J. Math. Biol.
DOI 10.1007/s00285-013-0671-9

Mathematical Biology

Efficient computation in the IM model

Lars Nørvang Andersen · Thomas Mailund · Asger Hobolth

Received: 27 September 2012 / Revised: 1 March 2013

Abstract In this paper we analyze the isolation-with-migration model in a continuous-time Markov chain framework, and derive analytical expressions for the probability densities of gene tree topologies with an arbitrary number of lineages. We combine these densities with both nucleotide-substitution and infinite sites mutation models and derive probabilities for use in maximum likelihood estimation. We demonstrate how to apply lumpability of continuous-time Markov chains to achieve a significant reduction in the size of the state-space under consideration. We use matrix exponentiation and spectral decomposition to derive explicit expressions for the case of two diploid individuals in two populations, when the data is given as alignment columns. We implement these expressions in order to carry out a maximum likelihood analysis and provide a simulation study to examine the performance of our method in terms of our ability to recover true parameters. Finally, we show how the performance depends on the parameters in the model.

Keywords Coalescence theory · Gene flow · Isolation-with-migration · Maximum likelihood · Speciation

Mathematics Subject Classification (2000) 92B05 · 92D99

L. N. Andersen (✉) · T. Mailund · A. Hobolth
Bioinformatics Research Centre, Aarhus University, C.F. Møllers Allé 8, Aarhus C, 8000 Aarhus, Denmark
e-mail: larsa@birc.au.dk

T. Mailund
e-mail: mailund@birc.au.dk

A. Hobolth
e-mail: asger@birc.au.dk

Published online: 16 April 2013

1 Introduction

The study of how speciation occurs is a central theme in population genetics. One aspect of the speciation process, which has been particularly scrutinized is the presence of gene flow. Recent advances in DNA sequencing technology have made available a large number of new alignments of whole-genome sequence data, which may potentially aid in this study, provided that we are able to describe models of such data and implement associated estimation methods. Traditional methods based on Markov Chain Monte Carlo (MCMC) algorithms (e.g. Hey and Nielsen 2004) are often very computationally demanding. An alternative to MCMC methods is to use maximum likelihood (ML) estimation, and this approach has been considered in recent papers (Wang and Hey 2010; Lohse et al. 2011; Zhu and Yang 2012). However, a drawback of such methods is that they are limited to few (two or three) lineages. In this paper we derive analytical expressions in a model with an arbitrary number of lineages in an arbitrary number of populations. This extension causes an explosion in the size of the state-space under consideration. We quantify this state-space explosion, and show how it can be reduced by using lumpability.

Lumpability of Markov chains was introduced in Kemeny and Snell (1960) and has found applications within queueing networks, stochastic Petri nets and stochastic process algebras but appears to be largely unnoticed within coalescent theory (but see Tian and Lin 2009). A further drawback of typical ML methods is that by way of considering loci of substantial length (typically $\sim$ 100 bp) these methods are potentially vulnerable to bias due to intra-locus recombination. In this paper we consider two possible mutation models, a time-reversible model and an infinite sites model. In the time-reversible case, our data is alignment columns and we are therefore able to avoid intra-locus recombination. In both cases we assume free recombination between loci. There is a potential loss of power when considering only alignment columns, and we provide a simulation study which shows how well we are able to recover parameters in the case of two diploid individuals in two populations. The true parameters of our simulation study are taken from Scally et al. (2012) where our method was used to provide a likelihood surface for migration and split time parameters between the Eastern and the Western Gorilla, and the simulation study also serves the purpose of examining the accuracy and sensitivity of the method in the part of the parameter-space, which is relevant when considering this data. The isolation-with-migration (IM) model we consider describes a single ancestral panmictic population with effective population size N_A which splits into P populations/demes at time T_A in the past. Migration is possible between populations, and in each generation we expect a proportion of $M_{i\to j}$ of the lineages in population i to migrate to population j (looking back in time, since the present is time $t = 0$). This model has been considered in Nielsen and Wakeley (2001) and Wang and Hey (2010) and is based on the classical theory of Kingman (1982).

2 The Wang-Hey model

Before introducing the general model, we consider the description of the IM model in Wang and Hey (2010) and Hobolth et al. (2011). The purpose of this is to introduce

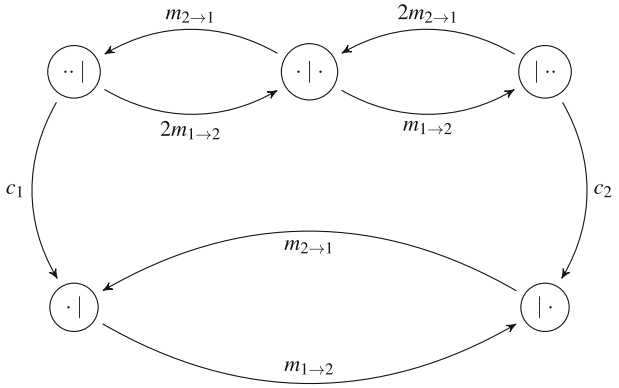


Fig. 1 The five states of the IM model with two lineages and two populations, prior to T_A. Each lineage is represented by a *dot* and a barrier separates the two populations. The states are $S_{11},\ S_{12},\ S_{22}$ (*upper row*) and $S_1,\ S_2$ (*lower row*)

our framework by means of a concrete example, and demonstrate how this can be used to give explicit expressions for the densities and probabilities of interest. We use the matrix notation presented in Sect. 8.1 in the Appendix. The given expressions are simplified versions of those derived for the general model in later sections.

In Wang and Hey (2010) the IM-model is given as a continuous-time Markov chain (CTMC) where, prior to the split time T_A, five states $S_{11}, S_{12}, S_{22}, S_1, S_2$ describe how many lineages is in each population, with S_{11} and S_{22} being two lineages in population 1 and 2 respectively, S_{12} being one lineage in each population and S_1 and S_2 being states, where the lineages have coalesced and are in population 1 and 2 respectively (see Fig. 1). Migration rates are denoted $m_{1\to2}$ for migration from population 1 to population 2 and $m_{2\to1}$ in the reverse direction, and coalescence rates are denoted $c_i,\ i = 1, 2$. The rate matrix is then given by

$$Q = \begin{array}{c} \\ S_{11} \\ S_{12} \\ S_{22} \\ S_1 \\ S_2 \end{array} \begin{array}{c} \begin{array}{ccccc} S_{11} & S_{12} & S_{22} & S_1 & S_2 \end{array} \\ \begin{pmatrix} -c_1 - 2m_{1\to2} & 2m_{1\to2} & 0 & c_1 & 0 \\ m_{2\to1} & -m_{1\to2} - m_{2\to1} & m_{1\to2} & 0 & 0 \\ 0 & 2m_{2\to1} & -c_2 - 2m_{2\to1} & 0 & c_2 \\ 0 & 0 & 0 & -m_{1\to2} & m_{1\to2} \\ 0 & 0 & 0 & m_{2\to1} & -m_{2\to1} \end{pmatrix} \end{array}.$$

After the split T_A, there are two states, S_{AA} corresponding to two ancestral lineages and S_A, which corresponds to a single ancestral lineage. The transition rate from S_{AA} to S_A is c_A, and 0 in the reverse direction. The coalescence time T is an absolutely continuous random variable, and in Hobolth et al. (2011) the formula for the coalescent density (the probability density of T), when the starting state is s, is given by

$$f(t) = \begin{cases} \left(\mathrm{e}^{Qt}\right)_{s,S_{11}} c_1 + \left(\mathrm{e}^{Qt}\right)_{s,S_{22}} c_2 & \text{for} \quad t < T_A \\ \left(\left(\mathrm{e}^{QT_A}\right)_{s,S_{11}} + \left(\mathrm{e}^{QT_A}\right)_{s,S_{12}} + \left(\mathrm{e}^{QT_A}\right)_{s,S_{22}}\right) c_A \mathrm{e}^{-c_A(t-T_A)} & \text{for} \quad t \geq T_A, \end{cases} \tag{1}$$

where $e^B = \sum_{n=0}^{\infty} B^n/n!$ is the matrix exponential of B. To provide a simplification of (1) we proceed by defining the set $\Upsilon = \{S_{11}, S_{12}, S_{22}\}$ of states as well as merging states S_1 and S_2 together in a dummy absorbing state A and considering the rate matrix

$$Q_{\Upsilon\cup A} := \begin{array}{c} \\ S_{11} \\ S_{12} \\ S_{22} \\ A \end{array} \begin{array}{c} \begin{array}{cccc} S_{11} & S_{12} & S_{22} & A \end{array} \\ \begin{pmatrix} -c_1 - 2m_{1\to 2} & 2m_{1\to 2} & 0 & c_1 \\ m_{2\to 1} & -m_{1\to 2} - m_{2\to 1} & m_{1\to 2} & 0 \\ 0 & 2m_{2\to 1} & -c_2 - 2m_{2\to 1} & c_2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

where we use the notation (31) in the Appendix.

The matrix $Q_{\Upsilon\cup A}$ may be written succinctly in terms of the submatrix Q_Υ and the vector $\mathbf{c} = (c_1, 0, c_2)$:

$$Q_{\Upsilon\cup A} := \begin{pmatrix} Q_\Upsilon & \mathbf{c}^* \\ \mathbf{0} & 0 \end{pmatrix}.$$

We observe that the coalescence time T for $t < T_A$ is simply the absorption time in A of the CTMC with rate matrix $Q_{\Upsilon\cup A}$ and hence has a Phase-type distribution with density

$$f(t) = \boldsymbol{\alpha} e^{Q_\Upsilon t}\mathbf{c}^*, \tag{2}$$

for initial vector $\boldsymbol{\alpha}$ (see Asmussen 2003 Prop. 4.1 Chap. III). For $t \geq T_A$ we have

$$f(t) = \mathbb{P}(T_A < T)c_A e^{-c_A(t-T_A)} = \left(\boldsymbol{\alpha} e^{Q_\Upsilon T_A}\mathbf{1}^*\right) c_A e^{-c_A(t-T_A)}, \tag{3}$$

where $\mathbf{1} = (1, 1, 1)$.

Note, that if we assume the starting state is s (i.e the sth entry of $\boldsymbol{\alpha}$ is 1), (2) may be written

$$f(t) = \sum_{\beta\in\Upsilon} \left(e^{Q_\Upsilon t}\right)_{s,\beta} (Q_{\Upsilon\cup A})_{\beta,A}, \tag{4}$$

that is, we sum over all possible states from which coalescence is possible and multiply with the appropriate rate.

We can use the formulas above to calculate the likelihood for a single locus given a mutation model, if they are combined with the fact that Q_Υ is diagonalizable (see Sect. 8.2 in the Appendix) so we may write $Q_\Upsilon = VDV^{-1}$ where V is a matrix of eigenvectors of Q_Υ, and $D = \text{diag}(\boldsymbol{\lambda})$ is a diagonal matrix with the associated eigenvalues $\boldsymbol{\lambda}$ of Q_Υ, and we have $e^{Q_\Upsilon t} = Ve^{Dt}V^{-1}$. If, for example, we assume a Jukes-Cantor model of substitution, the probability of observing the starting nucleotide after time t is $1/4 + (3/4)e^{-4t/3}$ and hence the probability of homozygosity is

$$
\begin{aligned}
&\int_0^\infty \left(\frac{1}{4}+\frac{3}{4}\mathrm{e}^{\frac{-4\cdot 2t}{3}}\right) f(t)\mathrm{d}t \\
&= \boldsymbol{\alpha} V \int_0^{T_A} \left(\frac{1}{4}+\frac{3}{4}\mathrm{e}^{\frac{-8t}{3}}\right) \mathrm{e}^{Dt}\mathrm{d}t\, V^{-1}\mathbf{c}^* + \left(\boldsymbol{\alpha} V \mathrm{e}^{T_A D} V^{-1}\mathbf{1}^*\right) \int_{T_A}^\infty \left(\frac{1}{4}+\frac{3}{4}\mathrm{e}^{\frac{-8t}{3}}\right) c_A \mathrm{e}^{-c_A(t-T_A)}\mathrm{d}t \\
&= \boldsymbol{\alpha} V \left(\frac{\mathrm{e}^{T_A D}-1}{4} D^{-1} + \frac{3}{4}\left(\mathrm{e}^{T_A(D-\mathrm{diag}(\frac{8}{3}\mathbf{1}))}-1\right)\left(D-\mathrm{diag}(\frac{8}{3}\mathbf{1})\right)^{-1}\right) V^{-1}\mathbf{c}^* \\
&\quad + \left(\boldsymbol{\alpha} V \mathrm{e}^{T_A D} V^{-1}\mathbf{1}^*\right)\left(\frac{1}{4}+\frac{3}{4}\frac{\mathrm{e}^{\frac{-8T_A}{3}} c_A}{\frac{8}{3}+c_A}\right).
\end{aligned}
$$

If we prefer, we can make the matrix-products above explicit by using the notation $V=(v)_{\alpha,\beta}$, $V^{-1}=(v^{-1})_{\alpha,\beta}$, for $\alpha,\beta\in\Upsilon$ to denote the entries of V and its inverse respectively, and letting $\boldsymbol{\lambda}=(\lambda_\gamma)_{\gamma\in\Upsilon}$ denote the eigenvalues of Q_Υ. Then, assuming the starting state is s, we have:

$$
\begin{aligned}
\int_0^\infty \left(\frac{1}{4}+\frac{3}{4}\mathrm{e}^{\frac{-4\cdot 2t}{3}}\right) f(t)\mathrm{d}t &= \sum_{\beta\in\Upsilon}\sum_{\gamma\in\Upsilon} v_{s,\gamma} v^{-1}_{\gamma,\beta} \left(Q_{\Upsilon\cup A}\right)_{\beta,A} \left(\frac{\mathrm{e}^{T_A\lambda_\gamma}-1}{4\lambda_\gamma}+\frac{3}{4}\frac{\mathrm{e}^{T_A(\lambda_\gamma-\frac{8}{3})}-1}{\lambda_\gamma-\frac{8}{3}}\right) \\
&\quad + \sum_{\beta\in\Upsilon}\sum_{\gamma\in\Upsilon} v_{s,\gamma} e^{-\lambda_\gamma T_A} v^{-1}_{\gamma,\beta}\left(\frac{1}{4}+\frac{3}{4}\frac{\mathrm{e}^{\frac{-8T_A}{3}} c_A}{\frac{8}{3}+c_A}\right).
\end{aligned}
$$

In Fig. 2 we show how the probability of homozygosity depends on the split time and migration rate in models where we assume symmetric migration rates and identical coalescence rates in the extant populations. The ancestral coalescence rate takes the values 0.75, 1 and 1.25.

Alternatively, we could use an infinite sites substitution model. In this case the number of mutations conditional on the coalescence time $T=t$ is $PO(2t,k)=(2t)^k\mathrm{e}^{-2t}/k!$, so that the probability of observing k mutations is

$$
\begin{aligned}
\int_0^\infty PO(2t,k) f(t)\mathrm{d}t &= \sum_{\beta\in\Upsilon}\sum_{\gamma\in\Upsilon} v_{s,\gamma} v^{-1}_{\gamma,\beta} \left(Q_{\Upsilon\cup A}\right)_{\beta,A} \int_0^{T_A} \frac{(2t)^k}{k!}\mathrm{e}^{-2t}\mathrm{e}^{-\lambda_\gamma t}\mathrm{d}t \\
&\quad + \sum_{\beta\in\Upsilon}\sum_{\gamma\in\Upsilon} v_{s,\gamma} e^{-\lambda_\gamma T_A} v^{-1}_{\gamma,\beta} \int_{T_A}^\infty \frac{(2t)^k}{k!}\mathrm{e}^{-2t} c_A \mathrm{e}^{-c_A(t-T_A)}\mathrm{d}t
\end{aligned}
$$

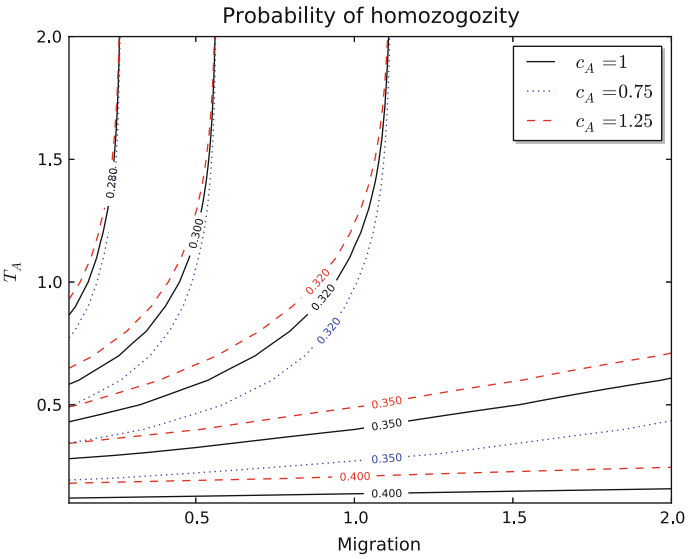


Fig. 2 Contour plots of the probability of homozygosity in models with identical coalescence rates (*black*, *solid*) and with ancestral coalescence rate 0.75 (*blue*, *dotted*) and 1.25 (*red*, *dashed*) times the coalescence rates of the extant populations, which are assumed identical and equal to 1

$$= \sum_{\beta\in\Upsilon}\sum_{\gamma\in\Upsilon} v_{s,\gamma} v^{-1}_{\gamma,\beta} \left(Q_{\Upsilon\cup A}\right)_{\beta,A} \frac{2^k}{(2+\lambda_\gamma)^{k+1}} \frac{1}{k!} \underline{\Gamma}(k+1, T_A(2+\lambda_\gamma))$$

$$\sum_{\beta\in\Upsilon}\sum_{\gamma\in\Upsilon} v_{s,\gamma} e^{-\lambda_\gamma T_A} v^{-1}_{\gamma,\beta} \frac{2^k}{(2+\lambda_\gamma)^{k+1}} \frac{1}{k!} \overline{\Gamma}\left(k+1, T_A\left(2+c_A\right)\right)$$

where we use the lower and upper incomplete gamma functions $\underline{\Gamma}(n,x) := \int_0^x t^{n-1}\mathrm{e}^{-t}\mathrm{d}t$ and $\overline{\Gamma}(n,x) := \int_x^\infty t^{n-1}\mathrm{e}^{-t}\mathrm{d}t$.

3 The general model

When defining the general model, we first need to define the state-space of the CTMC. We want the state-space to reflect which lineages are present, and to which population each lineage belongs. We do this by labeling each lineage with a subset of $[L] := \{1, 2, \ldots, L\}$, and combining the lineage with an integer indicating the population. The two form a tuple, and a state of the CTMC is a set of such tuples. The extant lineages are labeled with the singleton sets, and a coalescent event between two lineages is modeled by taking the union of the involved sets. An example of a state is the starting state in a situation where we consider two diploid individuals in two populations, that is with two lineages in each of the two populations—this state is $\{(1, \{1\}), (1, \{2\}), (2, \{3\}), (2, \{4\})\}$. From this state, both migration and coalescence is possible. For example, the lineage $\{2\}$ may migrate from population 1 to population 2, so that the state becomes $\{(1, \{1\}), (2, \{2\}), (2, \{3\}), (2, \{4\})\}$. Another possibility

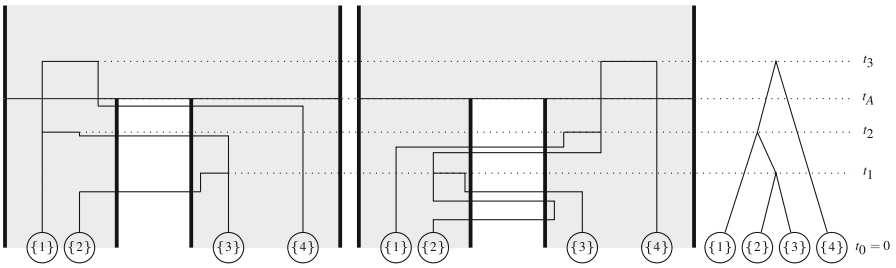


Fig. 3 Two different genealogies, which result in the same coalescent tree. Even though the two genealogies do not share migration events, and coalescence events in the extant populations take place in different populations, both genealogies belong to the coalescent tree to the right

is coalescence between the lineages {1} and {2}, which means that the state changes to $\{(1, \{1, 2\}), (2, \{3\}), (2, \{4\})\}$.

The example in Sect. 2 captures many of the elements of the general case: The density of the coalescence time is derived as an absorption time in a CTMC, and the likelihood for a single locus is in turn derived by combining the coalescence time with a mutation model and integrating explicitly using matrix exponentiation and diagonalization. One aspect of the general case is missing in the previous example, however, namely the role of the genealogy of the sample. We take "genealogy" to mean the whole history of the sample, including any migration events, and an important observation is that mutation probabilities are unaffected by migration. This motivates the mapping $F(\cdot)$ defined in (7) below, which simply removes any information about the populations from the states, so that if we continue our example with two diploid individuals in two populations we have:

$$F\big(\{(1, \{1\}), (1, \{2\}), (2, \{3\}), (2, \{4\})\}\big) = \{\{1\}, \{2\}, \{3\}, \{4\}\}$$
$$F\big(\{(1, \{1\}), (2, \{2\}), (2, \{3\}), (2, \{4\})\}\big) = \{\{1\}, \{2\}, \{3\}, \{4\}\}$$
$$F\big(\{(1, \{1, 2\}), (2, \{3\}), (2, \{4\})\}\big) = \{\{1, 2\}, \{3\}, \{4\}\}.$$

Note that F in this case takes its values in the partitions of $\{1, 2, 3, 4\}$. The mapping F allows us to define the so-called *coalescent trees*. Intuitively, the coalescent trees are unions of those genealogies that are identical when information about the population is removed. In Fig. 3 we have two different genealogies, which belong to the same coalescent tree.

We represent coalescent trees using a vector $\mathbf{t}$ of coalescent times and a vector $\mathbf{Y}$, which represents the lineages present at the time of the coalescent event. Figure 4 shows two such coalescent trees with different topologies. A main result of this paper is the derivation of the probability density for the coalescent trees in (8), which plays the same role in the general case, as the coalescent density did in the example in Sect. 2.

3.1 Formal definition of the IM-model

The IM-model is a CTMC $\{X(t)\}_{t \geq 0}$ pieced together from two time-homogeneous CTMCs, one for $0 \leq t < T_A$ and one for $T_A \leq t$. Each state is a set $\{(j_i, l_i) \mid i =$

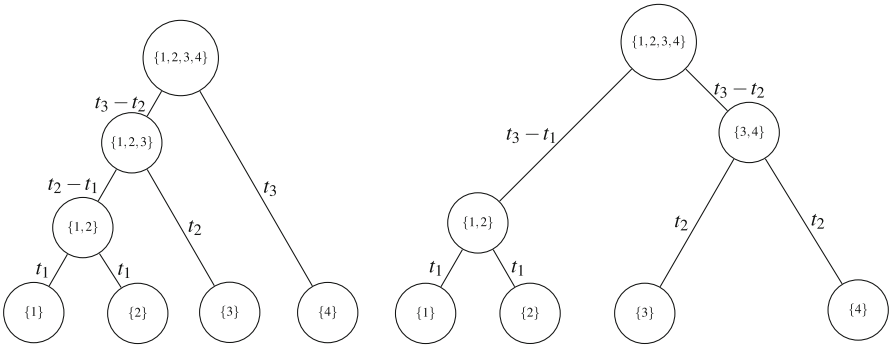


Fig. 4 Two topologically different coalescent trees $\mathscr{C}(\mathbf{t}, \mathbf{Y}_1)$ and $\mathscr{C}(\mathbf{t}, \mathbf{Y}_2)$ where $\mathbf{Y}_1 = (\{\{1\}, \{2\}, \{3\}, \{4\}\}, \{\{1, 2\}, \{3\}, \{4\}\}, \{\{1, 2, 3\}, \{4\}\})$ $(left)$ and $\mathbf{Y}_2 = (\{1\}, \{2\}, \{3\}, \{4\}\}, \{\{1, 2\}, \{3\}, \{4\}\}, \{\{1, 2\}, \{3, 4\}\})$ $(right)$—(see text)

$1, \ldots, m\}$ where $j_i \in [P]$ denotes the population, and the lineages $l_i \subseteq [L]$ form a partition of $[L]$. We denote the set of states $\mathfrak{S}$. The parameters of the IM-model are the migration rates, $m_{i \to j}$, the coalescent rates c_i, c_A and the split time T_A. For states $\alpha \neq \beta \in \mathfrak{S}$ and $0 \leq t < T_A$, the entries of the rate matrix of the CTMC are given by

$$Q_{\alpha,\beta} = \begin{cases} m_{i \to j} & \text{if } \alpha = \mathscr{S} \cup \{(i, l)\},\ \beta = \mathscr{S} \cup \{(j, l)\} \\ c_i & \text{if } \alpha = \mathscr{S} \cup \{(i, l_1)\} \cup \{(i, l_2)\},\ \beta = \mathscr{S} \cup \{(i, l_1 \cup l_2)\} \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

where $\mathscr{S}$ is of the form $\cup_i \{(j_i, l_i)\}$, and for $T_A \leq t$

$$\tilde{Q}_{\alpha,\beta} = \begin{cases} c_A & \text{if } \alpha = \mathscr{S} \cup \{(i, l_1)\} \cup \{(j, l_2)\},\ \beta = \mathscr{S} \cup \{(1, l_1 \cup l_2)\} \\ 0 & \text{otherwise} \end{cases}$$

where $i, j \in [P]$. In both cases the entries on the diagonal ($\alpha = \beta$) are determined by the requirement that the row sums are 0. Before T_A, lineages can only coalesce if they are in the same population, but migration is possible between populations. After T_A, we allow any pair of lineages to coalesce regardless of population, and no longer have migration events, so that $\{X_t\}_{t \geq T_A}$ behaves as a standard coalescent process. The probability of being in state β at time t given that the CTMC is in state α at time s is straightforward to compute, but depends on whether t and s are before or after T_A:

$$\mathbb{P}(X_t = \beta \mid X_s = \alpha) = \begin{cases} \left(\mathrm{e}^{Q(t-s)}\right)_{\alpha,\beta} & s < t < T_A \\ \sum_\gamma \left(\mathrm{e}^{Q(T_A - s)}\right)_{\alpha,\gamma} \left(\mathrm{e}^{\tilde{Q}(t - T_A)}\right)_{\gamma,\beta} & s < T_A \leq t \\ \left(\mathrm{e}^{\tilde{Q}(t-s)}\right)_{\alpha,\beta} & T_A \leq s < t. \end{cases} \tag{6}$$

The IM model allows us to assign a probability density to a *genealogy*, by which we mean the sample path traversed in the CTMC from the starting state s, until a

common ancestor is reached, meaning that we consider any migration events part of the genealogy. In order to define the coalescent trees, we introduce the mapping F defined on the states of the CTMC, $\mathfrak{S}$, and taking values in the partitions of $[L]$:

$$F\left(\cup_i \{(j_i, l_i)\}\right) = \cup_i \{l_i\}\ , \tag{7}$$

that is, F strips away population information from the states. Using F we define an equivalence relation on the sample paths of X by defining two sample paths $\{X(t)\}_{t\geq 0}$ and $\{X'(t)\}_{t\geq 0}$ to be equivalent if $\{F(X(t))\}_{t\geq 0} = \{F(X'(t))\}_{t\geq 0}$ and we call each equivalence class a *coalescent tree*. Coalescent trees are uniquely determined by the time of coalescence events and the set of lineages before and after these coalescences, and we represent a coalescent tree as $\mathscr{C}(\mathbf{t}, \mathbf{Y})$ where $\mathbf{t} = (t_i)_{i=1}^{L-1}$ is the vector of coalescence times with t_i denoting the i'th coalescent event and $\mathbf{Y} = (Y_i)_{i=1}^{L-1}$ is a vector such that Y_i are the lineages present immediately prior to t_i. Setting $t_0 = 0$, we have:

$$\{X(t)\}_{t\geq 0} \in \mathscr{C}(\mathbf{t}, \mathbf{Y}) \Leftrightarrow F(X(t)) = Y_i \quad \text{for } t_{i-1} \leq t < t_i.$$

We wish to calculate the probability density of a coalescent tree $\mathscr{C}(\mathbf{t}, \mathbf{Y})$, for $\mathbf{Y} = (Y_1, Y_2 \ldots, Y_{L-1})$. To do this, we sum over pairs of tuples of states $(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)$ where $\boldsymbol{\alpha}^1 = (\alpha_1^1, \alpha_2^1, \ldots, \alpha_{L-1}^1)$ is a tuple such that $\alpha_1^1 = s$ is the starting state, and $\alpha_i^1,\ i > 1$, are the possible states immediately after the $(i-1)$th coalescent event, and since these are the same as those immediately prior to the ith coalescent event we have $\alpha_i^1 \in F^{-1}(Y_i)$. The tuple $\boldsymbol{\alpha}^2 = (\alpha_1^2, \alpha_2^2, \ldots, \alpha_{L-1}^2)$ are the possible states immediately prior to the i'th coalescent event, that is $\alpha_i^2 \in F^{-1}(Y_i)$.

Thus we can compute the probability density of a coalescent tree, implicitly integrating over all elements in the equivalence class

$$f\left(\mathscr{C}(\mathbf{t}, \mathbf{Y})\right) = \sum_{\substack{(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2):\\ \alpha_1^1 = s \quad F(\alpha_1^2) = Y_1 \\ F(\alpha_i^1) = F(\alpha_i^2) = Y_i}} \mathbb{P}(\mathbf{t}, (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)) g(\mathbf{t}, (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)) \tag{8}$$

where

$$\mathbb{P}\left(\mathbf{t}, (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)\right) = \prod_{i=1}^{L-1} \mathbb{P}(X_{t_i} = \alpha_i^2 \mid X_{t_{i-1}} = \alpha_i^1)$$

using the transition probability calculated in (6), and

$$g(\mathbf{t}, (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)) = \prod_{i=1}^{L-2} \left(Q_{t_i}\right)_{\alpha_i^2, \alpha_{i+1}^1} \prod_{\beta: F(\beta) = [L]} \left(Q_{t_{L-1}}\right)_{\alpha_{L-1}^2, \beta}$$

where

$$(Q_s)_{x,y} = \begin{cases} Q_{x,y} & \text{if } s < T_A \\ \tilde{Q}_{x,y} & \text{if } T_A \leq s. \end{cases} \tag{9}$$

Note that (8) is a generalization of (3) and (4), and by way of being a sum over transition probabilities multiplied with coalescent rates, it has the same form. Not all vectors $\mathbf{t}$ and $\mathbf{Y}$ correspond to well-defined coalescent trees $\mathscr{C}(\mathbf{t}, \mathbf{Y})$. For $\mathbf{t}$ the requirement is that $\mathbf{t} \in \mathbb{R}^{L-1}$ with $t_1 < t_2 < \ldots t_{L-1}$, and for $\mathbf{Y}$, the requirement is that $Y_1 = \{\{i\} \mid 1 \leq i \leq L\}$ and for $i > 1$, Y_i is obtained by taking union of exactly two sets in Y_{i-1}. We use the notation $\mathfrak{Y}$ for the set of $\mathbf{Y}$, which correspond to well-defined coalescent trees.

3.2 Adding mutations

Next, we consider mutation models. We consider both an infinite sites model and a CTMC for nucleotide substitutions. First, we note that a given coalescent tree $\mathscr{C}(\mathbf{t}, \mathbf{Y})$ induces a binary weighted graph in an obvious way by taking as nodes the singleton sets $\{i\}$, $i = 1, \ldots, L$, and the lineages (as sets) formed at each coalescent event. The singleton sets are the leaf nodes, and whenever a coalescent event forms a lineage ℓ from ℓ_1 and ℓ_2, we put an edge between ℓ and ℓ_i, $i = 1, 2$. Hence, there are $2L - 1$ nodes, which we denote $V = (v_i)_{i=1}^{2L-1}$. We label $v_i = \{i\}$ for $1 \leq i \leq L$ and for $L < i \leq 2L - 1$, we label v_i with the lineage formed at the $(i - L)$'th coalescent event. In particular $v_{2L-1} = [L]$. We define the *age* of each node to be the time of the corresponding coalescent event, so that the age of v_i, $1 \leq i \leq L$, is 0 and the age of v_i $L < i \leq 2L-1$ is t_{i-L}. The weight of each edge is then the age of the parent, minus that of the child. Thus, we may write $\mathscr{C}(\mathbf{t}, \mathbf{Y}) = (V, E, W)$ where V, E and W are the nodes, edges and edge-weights respectively, of the induced graph, and we let w_e denote the weight of the edge e. Below, we discuss two mutation models: A Markov model for nucleotide substitutions and an infinite sites model. Since the former is not generally time-reversible, we will in this case consider the induced graph to be directed, with each edge going from parent to child. In the latter case, we consider the induced graph to be undirected.

3.2.1 Nucleotide substitutions

Suppose we wish to assign nucleotides to the nodes of a given coalescent tree. We assume that nucleotide substitutions follow a time-homogeneous Markov process with rate matrix given by $(\Delta)_{x,y}$, $x, y \in \{A, C, G, T\}$, and stationary distribution $\pi = (\pi_A, \pi_C, \pi_G, \pi_T)$. For a node $v_i \in V$ and nucleotide $y_i \in \{A, C, G, T\}$ we use the notation $v_i = y_i$ to mean the event that we observe y_i at v_i. Using this notation, the probability of observing a particular assignment of nucleotides y_i to nodes v_i of a given coalescent tree is:

$$\mathbb{P}(v_i = y_i, 1 \le i \le 2L-1 \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})) = \pi_{y_{2L-1}} \prod_{\substack{e \in E \\ e=(v_j, v_k)}} \left(e^{\Delta w_e}\right)_{y_j, y_k}. \quad (10)$$

When we wish to do parameter estimation, we face the problem that we cannot actually observe any of the nucleotide assignments in the past, so we sum these out:

$$\mathbb{P}(v_i = y_i, 1 \le i \le L \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})) = \sum_{\substack{y_i \in \\ \{A,C,G,T\} \\ L < i \le 2L-1}} \mathbb{P}(v_i = y_i, \quad 1 \le i \le 2L-1 \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})). \quad (11)$$

We combine (11) with (8) to obtain an unconditional formula:

$$\mathbb{P}(v_i = y_i, 1 \le i \le L) = \sum_{Y \in \mathfrak{Y}} \int_{\mathbf{t} \in \mathbb{T}} \sum_{\substack{y_i \in \\ \{A,C,G,T\} \\ L < i \le 2L-1}} \mathbb{P}(v_i = y_i 1 \le i \le 2L-1 \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})) f(\mathscr{C}(\mathbf{t}, \mathbf{Y})) \, \mathrm{d}\mathbf{t}, \quad (12)$$

where the domain of integration is $\mathbb{T} = \{\mathbf{t} \in \mathbb{R}^{L-1} \mid t_1 < t_2 < \cdots < t_{L-1}\}$.

3.2.2 Infinite sites model

Next, we consider an infinite sites model, and here we wish to assign a number of mutations to each edge of $\mathscr{C}(\mathbf{t}, \mathbf{Y}) = (V, E, W)$, which we take to be undirected since in this section, we wish to consider paths between leaves. In this case there is only one parameter in the mutation model, namely the mutation rate γ, and if we define $PO_\gamma(x, k) = (\gamma x)^k \mathrm{e}^{-\gamma x}/k!$, the probability of having k_e mutations on edge e is $PO_\gamma(w_e, k_e)$. We let $\mathbf{k} = (k_e, e \in E)$ denote a vector of non-negative integers indexed by the edges in the given coalescent tree. These integers correspond to a particular assignment of mutations on the branches of the tree. Then the probability of observing a particular assignment $\mathbf{k}$ is:

$$\mathbb{P}(\mathbf{k} \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})) = \prod_{e \in E} PO_\gamma(w_e, k_e). \quad (13)$$

Similarly to the previous section, we need to consider what is actually observable from the present. Here it is the total number of mutations on the path from node $\{i\}$ to node $\{j\}$, and we denote this number $d(i, j)$. Let $E(i, j)$ denote the set of edges on the path from $\{i\}$ to $\{j\}$ so that

$$d(i, j) = \sum_{e \in E(i,j)} k_e, \quad (14)$$

and let $\mathbf{d}$ denote the vector $(d(i, j) \mid i < j)$. We will call $\mathbf{d}$ *differences*, since for actual data it will be the count of nucleotide differences between genetic segments. At this

point, we have to deal with the identifiability issue, arising from the fact that mapping φ which maps $\mathbf{k}$ to $\mathbf{d}$ through Eq. (14) is not injective. When we wish to calculate the probability of observing a particular vector $\mathbf{d}$, we need to sum over the assignments of mutations which give rise to the observed differences:

$$\mathbb{P}(\mathbf{d} \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})) = \sum_{\mathbf{k}:\varphi(\mathbf{k})=\mathbf{d}} \mathbb{P}(\mathbf{k} \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})). \tag{15}$$

We may combine (15) with (8) to obtain an unconditional formula for observed differences:

$$\mathbb{P}(\mathbf{d}) = \sum_{Y \in \mathfrak{Y}} \int_{\mathbf{t} \in \mathbb{T}} \sum_{\mathbf{k}:\varphi(\mathbf{k})=\mathbf{d}} \mathbb{P}(\mathbf{k} \mid \mathscr{C}(\mathbf{t}, \mathbf{Y})) f\left(\mathscr{C}(\mathbf{t}, \mathbf{Y})\right) \mathrm{d}\mathbf{t}\,, \tag{16}$$

where, as in (12), $\mathbb{T} = \{\mathbf{t} \in \mathbb{R}^{L-1} \mid t_1 < t_2 < \cdots < t_{L-1}\}$.

4 State-space explosion and lumpability

We wish to count the number of states $|\mathfrak{S}|$. We do this by considering the elements in the image of F, which is the set of all partitions of $[L]$, and then counting $|\mathfrak{S}|$ by counting the number of elements in the pre-image of each element in $F(\mathfrak{S})$. The number of partitions $\rho \in F(\mathfrak{S})$ with $|\rho| = k$ for $1 \leq k \leq L$ is the Stirling Number of the second kind $S_2(L, k)$ (Stanley 2012). Since there are P populations, the number of elements in the pre-image of a partition with $|\rho| = k$ is $|F^{-1}(\rho)| = P^k$. Hence, we have

$$|\mathfrak{S}| = \sum_{k=1}^{L} \sum_{\substack{\rho \in F(\mathfrak{S}) \\ |\rho|=k}} P^k = \sum_{k=1}^{L} S_2(L, k) P^k = B_P(L)\,,$$

where $B_n(\cdot)$ is the n'th Bell polynomial. In Table 1 we give $B_2(\cdot)$ and $B_3(\cdot)$ evaluated at $L = 2, 3, 4, 5, 6$.

It is evident from Table 1 that simplifications are needed, in order for formulas such as (6) and (8), and in turn those in Sect. 3.2 to be useful for maximum likelihood estimation.

We present a simplification based on the concept of exact lumpability, and for ease of presentation, we restrict ourselves to the case $P = 2$, and sketch a generalization at

Table 1 The top two rows give the number of states in the full state-space for $P = 2, 3$ populations and $L = 2, 3, 4, 5, 6$. The bottom rows give the number of states in the reduced state-space for two and three populations

P/L	2	3	4	5	6
2	6	22	94	454	2,430
3	12	57	309	1,866	12,351
2	7	17	39	67	117
3	15	70	205	529	1,341

the end of the section. Intuitively, the simplification is based on the observation that for given states α, β with the same lineages (possibly distributed differently among populations), the probability of going from α to β in some time t only depends on the number of lineages switching populations, and not specifically *which* lineages switch populations. This means that we may calculate the desired probabilities by labeling the lineages with 1's or 2's and considering the new state-space $\hat{\mathfrak{S}}$, which consists of the relabeled states. Formally, we relabel the lineages by defining a map $\tilde{f}$ mapping the lineages to the set $\{1, 2\}$. This map then induces a map f from $\mathfrak{S}$ to $\hat{\mathfrak{S}}$ by

$$f : \mathfrak{S} \ni \alpha = \cup_i \{(j_i, l_i)\} \mapsto \cup_i \left\{\left(j_i, \tilde{f}(l_i)\right)\right\}.$$

Suppose that we wish to calculate a probability involving the density of a coalescent tree. It is evident from (8) that we need only consider probabilities of the form $\mathbb{P}(X_t = \beta \mid X_0 = \alpha) = \left(\mathrm{e}^{Qt}\right)_{\alpha,\beta}$ for states $\alpha, \beta \in \mathfrak{S}$, where $\beta \in \bar{\mathfrak{S}} := \{\gamma \in \mathfrak{S} \mid F(\alpha) = F(\gamma)\}$ Note that the states in $\bar{\mathfrak{S}}$ contain the same lineages. A first observation is that because α and β are transient states $\mathbb{P}(X_t = \beta \mid X_0 = \alpha)$ can be calculated by exponentiating the matrix $Q_{|\bar{\mathfrak{S}} \cup A}$, and a second observation, which is proved in Theorem 1 in the Appendix, is that this matrix is *exactly lumpable*. This implies—as it is proved in Corollary 1—that under the additional assumption $|f^{-1}(f(\alpha))| = 1$ we have

$$\left(\mathrm{e}^{Qt}\right)_{\alpha,\beta} = \mathbb{P}(X_t = \beta \mid X_0 = \alpha) = \frac{1}{|f^{-1}(f(\beta))|} \left(\mathrm{e}^{\hat{Q}_{|f(\bar{\mathfrak{S}})} t}\right)_{f(\alpha), f(\beta)} \tag{17}$$

where $\hat{Q}$ is defined in (32) in the Appendix.

As an example of relabelling, let $\alpha = \{(1, \{1\}), (1, \{2\}), (2, \{3\}), (2, \{4\})\}$ be the starting state in the case of two diploid individuals in two populations, which was also considered in Sect. 3, and let $\beta = \{(1, \{1\}), (2, \{2\}), (1, \{3\}), (2, \{4\})\}$ be the state where lineages $\{2\}$ and $\{3\}$ have migrated to population 2 and 1 respectively. In order to apply formula (17) we need to find $\tilde{f}$ such that $|f^{-1}(f(\alpha))| = 1$. This is achieved by setting

$$\tilde{f}(\{1\}) = 1 \quad \tilde{f}(\{2\}) = 1 \quad \tilde{f}(\{3\}) = 2 \quad \tilde{f}(\{4\}) = 2.$$

As noted, we need only consider exponentiation of the matrix $Q_{|\bar{\mathfrak{S}}}$, so we do not need to define $\tilde{f}$ on any of the other possible linages, e.g., $\{1, 2\}$. With this choice of $\tilde{f}$ we have $f(\beta) = \{(1, \{1\}), (2, \{1\}), (1, \{2\}), (2, \{2\})\}$, which is one of 9 states in $f(\bar{\mathfrak{S}}) \subset \hat{\mathfrak{S}}$. The remaining states are found by distributing the 2 1's and the 2 2's among the 2 populations. For arbitrary α and β, the condition $|f^{-1}(f(\alpha))| = 1$ is fulfilled by letting $\tilde{f}$ assign 1's to the lineages in the population with the most lineages in α and 2's to the lineages in the other population. If the two populations contain the same number of lineages, we assign 1's to the lineages in population 1 and 2's to those in population 2. We can now count the number of elements in the state-space $\hat{\mathfrak{S}}$, which consists of the union of states found, when applying the labelling scheme described above. To do this, we start by noticing that the number of ways m 2's and n

1's can be divided among the 2 populations is $(m+1)(n+1)$, so if we wish to count $|\{\hat{\alpha} \in \hat{\mathfrak{S}} \mid |\hat{\alpha}| = k\}|$, we have to count the number of assignments of m 2's and n 1's where $0 \leq m \leq n$ and $m+n=k$. This implies that we get two expressions for $|\{\hat{\alpha} \in \hat{\mathfrak{S}} \mid |\alpha| = k\}|$ depending on the parity of k:

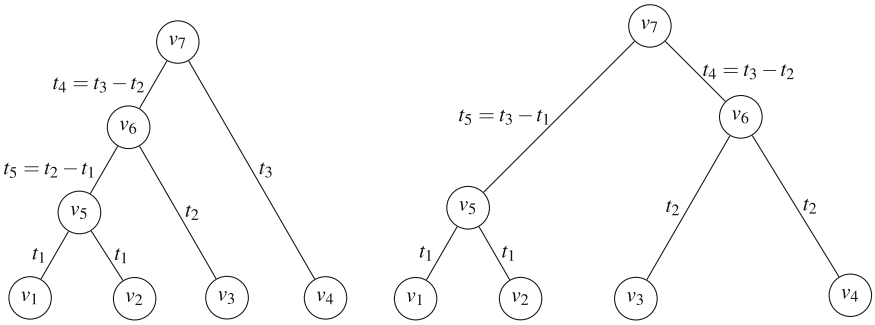
$$
\begin{aligned}
|\{\hat{\alpha} \in \hat{\mathfrak{S}} \mid |\alpha| = k\}| &= \sum_{\substack{0 \leq m \leq n: \\ m+n=k}} (m+1)(n+1) \\
&= \begin{cases} \sum_{m=0}^{k/2} (m+1)(k-m+1) = \frac{1}{24}(2+k)(4+k)(3+2k) & \text{for } k \text{ even} \\ \sum_{m=0}^{(k-1)/2} (m+1)(k-m+1) = \frac{1}{12}(1+k)(2+k)(3+k) & \text{for } k \text{ odd.} \end{cases}
\end{aligned}
$$

Hence, $|\hat{\mathfrak{S}}|$ depends on the parity of L and assuming L is even we find:

$$
\begin{aligned}
|\hat{\mathfrak{S}}| = \sum_{k=2}^{L} |\{\hat{\alpha} \in \hat{\mathfrak{S}} \mid |\alpha| = k\}| &= \sum_{k=1}^{L/2-1} |\{\hat{\alpha} \in \hat{\mathfrak{S}} \mid |\alpha| = 2k+1\}| \\
&+ \sum_{k=1}^{L/2} |\hat{\alpha} \in \hat{\mathfrak{S}} \mid |\alpha| = 2k\}| = \frac{1}{48}\left(-96 + 76L + 44L^2 + 11L^3 + L^4\right)
\end{aligned}
$$

and similarly, we find the expression $\frac{1}{48}(-114 + 61L + 4L^2 + 11L^3 + L^4)$ when L is odd. In Table 1 we list the number of elements of $\hat{\mathfrak{S}}$ for $P = 2, 3$ and $L = 2, 3, 4, 5, 6$. We see that the our lumping procedure does not reduce the size of the state-space under consideration in the case of two lineages, but that $|\hat{\mathfrak{S}}|$ is of the order $L^4/48$ and by comparing to the first rows of the same table (which increases super-exponentially), we note a very significant reduction in the number of states as L increases.

Generalizing the lumpability framework beyond the case of two populations is achievable, and can be treated by generalizing Theorem 1 to allow for P populations as well as the relabelling mapping $\tilde{f}$ to take values in an arbitrary set. That such a generalization is possible is reasonable in light of the fact that the proof of Theorem 1 relies on the rates of coalescence and migration being dependent only on the distribution of lineages among populations and not on which specific lineages are present in which populations. Let us consider the case of three populations: In this case the requirement that $|f^{-1}(f(\alpha))| = 1$ may be fulfilled similarly to the two population case by labelling the linages in α in descending order with respect to the number of lineages in each population, so that lineages in the population with most lineages get the label 1, lineages in the population with second-most lineages (if any) get the label 2, and the lineages in the remaining population (if any) get the label 3. The size of the induced state-space, can be calculated in a manner similar to the two population case: We first calculate the number of states $\hat{\alpha}$ with $|\hat{\alpha}| = k$ by noting that the number of ways, say, the n 1's can be distributed among the three populations is $\binom{n+2}{n}$, so if we assume we have n 1's, m 2's and l 3's with $l \leq m \leq n$ and $l+n+m=k$, then the number of such states is found by summing $\binom{l+2}{l}\binom{m+2}{m}\binom{n+2}{n}$ over the set



(a) Topology A with branch lengths and nucleotide assignments

(b) Topology B with branch lengths and nucleotide assignments

Fig. 5 The two topologies with branch lengths and nucleotide assignments corresponding to $\mathbf{Y}_1 = (\{\{1,2\},\{3\},\{4\}\},\{\{1,2,3\},\{4\}\})$ $(left)$ and $\mathbf{Y}_2 = (\{\{1,2\},\{3\},\{4\}\},\{\{1,2\},\{3,4\}\})$ $(right)$ – (see text)

$0 \le l \le m \le n$, $0 \le l+n+m = k$. The size of the state-space is then found by summing the number of $\hat{\alpha}$ with $|\hat{\alpha}| = k$ from 2 to L. We refrain from performing the explicit calculation, but provide the sizes of the reduced state-space for $L = 2, 3, 4, 5, 6$ in Table 1, and we note, as in the two population case, a polynomial order of increase.

5 Two diploid individuals in two populations

The aim of the this section is to derive an explicit expression for the log-likelihood function for data $\mathscr{X}$ given as alignment columns from two diploid individuals in two populations, where we assume a time-reversible nucleotide substitution model. We further assume free recombination between loci, and hence computing the log-likelihood of a given data set amounts to computing the probabilities of the 256 possible genotype combinations. This means expanding on the formulas (10), (11) and (12). The expressions derived here were implemented and applied in Scally et al. (2012). The extant lineages are labeled $\{1\}, \{2\}$ for the lineages in population 1 and $\{3\}, \{4\}$ for the lineages in population 2. It follows from the definition of coalescent trees that there are $\prod_{i=2}^{L} \binom{i}{2}$ well-defined coalescent trees for fixed branch lengths, and for $L = 4$ this means that $|\mathfrak{Y}| = 18$. Each coalescent tree induces a weighted graph which belongs to one of two topologies, shown in Fig. 4. In Fig. 5 we name the two topologies A and B and label the nodes v_i as described in Sect. 3.2. We wish to calculate the probability of observing $y_i \in \{A, C, G, T\}$ at node v_i, $1 \le i \le 4$. This probability was calculated in (12), and we have

$$\mathbb{P}(v_1 = y_1, v_2 = y_2, v_3 = y_3, v_4 = y_4) = \sum_{\mathbf{Y} \in \mathfrak{Y}} \int_0^{\infty} \int_0^{t_3} \int_0^{t_2} \mathbb{P}(v_1 = y_1, v_2 = y_2, v_3 = y_3, v_4 = y_4 | \mathscr{C}(\mathbf{t}, \mathbf{Y}))\, f(\mathscr{C}(\mathbf{t}, \mathbf{Y}))\, \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3. \tag{18}$$

We calculate the inner probability for the two coalescent trees $\mathscr{C}(\mathbf{t}, \mathbf{Y_i}), i = 1, 2$ shown in Fig. 5, where

$$\mathbf{Y}_1 = (\{\{1, 2\}, \{3\}, \{4\}\}, \{\{1, 2, 3\}, \{4\}\})$$
$$\mathbf{Y}_2 = (\{\{1, 2\}, \{3\}, \{4\}\}, \{\{1, 2\}, \{3, 4\}\}),$$

and $\mathbf{t} = (t_1, t_2, t_3)$. The expressions for the remaining 16 are found by permuting the nucleotide assignments on the leaves in each topology. Using formula (10) we find:

$$\mathbb{P}(v_i = y_i\,,\ i = 1 \ldots 7 \mid \mathscr{C}(\mathbf{t}, \mathbf{Y}_j)) = \begin{cases} \pi_{y_7} \left(\mathrm{e}^{\Delta t_3}\right)_{y_7,y_4} \left(\mathrm{e}^{\Delta t_4}\right)_{y_7,y_6} \left(\mathrm{e}^{\Delta t_2}\right)_{y_6,y_3} \left(\mathrm{e}^{\Delta t_5}\right)_{y_6,y_5} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_1} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_2} & \text{if} \quad j = 1 \\ \pi_{y_7} \left(\mathrm{e}^{\Delta t_4}\right)_{y_7,y_6} \left(\mathrm{e}^{\Delta t_5}\right)_{y_7,y_5} \left(\mathrm{e}^{\Delta t_2}\right)_{y_6,y_4} \left(\mathrm{e}^{\Delta t_2}\right)_{y_6,y_3} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_2} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_2} & \text{if} \quad j = 2. \end{cases}$$

We continue our calculations for $\mathbf{Y}_1$. We can only observe extant lineages so we sum over y_5, y_6, y_7 in the expression above:

$$\sum_{\substack{y_5,y_6,y_7 \in \\ \{A,C,G,T\}}} \mathbb{P}(v_i = y_i\,,\ i = 1 \ldots 7 \mid \mathscr{C}(\mathbf{t}, \mathbf{Y}_1))$$

$$= \sum_{\substack{y_5,y_6,y_7 \in \\ \{A,C,G,T\}}} \pi_{y_7} \left(\mathrm{e}^{\Delta t_3}\right)_{y_7,y_4} \left(\mathrm{e}^{\Delta t_4}\right)_{y_7,y_6} \left(\mathrm{e}^{\Delta t_2}\right)_{y_6,y_3} \left(\mathrm{e}^{\Delta t_5}\right)_{y_6,y_5} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_1} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_2} \tag{19}$$

$$= \sum_{\substack{y_5,y_6,y_7 \in \\ \{A,C,G,T\}}} \pi_{y_4} \left(\mathrm{e}^{\Delta t_3}\right)_{y_4,y_7} \left(\mathrm{e}^{\Delta t_4}\right)_{y_7,y_6} \left(\mathrm{e}^{\Delta t_2}\right)_{y_6,y_3} \left(\mathrm{e}^{\Delta t_5}\right)_{y_6,y_5} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_1} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_2} \tag{20}$$

$$= \sum_{\substack{y_5,y_6 \in \\ \{A,C,G,T\}}} \pi_{y_4} \left(\mathrm{e}^{\Delta (t_3+t_4)}\right)_{y_4,y_6} \left(\mathrm{e}^{\Delta t_2}\right)_{y_6,y_3} \left(\mathrm{e}^{\Delta t_5}\right)_{y_6,y_5} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_1} \left(\mathrm{e}^{\Delta t_1}\right)_{y_5,y_2} \tag{21}$$

where we use the assumption of time-reversibility in the step from Eqs. (19) to (20). The assumption of time-reversibility also ensures that Δ is diagonalizable (see Keilson 1979). This implies that we can write $\mathrm{e}^{\Delta t} = \bar{V} \mathrm{e}^{\bar{D} t} \bar{V}^{-1}$ where $\bar{V} = (\bar{v})_{x,y}$, $x, y \in \{A, C, G, T\}$ is a matrix of eigenvectors, $\bar{V}^{-1} = (\bar{v}^{-1})_{x,y}$ its inverse, and $\bar{D}$ is a diagonalmatrix containing the corresponding eigenvalues of Δ, which we denote $\bar{\lambda}_i, i \in \{A, C, G, T\}$. This gives us:

$$\left(\mathrm{e}^{\Delta t}\right)_{x,y} = \sum_{i \in \{A,C,G,T\}} \bar{v}_{x,i} \bar{v}^{-1}_{i,y} \mathrm{e}^{\bar{\lambda}_i t},$$

and using this, we may continue our calculation of (21):

$$(21) = \sum_{\substack{y_5,y_6 \in \\ \{A,C,G,T\}}} \sum_{\substack{i,j,k,l,m \in \\ \{A,C,G,T\}}} \Big(\pi_{y_4} \bar{v}_{y_4,i} \bar{v}^{-1}_{i,y_6} \mathrm{e}^{\lambda_i (t_3+t_4)} \bar{v}_{y_6,j} \bar{v}^{-1}_{j,y_3} \mathrm{e}^{\lambda_j t_2} \cdot \bar{v}_{y_6,k} \bar{v}^{-1}_{k,y_5} \mathrm{e}^{\lambda_k t_5} \bar{v}_{y_5,l} \bar{v}^{-1}_{l,y_1} \bar{v}_{y_5,m} \bar{v}^{-1}_{m,y_2} \mathrm{e}^{(\lambda_l+\lambda_m) t_1}\Big). \tag{22}$$

Defining

$$C_1 := \sum_{\substack{y_5,y_6\in \\ \{A,C,G,T\}}} \pi_{y_4}\bar{v}_{y_4,i}\bar{v}^{-1}_{i,y_6}\bar{v}_{y_6,j}\bar{v}^{-1}_{j,y_3}\bar{v}_{y_6,k}\bar{v}^{-1}_{k,y_5}\bar{v}_{y_5,l}\bar{v}^{-1}_{l,y_1}\bar{v}_{y_5,m}\bar{v}^{-1}_{m,y_2}, \tag{23}$$

and using that $t_4 = t_3 - t_2$ and $t_5 = t_2 - t_1$ we find:

$$(22) = \sum_{\substack{i,j,k,l,m\in \\ \{A,C,G,T\}}} C_1 \mathrm{e}^{\bar{\lambda}_i(2t_3-t_2)}\mathrm{e}^{(\bar{\lambda}_j+\bar{\lambda}_k)t_2}\mathrm{e}^{(\bar{\lambda}_l+\bar{\lambda}_m-\bar{\lambda}_k)t_1} =: \sum_{\substack{i,j,k,l,m\in \\ \{A,C,G,T\}}} C_1 f_1(t_1,t_2,t_3) \tag{24}$$

where we define function f_1 to be

$$f_1(t_1,t_2,t_3) := \mathrm{e}^{\bar{\lambda}_i(2t_3-t_2)}\mathrm{e}^{(\bar{\lambda}_j+\bar{\lambda}_k)t_2}\mathrm{e}^{(\bar{\lambda}_l+\bar{\lambda}_m-\bar{\lambda}_k)t_1}. \tag{25}$$

Similarly, we can find an expression when we condition on topology B:

$$\mathbb{P}(v_i = y_i\,,\ i = 1\ldots 4 \mid \mathscr{C}(\mathbf{t},\mathbf{Y}_2)) = \sum_{\substack{i,j,k,l,m\in \\ \{A,C,G,T\}}} C_2 \mathrm{e}^{\bar{\lambda}_i 2t_3}\mathrm{e}^{(\bar{\lambda}_j+\bar{\lambda}_k-\bar{\lambda}_i)t_2}\mathrm{e}^{(\bar{\lambda}_l+\bar{\lambda}_m-\bar{\lambda}_i)t_1} =: \sum_{\substack{i,j,k,l,m\in \\ \{A,C,G,T\}}} C_2 f_2(t_1,t_2,t_3) \tag{26}$$

where

$$C_2 := \sum_{\substack{y_5,y_6\in \\ \{A,C,G,T\}}} \pi_{y_5}\bar{v}_{y_5,i}\bar{v}^{-1}_{i,y_6}\bar{v}_{y_5,j}\bar{v}^{-1}_{j,y_1}\bar{v}_{y_5,k}\bar{v}^{-1}_{k,y_2}\bar{v}_{y_6,l}\bar{v}^{-1}_{l,y_3}\bar{v}_{y_6,m}\bar{v}^{-1}_{m,y_4} \tag{27}$$

and similar to f_1 we define:

$$f_2(t_1,t_2,t_3) := \mathrm{e}^{\bar{\lambda}_i 2t_3}\mathrm{e}^{(\bar{\lambda}_j+\bar{\lambda}_k-\bar{\lambda}_i)t_2}\mathrm{e}^{(\bar{\lambda}_l+\bar{\lambda}_m-\bar{\lambda}_i)t_1}. \tag{28}$$

The calculation of the conditional probabilities for the remaining 16 trees consists of permuting y_i $i = 1, 2, 3, 4$ in the definitions (23) and (27) of C_1 and C_2. Next, we turn our attention to $f(\mathscr{C}(\mathbf{t},\mathbf{Y}))$. This is given by formula (8) and we divide the formula into cases depending on how many coalescent events take place before T_A.

$$f(\mathscr{C}(\mathbf{t},\mathbf{Y})) = \sum \left(\mathrm{e}^{Qt_1}\right)_{s,\alpha_1^2}\left(\mathrm{e}^{Q(t_2-t_1)}\right)_{\alpha_2^1,\alpha_2^2}\left(\mathrm{e}^{Q(t_3-t_2)}\right)_{\alpha_3^1,\alpha_3^2} Q_{\alpha_1^2,\alpha_2^1}Q_{\alpha_2^2,\alpha_3^1}c_{\alpha_3^2} \quad \text{if} \quad t_3 < T_A$$

$$\sum\sum_{\beta\in\mathfrak{S}} \left(\mathrm{e}^{Qt_1}\right)_{s,\alpha_1^2}\left(\mathrm{e}^{Q(t_2-t_1)}\right)_{\alpha_2^1,\alpha_2^2}\left(\mathrm{e}^{Q(T_A-t_2)}\right)_{\alpha_3^1,\beta}\mathrm{e}^{-c_A(T_A-t_3)}Q_{\alpha_1^2,\alpha_2^1}Q_{\alpha_2^2,\alpha_3^1}c_A \quad \text{if} \quad t_2 < T_A \le t_3$$

$$\sum_{\beta\in\mathfrak{S}}\left(\mathrm{e}^{Qt_1}\right)_{s,\alpha_1^2}\left(\mathrm{e}^{Q(T_A-t_1)}\right)_{\alpha_2^1,\beta}\mathrm{e}^{-3c_A(t_2-T_A)}\mathrm{e}^{-c_A(t_3-t_2)}Q_{\alpha_1^2,\alpha_2^1}3c_A^2 \quad \text{if} \quad t_1 < T_A \le t_2$$

$$\sum_{\beta\in\mathfrak{S}}\left(\mathrm{e}^{Qt_1}\right)_{s,\beta}\mathrm{e}^{-6c_A(t_1-T_A)}\mathrm{e}^{-c_A(t_3-t_2)}18c_A^3 \quad \text{if} \quad T_A \le t_1.$$

where the outer sum in all cases is over the set

$$\{(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2)\mid \alpha_1^1 = s \quad F(\alpha_1^2) = Y_1 \quad F(\alpha_i^1) = F(\alpha_i^2) = Y_i\}$$

(as in (8)) and $c_{\alpha_3^2} = c_i$ if has α_3^2 has both its lineages in population i and 0 otherwise. Each of the matrix exponentials above is further simplified using (17), and since all the involved matrices are diagonalizable (see Sect. 8.2 in the Appendix) we apply the spectral decomposition to rewrite the expression above to

$$f(\mathscr{C}(\mathbf{t},\mathbf{Y})) = \begin{cases} \sum C_3 h_1(t_1,t_2,t_3) & t_3 < T_A \\ \sum C_4 h_2(t_1,t_2,t_3) & t_2 < T_A \le t_3 \\ \sum C_5 h_3(t_1,t_2,t_3) & t_1 < T_A \le t_2 \\ \sum C_6 h_4(t_1,t_2,t_3) & T_A \le t_1\,, \end{cases} \tag{29}$$

where the constants C_3, C_4, C_5, C_6 depend only on the eigenvectors of the lumped matrices $\hat{Q}$ and the coalescent rates c_1, c_2. The h_i functions are given by:

$$\begin{aligned} h_1(t_1,t_2,t_3) &:= \mathrm{e}^{\lambda_r t_1}\mathrm{e}^{\lambda_s(t_2-t_1)}\mathrm{e}^{\lambda_t(t_3-t_2)} \\ h_2(t_1,t_2,t_3) &:= \mathrm{e}^{\lambda_r t_1}\mathrm{e}^{\lambda_s(t_2-t_1)}\mathrm{e}^{\lambda t(T_A-t_2)}c_A\mathrm{e}^{-c_A(t_3-T_A)} \\ h_3(t_1,t_2,t_3) &:= \mathrm{e}^{\lambda_r t_1}\mathrm{e}^{\lambda_s(T_A-t_1)}c_A^2\mathrm{e}^{-3c_A(t_2-T_A)}\mathrm{e}^{-c_A(t_3-t_2)} \\ h_4(t_1,t_2,t_3) &:= \mathrm{e}^{\lambda_r T_A}c_A^3\mathrm{e}^{-6c_A(t_1-T_A)}\mathrm{e}^{-3c_A(t_2-t_1)}\mathrm{e}^{-c_A(t_3-t_2)}\,, \end{aligned}$$

where $\lambda_\cdot$ are the eigenvalues of the lumped matrix. We may now finish the calculation of (18) by combining (29) with (24) and (26). We see that the resulting integral involves only regular exponential functions, which, although notationally cumbersome, are easily integrated explicitly by dividing the integral into the cases $t_3 < T_A$, $t_2 < T_A \le t_3$, $t_1 < T_A \le t_2$ and $T_A \le t_1$. The involved integrals are calculated in Sect. 8.4 in the Appendix.

6 Simulation study

In this section we provide a simulation study, in order to examine how well our method is able to recover parameters in the case considered in Sect. 5. Time is scaled by the neutral mutation rate μ, so that we expect one mutation during one unit of time, and we let g denote the generation length. This scaling implies that coalescence occurs

in population i with rates $c_i = 1/(2N_i\mu g)$ for $i = 1, 2$ and $c_A = 1/(2N_A\mu g)$ for the ancestral population. The migration rates are scaled so that $m_{1\to 2} = M_{1\to 2}/(\mu g)$ and $m_{2\to 1} = M_{2\to 1}/(\mu g)$ where $M_{i\to j}$ is the proportion of population i expected to migrate to population j per generation. We performed simulation experiments with nine different sets of parameter values in a model where $c_1 = c_A$ and $m_{1\to 2} = m_{2\to 1}$. Our parameters are chosen to resemble those of the Eastern and Western Gorilla found in Scally et al. (2012), which means that $N_1 = N_A = 35,000$, $N_2 = 16,000$, $g = 20$ years and $\mu = 0.6 \cdot 10^{-9}$ expected mutations pr. year. The migration rate $m_{1\to 2} = m_{2\to 1} = m$ belongs to the set $\{125, 250, 500\}$ and T_A belongs to the set the set $\{10^{-4}, 3 \cdot 10^{-4}, 6 \cdot 10^{-4}\}$. The parameters $m = 250$ and $T_A = 3 \cdot 10^{-4}$ correspond to those found by maximum likelihood estimation in Scally et al. (2012), and hence we examine the performance of our method in a neighborhood of these parameters. The substitution matrix was estimated from the data of Scally et al. (2012). For each of the 9 sets of parameter values we performed 5 simulations, each containing 10^7 loci, where each locus is an alignment column i.e. one basepair long. When maximizing the likelihood function we used the Nelder-Mead simplex method with five different starting values chosen uniformly in the interval $[0.5\tau, 1.5\tau]$ where τ is the true parameter. The results of the simulation study is displayed in Figs. 6 and 7 and Table 2.

Figure 6 shows likelihood surfaces for migration and split time parameters from one of the five experiments (the remaining parameters are fixed at their true values). We see that for a small split time ($T_A = 10^{-4}$), we are unable to recover the migration rate, while the split time itself is easily recovered. For the split time ($T_A = 3 \cdot 10^{-4}$) we are able to recover both migration rates and split times reasonably well, and for the large split time ($T_A = 6 \cdot 10^{-4}$), we are able to recover both migration rate and split time very well.

The likelihood surfaces in Fig. 6 are made under the assumption that the true population sizes are known. If we wish to estimate all four parameters of the model we see the flat likelihood makes parameter estimation of migration rates difficult for small values of T_A. In Table 2 we give the standardized biases ($\mathbb{E}[\hat{\tau} - \tau]/\tau$) and standardized root-mean-square errors ($\sqrt{\mathbb{E}[(\hat{\tau} - \tau)^2]}/\tau$). We see that the coalescence rates are very well estimated for all parameters, and that the split times are reasonably well estimated. We also note that for the smaller values of migration, the estimate of T_A when $T_A = 10^{-4}$ is actually better than the estimate when $T_A = 3 \cdot 10^{-4}$, while this is not the case for the large value of the migration rate. Finally we note that we are unable to recover the migration rate when $T_A = 10^{-4}$, and the estimates are still quite poor when $T_A = 3 \cdot 10^{-4}$. Only for the large value of the split time are we able to properly recover the migration rate.

7 Discussion

In this paper we have considered an IM model in which a panmictic ancestral population split into P subpopulations at some time T_A in the past, and we have shown how to deal with the issues which arise, when one wants to incorporate an arbitrary number of lineages and an arbitrary number of populations. We have shown how to define unions of genealogies – so-called coalescent trees – which are relevant for combination with

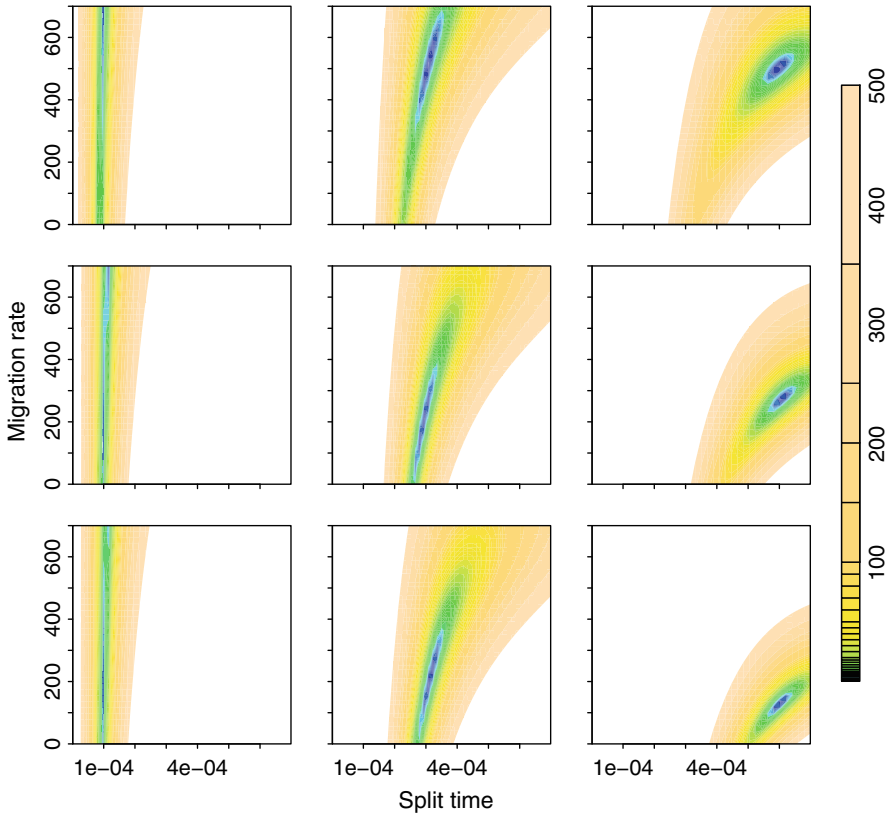


Fig. 6 log-likelihood surfaces—the true split times are: $10^{-4}, 3 \cdot 10^{-4}, 6 \cdot 10^{-4}$ (*columns from left to right*). The true migration rates are (*rows from bottom to top*) 125, 250, 500

different mutation models, and we have derived analytical explicit expressions for the probability densities of such unions. Furthermore, we have combined these expressions with mutation models and derived analytical expressions, which are relevant for maximum likelihood estimation. The derived expressions are generalizations to an arbitrary number of lineages of expressions found in the literature (e.g. Takahata et al. 1995 and Wilkinson-Herbots 2008). We have implemented these expressions in the case of four lineages in two populations. Extension beyond 4 lineages is possible, but is made difficult by the inherent complexity of the involved expressions (see Sect. 8.4) as well as fact the number of unlabeled topologies increases. In our case there were 2 unlabeled topologies (see Fig. 5). For five lineages there are three unlabeled topologies, for six lineages there are six topologies and for seven there are eleven (see Rosenberg 2007), so the number of different expressions one needs to consider increases considerably. In Sect. 2 we derived an explicit expression for the probability of observing k mutations in the case of 2 lineages with an infinite sites model, and we saw that these expressions involve the lower and upper incomplete Gamma functions. Extension beyond two lineages is in this case made difficult by the need to consider more complicated special functions, namely the confluent

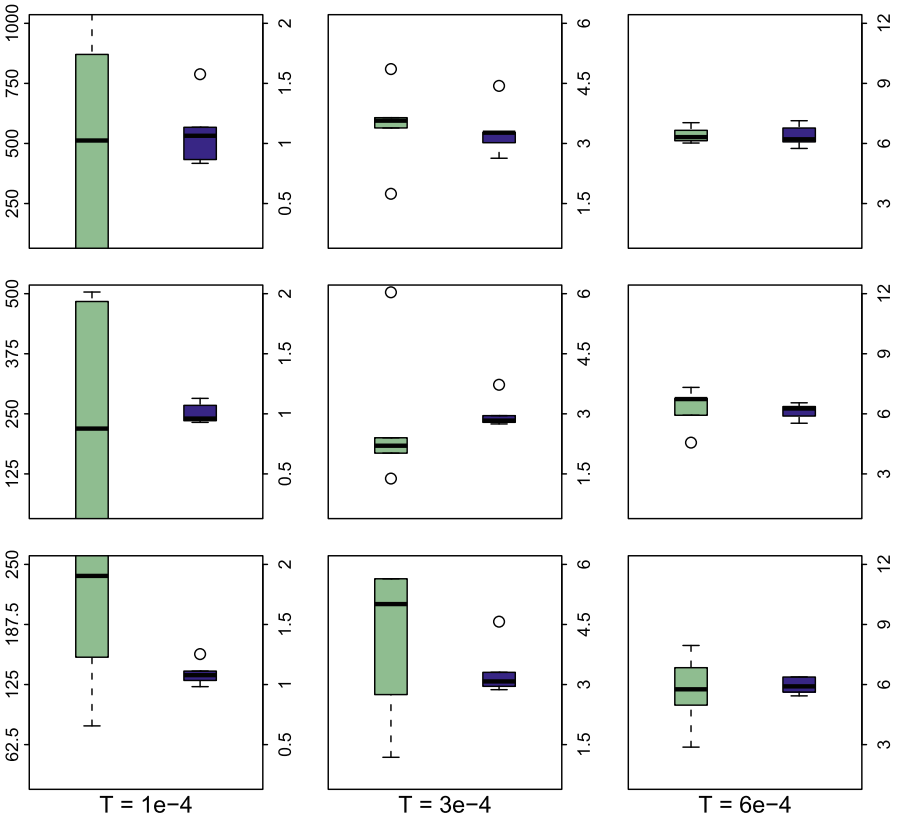


Fig. 7 Boxplots for maximum likelihood estimates of the migration rate (*green*, axis on *left* of each subplot) and the split time (*blue*, axis on *right* of each subplot, which is labeled in units of 10^{-4}). The true split times are: 10^{-4} , $3 \cdot 10^{-4}$, $6 \cdot 10^{-4}$ (*columns from left to right*). The true migration rates are (*rows from bottom to top*) 125 , 250 , 500

hypergeometric function ${}_1F_1$ (Abramowitz and Stegun 1984). Extensions to our model is also possible. In particular, it is relatively easy to extend our model to a situation where P extant populations have split from ancestral populations at different split times in the past. This can be achieved by defining the appropriate number of rate matrices like (5) and then changing formula (6) accordingly. By setting the appropriate migration rates to 0 this will also allow us to extend our model to the "isolation with an initial period of migration" model considered in Wilkinson-Herbots (2012). Furthermore, it is easy to define other models of divergence, e.g. the time-dependent migration rates of Innan and Watanabe (2006), as a CTMC on the state-space considered in this paper, although explicit expressions are only available in the case of piecewise constant rates considered in this paper.

Acknowledgments The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The research of LNA was supported by the Lundbeck Foundation, and the Villum Foundation.

Table 2 standardized biases and root-mean-square errors. The 9 values of each subtable correspond to the true split times: 10^{-4} , $3 \cdot 10^{-4}$, $6 \cdot 10^{-4}$ (columns from left to right), and true migration rates (rows from bottom to top) 125 , 250 , 500

Empirical standardized biases					
c_1			m		
−0.0005	0.01132	0.01133	0.18149	0.14711	0.07229
0.00182	0.00010	0.00578	−0.02725	−0.06350	0.04417
0.01073	0.01621	−0.00194	1.68154	0.78169	−0.05321
c_2			T_A		
0.01642	0.00191	0.00461	0.09536	0.10925	0.06488
−0.00566	0.00532	0.00916	0.00627	0.00336	0.020046
−0.04709	−0.00233	0.00266	0.09230	0.11894	−0.00917
Empirical standardized root-mean-square errors					
c_1			m		
0.01723	0.02642	0.02558	1.18510	0.36274	0.09507
0.00738	0.01895	0.01861	0.87655	0.55347	0.16563
0.01612	0.03071	0.01600	2.82947	1.47540	0.29252
c_2			T_A		
0.10623	0.015978	0.01847	0.28313	0.22908	0.10532
0.05253	0.020269	0.01280	0.07927	0.12194	0.06407
0.06348	0.02109	0.01011	0.13028	0.23995	0.06464

8 Appendix

8.1 Notation

For a vector or matrix $\mathbf{x}$ we let $\mathbf{x}^*$ denote the transpose of $\mathbf{x}$. For a set A we let $|A|$ denote the number of elements in A, and we use the notation $[n] := \{1, 2, \ldots, n\}$ to mean the natural numbers less than or equal to n, excluding 0. A *partition* of a set Ω is a set of non-empty, disjoint subsets of Ω, whose union is Ω. We let $I(\cdot)$ denote the indicator function. For a vector $\mathbf{v}$ we let $\text{diag}(\mathbf{v})$ denote the diagonal matrix with the entries of $\mathbf{v}$ on the diagonal.

Let $M = \{m_{s,s'}\}$, $s, s' \in S$ be a matrix indexed by a set S and let $S_0 \subseteq S$ be a subset of S. We let $M_{|S_0}$ denote the submatrix of M indexed by the elements of S_0 i.e.

$$\left(M_{|S_0}\right)_{s,s'} := m_{s,s'} \qquad s, s' \in S_0. \tag{30}$$

Occasionally, we wish to consider matrices, which are not proper submatrices of M, because they include an absorbing dummy state A which is not an element of S, so that $S_1 = S_0 \cup A$. In this case:

$$\left(M_{|S_1}\right)_{s,s'} = \begin{cases} m_{s,s'} & s, s' \in S_0 \\ 0 & s = A, s' \in S_0 \\ -\sum_{s'' \in S_0} m_{s,s''} & s \in S_0, s' = A. \end{cases} \tag{31}$$

8.2 On diagonalization of rate matrices

First, consider a *tridiagonal matrix* Q:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & 0 & 0 & \dots & 0 & 0 \\ q_{2,1} & q_{2,2} & q_{2,3} & 0 & \dots & 0 & 0 \\ 0 & q_{3,2} & q_{3,3} & q_{3,4} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & q_{m,m-1} & q_{m,m} \end{pmatrix}$$

and assume $q_{i+1,i}, q_{i,i+1} > 0$. Note that Q_Υ in Sect. 2 is of this form. Such a matrix is diagonalizable. This is seen by defining the vector $\mathbf{d}$ to have entries $d_1 = 1$ and $d_{i+1} = \sqrt{\frac{q_{i+1,i}}{q_{i,i+1}}} d_i$. Then

$$\left(\mathrm{diag}(\mathbf{d})^{-1} \cdot Q \cdot \mathrm{diag}(\mathbf{d})\right)_{i,j} = \begin{cases} q_{i,i} & \text{if} \quad i = j \\ \frac{d_{i+1}}{d_i} q_{i,i+1} = \sqrt{q_{i+1,i} q_{i,i+1}} & \text{if} \quad j = i+1 \\ \frac{d_i}{d_{i+1}} q_{i+1,i} = \sqrt{q_{i+1,i} q_{i,i+1}} & \text{if} \quad j = i-1 \\ 0 & \text{otherwise} \end{cases}.$$

We see that Q is similar to a symmetric matrix and hence it is diagonalizable. More generally we see that a *tri-diagonal block matrix*

$$Q = \begin{pmatrix} Q_{1,1} & Q_{1,2} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ Q_{2,1} & Q_{2,2} & Q_{2,3} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Q_{3,2} & Q_{3,3} & Q_{3,4} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & Q_{m-1,m-2} & Q_{m-1,m-1} & Q_{m-1,m} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & Q_{m,m-1} & Q_{m,m} \end{pmatrix}$$

where each $Q_{i,i}$ is an $n_i \times n_i$ diagonal matrix, and the off-diagonal blocks $Q_{i,j}$ have entries which are either 0 or $q_{i,j}$ for some $q_{i,j} > 0$ and furthermore fulfill $\frac{1}{q_{i,i+1}} Q_{i,i+1} = \left(\frac{1}{q_{i+1,i}} Q_{i+1,i}\right)^*$, is similar to a diagonal matrix, .This is seen by defining the numbers d_i recursively as before and then let $\mathbf{d}$ be the vector where d_i is repeated n_i times. Finally, we notice that if we are given a lumpable CTMC where the rate matrix is of the form

$$Q = \begin{pmatrix} \tilde{Q} & \mathbf{c}^* \\ \mathbf{0} & 0 \end{pmatrix}$$

where $\tilde{Q}$ a tri-diagonal block matrix then the rate matrix of lumped process $\hat{Q}$ will be diagonalizable. To see this, assume Q is an $n \times n$ matrix and $\hat{Q}$ in an $m \times m$ matrix. By assumption, we may write $\hat{Q} = UQV$ where the j-th column vector of V is 1 in the entries corresponding to the jth-partition and 0 otherwise (see Kemeny and Snell

1960 together with the uniformization argument in Buchholz 1994). We note that Q is diagonalizable and hence the matrix C whose rows are the left eigenvectors of Q has rank n. Now, if we combine the observation in Barr and Thomas (1977) that if x is a left eigenvector of Q then xV is a left eigenvector of $\hat{Q}$ if xV is not the null vector with Sylvester's inequality for rank of the product of two matrices (inequality 0.4.5 (c) p. 13 in Horn and Johnson 1985), we can see that the matrix CV has rank m and hence there are m linearly independent vectors in the set $\{xV \mid x \text{ is a left eigenvector of } Q\}$ so $\hat{Q}$ is diagonalizable.

8.3 The lumped state-space

We define the state-space $\hat{\mathfrak{S}}$ to consist of states of the form $[(j_i, l_i) \mid i = 1, \ldots, m]$ where $j_i \in \{1, 2\}$ and $l_i \in \{1, 2\}$. Note that, as is implied by that notation $[\cdot]$, the states of $\hat{\mathfrak{S}}$ are *multisets* i.e. identical elements of each state are allowed to appear more than once, but, unlike vectors, the order of elements is not important. We use the notation $\#(i, j) \in \hat{\alpha}$ to mean the multiplicity of (i, j) and let the notation $\#(i, \cdot) \in \hat{\alpha}$ mean the number of elements in population i in the state $\hat{\alpha}$. We let A be a dummy absorbing state and define a rate matrix indexed by the states of $\hat{\mathfrak{S}} \cup A$:

$$\hat{Q}_{\hat{\alpha},\hat{\beta}} = \begin{cases} m_{1\to 2} \cdot \#(1, l) \in \hat{\alpha} & \text{if } \hat{\alpha} = \mathscr{S} \cup (1, l)\,,\ \hat{\beta} = \mathscr{S} \cup (2, l) \\ m_{2\to 1} \cdot \#(2, l) \in \hat{\alpha} & \text{if } \hat{\alpha} = \mathscr{S} \cup (2, l)\,,\ \hat{\beta} = \mathscr{S} \cup (1, l) \\ \binom{\#(1,\cdot)\in\hat{\alpha}}{2} c_1 + \binom{\#(2,\cdot)\in\hat{\alpha}}{2} c_2 & \text{if } \hat{\beta} = A \\ 0 & \text{otherwise}\,, \end{cases} \tag{32}$$

where the diagonal entries are defined by the requirement that $\hat{Q}$ is a rate matrix, i.e. the rows sum to 0.

Theorem 1 *Let $\alpha, \beta \in \mathfrak{S}$ with $F(\alpha) = F(\beta)$ and set $\bar{\mathfrak{S}} = \{\gamma \in \mathfrak{S} \mid F(\alpha) = F(\gamma)\} \cup A$. Let $\tilde{f}$ be any mapping from the set of lineages to $1, 2$ and let $f : \mathfrak{S} \cup A \to \hat{\mathfrak{S}} \cup A$ be the mapping induced by $\tilde{f}$, with $f(A) = A$. Then $Q_{|\bar{\mathfrak{S}}}$ is exactly lumpable with respect to the partition generated by f and the transition matrix of the lumped process is given by $\hat{Q}_{|f(\bar{\mathfrak{S}})}$*

Corollary 1 *Under the conditions from Theorem 1 and the additional assumption $|f^{-1}(f(\alpha))| = 1$ we have*

$$\mathbb{P}(X_t = \beta \mid X_0 = \alpha) = \frac{1}{|f^{-1}(f(\beta))|} \left(\mathrm{e}^{\hat{Q}_{|f(\bar{\mathfrak{S}})}}\right)_{f(\alpha), f(\beta)}$$

Proof First, we prove exact lumpability. Formally, the partition generated by f is the quotient set $\bar{\mathfrak{S}}/\!\sim_f$ where $\alpha \sim_f \beta \overset{def}{\Leftrightarrow} f(\alpha) = f(\beta)$. $\bar{\mathfrak{S}}/\!\sim_f$ is naturally identified with the image of $\bar{\mathfrak{S}}$ under f and since $f(\bar{\mathfrak{S}}) \subseteq \hat{\mathfrak{S}}$, we will label the elements of $\bar{\mathfrak{S}}/\!\sim_f$ with the elements of $\hat{\alpha} \in \hat{\mathfrak{S}} \cup A$. In particular, for $\hat{\alpha} \in \hat{\mathfrak{S}}$ we write "$\alpha \in \hat{\alpha}$" if $\alpha \in \{\gamma \mid f(\gamma) = \hat{\alpha}\}$. We need to check that for $\hat{\alpha}, \hat{\beta} \in \bar{\mathfrak{S}}/\!\sim_f \cup A$:

$$\beta, \beta' \in \hat{\beta} : \sum_{\alpha \in \hat{\alpha}} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\alpha,\beta} = \sum_{\alpha \in \hat{\alpha}} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\alpha,\beta'} \tag{33}$$

That is, the column sums must be equal for all partitions.

First, we note that (33) is fulfilled if either $\hat{\alpha} = A$ or $\hat{\beta} = A$ since in the former case all column sums are 0 (since A is absorbing) and in the latter there is only one column. Let $\hat{\alpha}, \hat{\beta} \in \tilde{\mathfrak{S}}/\sim_f$ be given. Write $\hat{\alpha} = (a_1, a_2)$, $\hat{\beta} = (b_1, b_2)$ where $a_1, a_2 \in \mathbb{N}_0$ are respectively the number of 1's and 2's in population 1 in $\hat{\alpha}$, and similarly b_1, b_2 are the number of 1's and 2's in population 1 in $\hat{\beta}$. Note that this notation unambiguously identifies $\hat{\alpha}$ since e.g. the number of 1's in population 2 in $\hat{\alpha}$ is $|\tilde{f}^{-1}(1)| - a_1$. We check that (33) holds in the three exhaustive cases $|a_1 - b_1| + |a_2 - b_2| = 0$, $|a_1 - b_1| + |a_2 - b_2| = 1$ and $|a_1 - b_1| + |a_2 - b_2| \geq 2$.

$$|a_1 - b_1| + |a_2 - b_2| = 0 :$$

This condition implies $\hat{\alpha} = \hat{\beta}$. We rewrite (33) by isolating the diagonal elements:

$$\left(Q_{|\tilde{\mathfrak{S}}} \right)_{\beta,\beta} + \sum_{\substack{\beta'' \in \hat{\beta} \\ \beta'' \neq \beta}} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\beta'',\beta} = \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\beta',\beta'} + \sum_{\substack{\beta'' \in \hat{\beta} \\ \beta'' \neq \beta'}} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\beta'',\beta'} \tag{34}$$

Note that f preserves number of elements in each population so for any $\gamma, \gamma' \in \mathfrak{S}$: $f(\gamma) = f(\gamma')$ implies that γ and γ' have the same number of lineages in population 1 and 2 respectively. This in turn implies

$$\gamma, \gamma' \in \hat{\beta} \overset{def}{\Leftrightarrow} f(\gamma) = f(\gamma') \Rightarrow \begin{cases} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\gamma,\gamma} = \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\gamma',\gamma'} & \text{and} \\ \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\gamma,\gamma'} = 0 \end{cases}$$

so that the sums in (34) are equal.

$$|a_1 - b_1| + |a_2 - b_2| = 1$$

This case covers the four sub-cases $a_1 - b_1 = 1$, $b_1 - a_1 = 1$, $a_2 - b_2 = 1$ and $b_2 - a_2 = 1$. We treat the first, as the rest are similar. The condition $a_1 - b_1 = 1$ means $\hat{\beta}$ is obtained from $\hat{\alpha}$ by moving a 1 from population 1 to population 2. Writing $\alpha = (A_1, A_2)$ where $A_1 \subseteq \tilde{f}^{-1}(1)$ and $A_2 \subseteq \tilde{f}^{-1}(2)$ are the lineages in population 1 and similarly for $\beta = (B_1, B_2)$, the assumed condition translates into $|A_1| - |B_1| = 1$, $|A_2| = |B_2|$ and we rewrite the sum $\sum_{\alpha \in \hat{\alpha}} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\alpha,\beta}$ in terms of the A_i's B_i's:

$$\sum_{\alpha \in \hat{\alpha}} \left(Q_{|\tilde{\mathfrak{S}}} \right)_{\alpha,\beta} = m_{1\to 2} \cdot \sum_{\substack{\alpha \in \hat{\alpha} \\ |A_1| - |B_1| = 1 \\ |A_2| = |B_2|}} I\left(\left(Q_{|\tilde{\mathfrak{S}}} \right)_{\alpha,\beta} > 0 \right) \tag{35}$$

The condition $|A_1| - |B_1| = 1$ implies that at least one lineage switches populations, and by the definition of Q given in (5) entries in $Q_{|\tilde{\mathfrak{S}}}$ will be 0 if they correspond to states where more than one lineage switches population. This will be the case if $A_2 \neq B_2$ or $B_1 \not\subseteq A_1$ Hence:

$$(35) = m_{1\to 2} \cdot \sum_{\substack{\alpha\in\hat{\alpha}\,,\,\alpha=(A_1,A_2)\\ B_1\subseteq A_1,|A_1|-|B_1|=1\\ A_2=B_2}} I\left(\left(Q_{|\tilde{\mathfrak{S}}}\right)_{\beta,\alpha} > 0\right) \tag{36}$$

Finally, observe that each summand in (36) is positive and that there are $|f^{-1}(1)|-|B_1|$ such summands, so that

$$(36) = (|\tilde{f}^{-1}(1)| - |B_1|)m_{1\to 2}. \tag{37}$$

By combining the equations above, we see that the right hand side of (35) depends on β through $|B_1|$ and by the remark that f preserves the number of elements in a given population this implies that the sum is unchanged if we replace β with β', which proves (33) in this case.

Finally, in the case $|a_1 - b_1| + |a_2 - b_2| \geq 2$ both sides (35) are 0 by the remark that f preserves the number of element in each population.

This proves exact lumpability. According to Proposition 7 of Baarir et al. (2011) the entries of the transition matrix of the lumped process are given by $\frac{|\hat{\beta}|}{|\hat{\alpha}|}\sum_{\alpha\in\hat{\alpha}}\left(Q_{|\tilde{\mathfrak{S}}}\right)_{\alpha,\beta}$. The number of elements of a partition $\hat{\alpha} = (a_1, a_2)$ is

$$|\hat{\alpha}| = \binom{|\tilde{f}^{-1}(1)|}{a_1}\binom{|\tilde{f}^{-1}(2)|}{a_2}$$

so in the case $a_1 - b_1 = 1$, $a_1 = b_2$ discussed above we have

$$\begin{aligned}
&\frac{|\hat{\beta}|}{|\hat{\alpha}|}\sum_{\alpha\in\hat{\alpha}}\left(Q_{|\tilde{\mathfrak{S}}}\right)_{\alpha,\beta}\\
&= \frac{|\hat{\beta}|}{|\hat{\alpha}|}(|\tilde{f}^{-1}(1)| - |B_1|)m_{1\to 2}\\
&= \frac{\binom{|\tilde{f}^{-1}(1)|}{b_1}\binom{|\tilde{f}^{-1}(2)|}{b_2}}{\binom{|\tilde{f}^{-1}(1)|}{a_1}\binom{|\tilde{f}^{-1}(2)|}{a_2}}(|\tilde{f}^{-1}(1)| - |B_1|)m_{1\to 2}\\
&= \frac{\binom{|\tilde{f}^{-1}(1)|}{a_1-1}}{\binom{|\tilde{f}^{-1}(1)|}{a_1}}(|\tilde{f}^{-1}(1)| - |B_1|)m_{1\to 2}\\
&= \frac{a_1}{|\tilde{f}^{-1}(1)| - (a_1-1)}(|\tilde{f}^{-1}(1)| - |B_1|)m_{1\to 2} = a_1 m_{1\to 2}
\end{aligned}$$

which is the first case in (32). The rest of the sub-cases of $|a_1 - b_1| + |a_2 - b_2| = 1$ are identical, and the remaining two cases ($|a_1 - b_1| + |a_2 - b_2| = 0$, $|a_1 - b_1| + |a_2 - b_2| \geq 2$) are trivial. □

Proof of corollary 1 To obtain the corollary we observe that exact lumpability implies that the CTMC remains equiprobable with respect to the partition generated by f if the starting distribution is equiprobable, and that the condition $|f^{-1}(f(\alpha))| = 1$ ensures that this is case for the starting distribution $\mathbb{P}(X_0 = \alpha) = 1$. Hence

$$\begin{aligned}\mathbb{P}(X_t = \beta \mid X_0 = \alpha) &= \frac{1}{|f^{-1}(f(\beta))|}\mathbb{P}(X_t \in f^{-1}(f(\beta)) \mid X_0 = \alpha) \\ &= \frac{1}{|f^{-1}(f(\beta))|}\left(\mathrm{e}^{\hat{Q}_{|f(\tilde{\mathcal{S}})}}\right)_{f(\alpha),f(\beta)}.\end{aligned}$$

□

8.4 The explicit integrals of Sect. 5

$$\begin{aligned}&\int_0^{T_A}\int_0^{t_3}\int_0^{t_2} f_1(t_1,t_2,t_3)h_1(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3 \\ &= \frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r)}}{(\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r)(-\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_s)(-\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r-\lambda_t)} \\ &\quad + \frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_k+\lambda_s)}}{(\lambda_i+\lambda_j+\lambda_k+\lambda_s)(\lambda_k-\lambda_l-\lambda_m-\lambda_r+\lambda_s)(-\lambda_i+\lambda_j+\lambda_k+\lambda_s-\lambda_t)} \\ &\quad - \frac{1}{(\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r)(\lambda_i+\lambda_j+\lambda_k+\lambda_s)(2\lambda_i+\lambda_t)} \\ &\quad - \frac{\mathrm{e}^{T_A(2\lambda_i+\lambda_t)}}{(-\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r-\lambda_t)(2\lambda_i+\lambda_t)(\lambda_i-\lambda_j-\lambda_k-\lambda_s+\lambda_t)} \\ &\int_{T_A}^{\infty}\int_0^{T_A}\int_0^{t_2} f_1(t_1,t_2,t_3)h_2(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3 \\ &= \frac{1}{\theta_a-2\lambda_i}\theta_a\left(\frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r)}}{(-\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_s)(-\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r-\lambda_t)}\right. \\ &\quad + \frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_k+\lambda_s)}}{(\lambda_k-\lambda_l-\lambda_m-\lambda_r+\lambda_s)(-\lambda_i+\lambda_j+\lambda_k+\lambda_s-\lambda_t)} \\ &\quad \left. - \frac{\mathrm{e}^{T_A(2\lambda_i+\lambda_t)}}{(-\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r-\lambda_t)(\lambda_i-\lambda_j-\lambda_k-\lambda_s+\lambda_t)}\right) \\ &\int_{T_A}^{\infty}\int_{T_A}^{t_3}\int_0^{T_A} f_1(t_1,t_2,t_3)h_3(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3 \\ &= \frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j)}(\mathrm{e}^{T_A(\lambda_l+\lambda_m+\lambda_r)} - \mathrm{e}^{T_A(\lambda_k+\lambda_s)}\theta_a^2}{(-\theta_a+2\lambda_i)(-3\theta_a+\lambda_i+\lambda_j+\lambda_k)(-\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_s)}\end{aligned}$$

$$\int_{T_A}^{\infty}\int_{T_A}^{t_3}\int_{T_A}^{t_2} f_1(t_1,t_2,t_3)h_4(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3$$
$$= \frac{-\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_l+\lambda_m+\lambda_r)}\theta_a^3}{(-\theta_a+2\lambda_i)(-3\theta_a+\lambda_i+\lambda_j+\lambda_k)(-6\theta_a+\lambda_i+\lambda_j+\lambda_l+\lambda_m)}$$

$$\int_{0}^{T_A}\int_{0}^{t_3}\int_{0}^{t_2} f_2(t_1,t_2,t_3)h_1(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3$$
$$= \frac{\mathrm{e}^{T_A(\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r)}}{(\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r)(-\lambda_i+\lambda_l+\lambda_m+\lambda_r-\lambda_s)(-2\lambda_i+\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_t)}$$
$$- \frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_k+\lambda_s)}}{(\lambda_i+\lambda_j+\lambda_k+\lambda_s)(\lambda_i-\lambda_l-\lambda_m-\lambda_r+\lambda_s)(\lambda_i-\lambda_j-\lambda_k-\lambda_s+\lambda_t)}$$
$$- \frac{1}{(\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r)(\lambda_i+\lambda_j+\lambda_k+\lambda_s)(2\lambda_i+\lambda_t)}$$
$$- \frac{\mathrm{e}^{T_A(2\lambda_i+\lambda_t)}}{(-2\lambda_i+\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_t)(2\lambda_i+\lambda_t)(\lambda_i-\lambda_j-\lambda_k-\lambda_s+\lambda_t)}$$

$$\int_{T_A}^{\infty}\int_{0}^{T_A}\int_{0}^{t_2} f_2(t_1,t_2,t_3)h_2(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3$$
$$= \frac{\theta_a}{(\theta_a-2\lambda_i)(\lambda_i-\lambda_l-\lambda_m-\lambda_r+\lambda_s)}\Bigg(-\frac{\mathrm{e}^{T_A(\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r)}}{-2\lambda_i+\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_t}$$
$$- \frac{\mathrm{e}^{T_A(\lambda_i+\lambda_j+\lambda_k+\lambda_s)}}{\lambda_i-\lambda_j-\lambda_k-\lambda_s+\lambda_t} + \frac{\mathrm{e}^{T_A(2\lambda_i+\lambda_t)}}{-2\lambda_i+\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r-\lambda_t}$$
$$+ \frac{\mathrm{e}^{T_A(2\lambda_i+\lambda_t)}}{\lambda_i-\lambda_j-\lambda_k-\lambda_s+\lambda_t}\Bigg)$$

$$\int_{T_A}^{\infty}\int_{T_A}^{t_3}\int_{0}^{T_A} f_2(t_1,t_2,t_3)h_3(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3$$
$$= \frac{\mathrm{e}^{(T_A(\lambda_j+\lambda_k))}(\mathrm{e}^{T_A(\lambda_l+\lambda_m+\lambda_r)}-\mathrm{e}^{T_A(\lambda_i+\lambda_s)})\theta_a^2}{(-\theta_a+2\lambda_i)(-3\theta_a+\lambda_i+\lambda_j+\lambda_k)(-\lambda_i+\lambda_l+\lambda_m+\lambda_r-\lambda_s)}$$

$$\int_{T_A}^{\infty}\int_{T_A}^{t_3}\int_{T_A}^{t_2} f_2(t_1,t_2,t_3)h_4(t_1,t_2,t_3)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3$$
$$= \frac{\mathrm{e}^{T_A(\lambda_j+\lambda_k+\lambda_l+\lambda_m+\lambda_r)}\theta_a^3}{(\theta_a-2\lambda_i)(-3\theta_a+\lambda_i+\lambda_j+\lambda_k)(-6\theta_a+\lambda_j+\lambda_k+\lambda_l+\lambda_m)}$$

References

Abramowitz M, Stegun IA (eds) (1984) Handbook of mathematical functions with formulas, graphs, and mathematical tables. A Wiley-Interscience Publication, Wiley, New York, reprint of the 1972 edition, Selected Government Publications

Asmussen S (2003) Applied Probability and Queues, Applications of Mathematics (New York), stochastic Modelling and Applied Probability. vol 51, 2nd edn. Springer, New York

Baarir S, Beccuti M, Dutheillet C, Franceschinis G, Haddad S (2011) Lumping partially symmetrical stochastic models. Perform Eval 68(1):21–44. doi:10.1016/j.peva.2010.09.002

Barr DR, Thomas MU (1977) An eigenvector condition for Markov chain lumpability. Operat Res 25(6):1028–1031, http://www.jstor.org/stable/169878

Buchholz P (1994) Exact and ordinary lumpability in finite Markov chains. J Appl Probab 31(1):59–75

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of drosophila pseudoobscura and d. persimilis. Genetics 167(2):747–760, doi:10.1534/genetics.103.024182, http://www.genetics.org/content/167/2/747.abstract, http://www.genetics.org/content/167/2/747.full.pdf+html

Hobolth A, Andersen LN, Mailund T (2011) On computing the coalescence time density in an isolation-with-migration model with few samples. Genetics 187(4):1241–1243, doi:10.1534/genetics.110.124164, http://www.genetics.org/content/187/4/1241.short, http://www.genetics.org/content/187/4/1241.full.pdf+html

Horn RA, Johnson CR (1985) Matrix analysis. Cambridge University Press, Cambridge

Innan H, Watanabe H (2006) The effect of gene flow on the coalescent time in the human-chimpanzee ancestral population. Molecular Biology and Evolution 23(5):1040–1047, doi:10.1093/molbev/msj109, http://mbe.oxfordjournals.org/content/23/5/1040.abstract, http://mbe.oxfordjournals.org/content/23/5/1040.full.pdf+html

Keilson J (1979) Markov Chain Models-Rarity and Exponentiality. No. vb. 28 in Applied Mathematical Sciences, Springer. http://books.google.dk/books?id=X6SjQgAACAAJ

Kemeny JG, Snell JL (1960) Finite Markov chains. The University Series in Undergraduate Mathematics, D. Van Nostrand Co., Inc., Princeton

Kingman J (1982) The coalescent. Stochastic Processes and their Applications 13(3):235–248. doi:10.1016/0304-4149(82)90011-4, http://www.sciencedirect.com/science/article/pii/0304414982900114

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain monte carlo approach. Genetics 158(2):885–896. http://www.genetics.org/content/158/2/885.abstract, http://www.genetics.org/content/158/2/885.full.pdf+html

Rosenberg NA (2007) Counting coalescent histories. J Comput Biol 14(3):360–377

Scally A et al (2012) Insights into hominid evolution from the gorilla genome sequence. Nature 483(7388):169–175. doi:10.1038/nature10842

Stanley RP (2012) Enumerative combinatorics. vol 1, Cambridge Studies in Advanced Mathematics, vol 49, 2nd edn. Cambridge University Press, Cambridge

Takahata N, Satta Y, Klein J (1995) Divergence time and population size in the lineage leading to modern humans. Theoretical Population Biology 48(2):198–221. doi:10.1006/tpbi.1995.1026, http://www.sciencedirect.com/science/article/pii/S004058098571026X

Tian J, Lin XS (2009) The mutation process in colored coalescent theory. Bull Math Biol 71:1873–1889. doi:10.1007/s11538-009-9428-4

Wang Y, Hey J (2010) Estimating divergence parameters with small samples from a large number of loci. Genetics 184(2):363–379. doi:10.1534/genetics.109.110528, http://www.genetics.org/content/184/2/363.abstract, http://www.genetics.org/content/184/2/363.full.pdf+html

Wilkinson-Herbots HM (2008) The distribution of the coalescence time and the number of pairwise nucleotide differences in the "isolation with migration" model. Theor Popul Biol 73(2):277–288. doi:10.1016/j.tpb.2007.11.001

Wilkinson-Herbots HM (2012) The distribution of the coalescence time and the number of pairwise nucleotide differences in a model of population divergence or speciation with an initial period of gene flow. Theore Popul Biol 82(2):92–108. doi:10.1016/j.tpb.2012.05.003, http://www.sciencedirect.com/science/article/pii/S0040580912000524

Lohse K, Harrison RJ, Barton NH (2011) A general method for calculating likelihoods under the coalescent process. Genetics 189(3):977–987. doi:10.1534/genetics.111.129569, http://www.genetics.org/content/189/3/977.abstract, http://www.genetics.org/content/189/3/977.full.pdf+html

Zhu T, Yang Z (2012) Maximum likelihood implementation of an isolation-with-migration model with three species for testing speciation with gene flow. Mol Biol Evolut. doi:10.1093/molbev/mss118, http://mbe.oxfordjournals.org/content/early/2012/04/13/molbev.mss118.abstract, http://mbe.oxfordjournals.org/content/early/2012/04/13/molbev.mss118.full.pdf+html