus 2\zeta_2 \oplus 2\xi$		
10.12	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_1^\xi, B_2^\xi, B_3^\xi, B_4^\xi \rangle_{\mathbb{C}}$	$(4, 0, 1)c, (4, 0, 1)e, (4, 1, 0)a$
10.34	$\langle B_1^{\text{id}}, B_2^{\text{id}}, B^{\text{sgn}}, B^{\zeta_1}, B_1^{\zeta_2}, B_2^{\zeta_2}, B_1^\xi, B_2^\xi, B_5^\xi, B_6^\xi \rangle_{\mathbb{C}}$	$(2, 1, 0), (4, 0, 1)d, (4, 0, 1)e,$ $(4, \frac{1}{3}, \frac{2}{3})a, (4, \frac{1}{3}, \frac{2}{3})b,$ $(8, \frac{1}{3}, \frac{2}{3})a, (8, 1, 1)$
$2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus 2\zeta_2 \oplus 3\xi$		
12.12	\mathcal{L}_{GM}	$(4, 1, 0)a, (8, 0, 1)a$

The first column gives the name of the model, while the second and the third column provide a basis of the corresponding Lie subalgebra and the ray-orbits of the corresponding stochastic cone, respectively (see Table 8)

matrices of HKY model describes a variety that is not linear and contains singular points. The deep connection between Lie Markov models and submodels of the general time reversible model appears as a beautiful line of research that will deserve some attention from the authors in the future. \diamond

Remark 12 As already noted in Remark 4, a number of models in the above list have more symmetries than those requested by the group \mathcal{G} , and they already appeared as Lie Markov models with \mathfrak{S}_4 symmetry (see Sumner et al. 2012a). For those models, the decomposition into irreducible representations of \mathcal{G} can be obtained from the decomposition into irreducible representations of \mathfrak{S}_4 by applying the branching rule of Table 3 (cf. Sumner et al. 2012a, Table 2). Since there are no subgroups between \mathcal{G} and \mathfrak{S}_4 , we can conclude that the rest of models listed here do not have further symmetries. \diamond

Remark 13 A vector subspace \mathcal{L} in \mathcal{L}_{GM} is a *matrix algebra* if its multiplicatively closed, that is, if the product XY lies in \mathcal{L} for any couple $X, Y \in \mathcal{L}$. Of course, this condition is stronger than that of a Lie algebra. The reader may wonder which of the models in the above list are actually algebras. The authors were surprised to find that the only Lie algebras which are not algebras correspond to the Lie Markov models that appear in families depending on some parameters a, b in the sense of Remark 9, that is, the Lie Markov models corresponding to the following decompositions:

id (for dimension 1);
 $\text{id} \oplus \text{sgn}$, $\text{id} \oplus \zeta_1$ and $\text{id} \oplus \zeta_2$ (for dimension 3);
 $\text{id} \oplus \zeta_2 \oplus \xi$, $\text{id} \oplus \text{sgn} \oplus \xi$, $\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2$ (for dimension 4); and
 $\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2 \oplus \xi$ (for dimension 8).

Notice that these decompositions correspond exactly to the decompositions of Table 4, that is, the irreducible permutation representations $(\mathcal{G}/H)_{\mathbb{C}}$ for a subgroup H of \mathcal{G} . \diamond

Remark 14 The reader may notice that not all decompositions listed in Table 5 give rise to Lie Markov models. Although, there exists Lie subalgebras of the general Markov model that decompose according to $2\text{id} \oplus \text{sgn} \oplus \zeta_1$ and $2\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \xi$, their relative position to the positive orthant causes a drop in the dimension of the convex polyhedral cone obtained when imposing the stochastic restrictions (namely, non-diagonal rates have to be non-negative). Such situations are not desired as explained in Remark 2 and do not correspond to our definition of Lie Markov model.

6 Discussion

Following the ideas of our previous work (Sumner et al. 2012a), in this paper we have discussed Lie Markov models with purine/pyrimidine symmetry. This symmetry was mathematically expressed by taking the group of nucleotide permutations

$$\mathcal{G} = \{e, (AG), (CT), (AG)(CT), (AC)(GT), (AT)(GC), (ACGT), (ATGC)\}.$$

Our main motivation is that this symmetry may be of special interest to the biologist who wishes to deal with evolutionary models preserving the specific grouping of nucleotides into purines and pyrimidines. In Sect. 2 we recalled some of the basic definitions on Lie Markov models and the required tools arising from representation theory of groups. We also show that any rate-matrix model being locally multiplicatively closed is necessarily a Lie Markov model. Also in this section, we introduce a new concept which is the *stochastic cone* of a Lie Markov model, being the set of stochastic rate matrices of the Lie Markov model. In Sect. 3 we explained how to derive Lie Markov models with prescribed symmetry and discussed the geometry of the corresponding cone of stochastic rate matrices. In Sect. 4 we took the permutation group \mathcal{G} and decomposed the space of all rate matrices into irreducible modules of \mathcal{G} and provided a basis consistent with this decomposition. In Sect. 5 we gave the full list of all Lie Markov models with \mathcal{G} symmetry, arranged by their dimension.

From an applied point of view, a natural question is which Lie Markov models are biologically interesting. This is a crucial point that will deserve special attention from the authors. Computer simulations to compare phylogenetic estimation using Lie Markov models with other evolutionary models are being designed and will appear in a future publication. At the same time, this leads to more theoretical questions as the connection between well-established evolutionary models and Lie Markov models, like the HKY model and the Lie Markov model 5.6b (see Remark 11).

We have considered models from a rate matrix perspective as some well-defined subset \mathcal{L} of the space \mathcal{L}_{GM} of all rate matrices. We could have adopted a more *algebraic*

Table 7 The Lie brackets of the basis $\{B_j^*\}$ of \mathcal{L}_{GM}

	$B_1^{i\text{d}}$	$B^{i\text{d}}$	B^{sgn}	B^{ζ_1}	$B_1^{\zeta_2}$	B^{ζ_2}	B_1^{ξ}	B_2^{ξ}	B_3^{ξ}	B_4^{ξ}	B_5^{ξ}	B_6^{ξ}
$B_1^{i\text{d}}$	0	0	0	0	0	0	$-2B_1^{\xi}$	$-2B_2^{\xi}$	$-2B_3^{\xi}$	$-2B_4^{\xi}$	$2B_5^{\xi}$	$2B_6^{\xi}$
$B^{i\text{d}}$		0	0	0	0	$-4B^{\zeta_2}$	$-4B_1^{\xi}$	$-4B_2^{\xi}$	0	0	0	0
B^{sgn}			0	$4B_1^{\zeta_2}$	$4B^{\zeta_1}$	0	$2B_2^{\xi}$	$2B_1^{\xi}$	$-2B_4^{\xi}$	$-2B_3^{\xi}$	$2B_6^{\xi}$	$2B_5^{\xi}$
B^{ζ_1}				0	$4B^{\text{sgn}}$	0	$-2B_2^{\xi}$	$2B_1^{\xi}$	$2B_4^{\xi}$	$-2B_3^{\xi}$	$2B_6^{\xi}$	$-2B_5^{\xi}$
$B_1^{\zeta_2}$					0	0	$-2B_1^{\xi}$	$2B_2^{\xi}$	$-2B_3^{\xi}$	$2B_4^{\xi}$	$2B_5^{\xi}$	$-2B_6^{\xi}$
B^{ζ_2}						0	0	0	$4B_1^{\xi}$	$-4B_2^{\xi}$	0	0
B_1^{ξ}							0	0	0	0	$2B^{\zeta_2}$	0
B_2^{ξ}								0	0	0	0	$-2B_2^{\zeta_2}$
B_3^{ξ}									0	0	$E_{3,5}$	$E_{3,6}$
B_4^{ξ}										0	$E_{4,5}$	$E_{4,6}$
B_5^{ξ}											0	0
B_6^{ξ}												0

The entries not included are easily determined by applying the rule $[X, Y] = -[Y, X]$. Here we use the notation $E_{3,5} := -6B_1^{i\text{d}} + 2B_2^{i\text{d}} - 2B_1^{\zeta_2}$, $E_{3,6} := -2B^{\text{sgn}} - 2B^{\zeta_1}$, $E_{4,5} := -2B^{\text{sgn}} + 2B^{\zeta_1}$, $E_{4,6} := -6B_1^{i\text{d}} + 2B_2^{i\text{d}} + 2B_1^{\zeta_2}$

perspective and deal with the substitution matrices instead, and keeping in mind the importance of substitution matrices being multiplicatively closed (see Sumner et al. 2012a), define “evolutionary model” as some well-defined groups \mathfrak{M} of matrices in $M_n(\mathbb{R})$. Then, when we restrict to the stochastic setting, we would be led to consider the intersection of \mathfrak{M} with the *stochastic* polytope:

$$\mathbb{P}_{\text{sto}} := \left\{ M = (m_{ij}) \in \mathfrak{M} : m_{ij} \geq 0, \sum_i m_{ij} = 1 \right\}.$$

This is a compact polytope with the identity matrix in one of the vertices. This polytope is cut into several connected components by the algebraic hypersurface of equation $\det(M) = 0$. We are mainly interested in the connected component that contains the identity matrix. This is because, by continuity arguments, this connected component contains the exponential of the stochastic rate matrices of the model. In this paper, we have preferred to introduce evolutionary models from the point of view of rate matrices because both the definition of Lie Markov models and the procedure to construct them appear in a natural way in this setting. However, the connection between rate matrices and substitution models is not completely clear, and it deserves further attention. For example, it is known that the image of the exponential map restricted to the stochastic cone does not cover in general the whole connected component of the identity. These issues are related to the convergence of the Baker–Campbell–Hausdorff formula (Blanes and Casas 2004) and the Elfving’s “embedding problem” (Davies 2010): given a Markov matrix M , there exists a matrix Q such that $M = e^Q$

Table 8 Ray-orbits of the Lie Markov models with \mathcal{G} symmetry with the corresponding generators

Orbit	Matrices	Dec. as an abstract set / dec. in \mathfrak{L}_{GM}
(1, 1, 0)	$L_{(12)(34)}$	id
(1, 0, 1)	$L_{(13)(24)} + L_{(14)(23)}$	id
$(1, \frac{1}{3}, \frac{2}{3})$	$L_{(12)(34)} + L_{(13)(24)} + L_{(14)(23)}$	id
(2, 0, 1) <i>a</i>	$L_{(13)(24)}, L_{(14)(23)}$	id \oplus sgn
(2, 0, 1) <i>b</i>	$L_{(1324)}, L_{(1423)}$	id \oplus ζ_1
(2, 0, 1) <i>c</i>	Ch ₁₂ , Ch ₃₄	id \oplus ζ_2
(2, 1, 0)	$L_{(12)}, L_{(34)}$	id \oplus ζ_2
$(2, \frac{1}{3}, \frac{2}{3})$	R_{12}^+, R_{34}^+	id \oplus ζ_2
(4, 0, 1) <i>a</i>	$V_1 + H_2 - 2L_{12}, V_2 + H_1 - 2L_{12}, V_3 + H_4 - 2L_{34}, V_4 + H_3 - 2L_{34}$	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
(4, 0, 1) <i>b</i>	$H_{13}^+ - (L_{21} + L_{43}), H_{14}^+ - (L_{21} + L_{34}), H_{23}^+ - (L_{12} + L_{43}), H_{24}^+ - (L_{12} + L_{34})$	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
(4, 0, 1) <i>c</i>	$L_{13} + L_{14}, L_{23} + L_{24}, L_{31} + L_{32}, L_{41} + L_{42}$	id \oplus $\zeta_2 \oplus \xi$
(4, 0, 1) <i>d</i>	$L_{13} + L_{42}, L_{23} + L_{41}, L_{31} + L_{24}, L_{32} + L_{14}$	id \oplus sgn $\oplus \xi$
(4, 0, 1) <i>e</i>	$L_{13} + L_{24}, L_{23} + L_{14}, L_{31} + L_{42}, L_{32} + L_{41}$	id \oplus sgn $\oplus \zeta_1 \oplus \zeta_2$
(4, 0, 1) <i>f</i>	$L_{(13)}, L_{(14)}, L_{(23)}, L_{(24)}$	id \oplus $\zeta_2 \oplus \xi$
(4, 1, 0) <i>a</i>	$L_{12}, L_{21}, L_{34}, L_{43}$	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
(4, 1, 0) <i>b</i>	$L_{12} + L_{34}, L_{12} + L_{43}, L_{21} + L_{34}, L_{21} + L_{43}$	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{3}, \frac{2}{3})$ <i>a</i>	R_1, R_2, R_3, R_4	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{3}, \frac{2}{3})$ <i>b</i>	V_1, V_2, V_3, V_4	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{3}, \frac{2}{3})$ <i>c</i>	H_1, H_2, H_3, H_4	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{3}, \frac{2}{3})$ <i>d</i>	$R_{13}^+, R_{14}^+, R_{23}^+, R_{24}^+$	id \oplus sgn $\oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{3}, \frac{2}{3})$ <i>e</i>	$H_{13}^+, H_{14}^+, H_{23}^+, H_{24}^+$	id \oplus sgn $\oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{3}, \frac{2}{3})$ <i>f</i>	$V_{13}^+, V_{14}^+, V_{23}^+, V_{24}^+$	id \oplus sgn $\oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{7}, \frac{6}{7})$	$V_1 + \text{Ch}_{12}, V_2 + \text{Ch}_{12}, V_3 + \text{Ch}_{34}, V_4 + \text{Ch}_{34}$	id \oplus $\zeta_2 \oplus \xi$ / id $\oplus \xi$
$(4, \frac{1}{2}, \frac{1}{2})$ <i>a</i>	$R_{13}^+ - (L_{14} + L_{32}), R_{14}^+ - (L_{13} + L_{42}), R_{23}^+ - (L_{24} + L_{31}), R_{24}^+ - (L_{23} + L_{41})$	id \oplus $\zeta_2 \oplus \xi$
$(4, \frac{1}{2}, \frac{1}{2})$ <i>b</i>	$L_{(1234)}, L_{(1243)}, L_{(1324)}, L_{(1342)}$	id \oplus sgn $\oplus \xi$
$(4, \frac{1}{5}, \frac{4}{5})$ <i>a</i>	$2R_1 + \text{Ch}_{34}, 2R_2 + \text{Ch}_{34}, 2R_3 + \text{Ch}_{12}, 2R_4 + \text{Ch}_{12}$	id \oplus $\zeta_2 \oplus \xi$
$(4, \frac{1}{5}, \frac{4}{5})$ <i>b</i>	$2H_1 + \text{Ch}_{34}, 2H_2 + \text{Ch}_{34}, 2H_3 + \text{Ch}_{12}, 2H_4 + \text{Ch}_{12}$	id \oplus $\zeta_2 \oplus \xi$
$(4, \frac{1}{5}, \frac{4}{5})$ <i>c</i>	$2V_1 + \text{Ch}_{12}, 2V_2 + \text{Ch}_{12}, 2V_3 + \text{Ch}_{34}, 2V_4 + \text{Ch}_{34}$	id \oplus $\zeta_2 \oplus \xi$
$(4, \frac{3}{5}, \frac{2}{5})$	$V_1 + L_{(34)}, V_2 + L_{(34)}, V_3 + L_{(12)}, V_4 + L_{(12)}$	id \oplus $\zeta_2 \oplus \xi$
(8, 0, 1) <i>a</i>	$L_{13}, L_{14}, L_{23}, L_{24}, L_{31}, L_{32}, L_{41}, L_{42}$	id \oplus sgn $\oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$
(8, 0, 1) <i>b</i>	$L_{(13)(24)} - L_{13} + L_{23}, L_{(13)(24)} - L_{24} + L_{14}, L_{(13)(24)} - L_{31} + L_{41}, L_{(13)(24)} - L_{42} + L_{32}, L_{(14)(23)} - L_{14} + L_{24}, L_{(14)(23)} - L_{23} + L_{13}, L_{(14)(23)} - L_{32} + L_{42}, L_{(14)(23)} - L_{41} + L_{31}$	id \oplus sgn $\oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$ / id \oplus sgn $\oplus \zeta_1 \oplus \xi$
$(8, \frac{1}{3}, \frac{2}{3})$ <i>a</i>	$R_1 - L_{14} + L_{41}, R_1 - L_{13} + L_{31}, R_2 - L_{24} + L_{42}, R_2 - L_{23} + L_{32}, R_3 - L_{31} + L_{13}, R_2 - L_{32} + L_{23}, R_4 - L_{41} + L_{14}, R_4 - L_{42} + L_{24}$	id \oplus sgn $\oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$ / id $\oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$

Table 8 continued

Orbit	Matrices	Dec. as an abstract set / dec. in \mathfrak{L}_{GM}
$(8, \frac{1}{3}, \frac{2}{3})b$	$L_{(123)}, L_{(124)}, L_{(132)}, L_{(134)}, L_{(142)}, L_{(143)},$ $L_{(234)}, L_{(243)}$	$\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2$ $\oplus 2\xi /$ $\text{id} \oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$
$(8, \frac{1}{2}, \frac{1}{2})$	$L_{12} + L_{34} + 2L_{13}, L_{12} + L_{34} + 2L_{31}, L_{12}$ $+ L_{43} + 2L_{14}, L_{12} + L_{43} + 2L_{41}, L_{21}$ $+ L_{34} + 2L_{23}, L_{21} + L_{34} + 2L_{32}, L_{21}$ $+ L_{43} + 2L_{24}, L_{21} + L_{43} + 2L_{42}$	$\text{id} \oplus \text{sgn} \oplus \zeta_1 \oplus \zeta_2$ $\oplus 2\xi /$ $\text{id} \oplus \zeta_1 \oplus \zeta_2 \oplus 2\xi$

The 3rd column describes the decomposition into irreducible representations of both the abstract set generated by these orbits and the subspace of \mathfrak{L}_{GM} spanned by them (see Remark 8). When both decompositions are equal, we write it down only once

and e^{tQ} is a Markov matrix for all $t \geq 0$. We want to explore this question in the future to clarify the connection between substitution and rate matrices of evolutionary models.

Although we have kept the original definition of symmetry for a Lie Markov model from Sumner et al. (2012a), an interesting question arises if one tries to expand this definition. Namely, we could investigate evolutionary models which are invariant under the action of some permutation subgroup G of \mathfrak{S}_4 without the additional request that they have a permutation basis. From an applied point of view, we do not find any particular reason not to consider this expanded definition, which would lead to a huge number of possible models. For example, under this expanded definition, we would admit the complex span of the ray-orbits $(4, \frac{1}{3}, \frac{2}{3})d : \mathfrak{L} = \langle R_{13}^+, R_{14}^+, R_{23}^+, R_{24}^+ \rangle_{\mathbb{C}}$, $(4, \frac{1}{3}, \frac{2}{3})e : \mathfrak{L} = \langle H_{13}^+, H_{14}^+, H_{23}^+, H_{24}^+ \rangle_{\mathbb{C}}$ and $(4, \frac{1}{3}, \frac{2}{3})f : \mathfrak{L} = \langle V_{13}^+, V_{14}^+, V_{23}^+, V_{24}^+ \rangle_{\mathbb{C}}$ (see Table 8) as models with symmetry \mathcal{G} and decomposition $\text{id} \oplus \xi$ (note that this decomposition does not appear in the list of Table 5). More interestingly, as noted in Remark 10, the set of doubly stochastic rate matrices has \mathfrak{S}_4 symmetry under this expanded definition, and moreover they form a Lie algebra. The authors keep back this line of research for future publication.

Acknowledgments JFS was partially supported by Ministerio de Educación y Ciencia MTM2009-14163-C02-02, MTM2012-38122-C03-01 and Generalitat de Catalunya, 2009 SGR 1284. JGS and PDJ were partially supported by Australian Research Council grant DP0877447. MDW was partially supported by Australian Research Council Grant FT100100031.

References

Alexandrov AD (2005) Convex polyhedra. Springer Monographs in Mathematics. Springer, Berlin. ISBN 3-540-23158-7 (translated from the 1950 Russian edition by N. S. Dairbekov, S. S. Kutateladze and A. B. Sossinsky, with comments and bibliography by V. A. Zalgaller and appendices by L. A. Shor and Yu. A. Volkov)

Birkhoff G (1938) Analytical groups. Trans Am Math Soc 43(1):61–101. ISSN 0002–9947. doi:10.2307/1989902

Blanes S, Casas F (2004) On the convergence and optimization of the Baker–Campbell–Hausdorff formula. Linear Algebra Appl 378:135–158. ISSN 0024–3795. doi:10.1016/j.laa.2003.09.010

- Bogopolski O (2008) Introduction to group theory. EMS Textbooks in Mathematics, European Mathematical Society (EMS), Zürich. ISBN 978-3-03719-041-8. doi:10.4171/041 (translated, revised and expanded from the Russian original)
- Campbell JE (1897) On a law of combination of operators (second paper). Proc Lond Math Soc 28:381–390
- Casanellas M, Fernández-Sánchez J (2010) Relevant phylogenetic invariants of evolutionary models. J Math Pure Appl 96:207–229
- Casanellas M, Sullivant S (2005) The strand symmetric model. In: Algebraic statistics for computational biology. Cambridge University Press, New York, pp 305–321. doi:10.1017/CBO9780511610684.020
- Casanellas M, Fernández-Sánchez J, Kedzińska A (2012) The space of phylogenetic mixtures for equivariant models. Algorithms Mol Biol 7:33
- Davies EB (2010) Embeddable Markov matrices. Electron J Probab 15(47):1474–1486. ISSN 1083–6489. doi:10.1214/EJP.v15-733
- Donten-Bury M, Michałek M (2012) Phylogenetic invariants for group-based models. J Algebr Stat 3(1):44–63. ISSN 1309–3452
- Drainsma J, Kuttler J (2008) On the ideals of equivariant tree models. Math Ann 344:619–644
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376
- Fernández-Sánchez J (2013) Code for lie markov models with purine/pyrimidine symmetry. http://www.pagines.ma1.upc.edu/jfernandez/purine_pyrimidine.html
- Hasegawa M, Kishino H, Yano T (1988) Phylogenetic inference from DNA sequence data. Statistical theory and data analysis, II (Tokyo, 1986). North-Holland, Amsterdam
- James G, Liebeck M (2001) Representations and characters of groups, 2nd edn. Cambridge University Press, New York
- Johnson JE (1985) Markov-type Lie groups in $GL(n, R)$. J Math Phys 26:252–257
- Jukes T, Cantor C (1969) Evolution of protein molecules. In: Mammalian protein, metabolism, pp 21–132
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16:111–120
- Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci 78:1454–1458
- Michałek M (2011) Geometry of phylogenetic group-based models. J Algebra 339:339–356. ISSN 0021-8693. doi:10.1016/j.jalgebra.2011.05.016
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818
- Rotman J (1995) An introduction to the theory of groups, 4th edn, volume 148 of Graduate Texts in Mathematics. Springer, New York. ISBN 0-387-94285-8
- Sagan BE (2001) The symmetric group: representations, combinatorial algorithms, and symmetric functions, 2nd edn., Graduate Texts in Mathematics. Springer, Berlin
- Semple C, Steel M (2003) Phylogenetics. Oxford Press, Oxford
- Stein W et al (2012) Sage Mathematics Software (Version 4.8). The Sage Development Team. <http://www.sagemath.org>
- Sumner JG, Fernández-Sánchez J, Jarvis PD (2012a) Lie Markov models. J Theor Biol 298:16–31. ISSN 0022-5193. doi:10.1016/j.jtbi.2011.12.017
- Sumner JG, Jarvis PD, Fernández-Sánchez J, Kaine BT, Woodhams MD, Holland BR (2012b) Is the general time-reversible model bad for molecular phylogenetics? Syst Biol 61:1069–1074
- Tavaré S (1986) Some probabilistic and statistical problems in the analysis of dna sequences. Lect Math Life Sci (American Mathematical Society) 17:57–86
- Yap V, Pachter L (2004) Identification of evolutionary hotspots in the rodent genomes. Genome Res 14(4):574–579