A model for isolation with migration avoiding explicit migration events

Graham Jones

2022-10-31, March 9, 2023

art@gjones.name, www.indriid.com

THIS IS PRELIMINARY

1 Introduction

This note describes a model for isolation and migration. In DENIM and IMa3 (Hey et al. (2018)), at every time t, every branch of every gene tree existing at t is assigned to a branch in the species tree, and migration events (times, source and destination branches) are parameters in the model which are sampled during the MCMC sampling. The model here assigns gene tree nodes (coalescences) to species tree branches, and uses an approximation to integrate out the migration events analytically.

My current name for the model is LUCAS = Lineages Unassigned and Coalescences Assigned to Species

1.1 Notation

(This is mainly for organizing names; full definitions in text.)

- G gene tree topology and node (coalescence) times
- Λ parameters assigning coalescences to species tree branches
- never name the species tree?
- m migration rate.
- θ_b population parameter in branch b
- t time measured backwards
- b, c, d indices for species tree branches
- s number of current branches
- $\mathcal{L}_b(t)$ lineages in branch b at time t.
- $\mathcal{L}(t)$ all lineages at time t.
- i, j, k indices for lineages (=gene tree branches)

- $\breve{A}_i$ event that lineage i goes from start branch and time to end branch and time
- $\hat{t}(i)$, $\breve{t}(i)$ start and end time of lineage i.
- $\hat{b}(i)$, $\breve{b}(i)$ start and end branch of lineage i.
- $\widehat{P}_{bi}(t)$ prob that i is in b at t.
- α average coalescence rate.
- Φ a stochastic rate matrix for states coalesced, together, apart.
- $T_{ij}(t)$ event that $\mathsf{path}(i)$ and $\mathsf{path}(j)$ do not coalesce before t
- $T_{ij}(k,t)$ event that neither $\mathsf{path}(i)$ and $\mathsf{path}(k)$ nor $\mathsf{path}(j)$ and $\mathsf{path}(k)$ coalesce before t
- N_i number of migrations in lineage i.
- $\mathsf{path}(i)$ path rootwards from lineage i.
- $\mathsf{tmrca}(i,j)$ the most recent common ancestor of i and j.
- $\mathsf{allS}(i)$ part of species tree that co-exists with i.
- $\mathsf{ancS}(i)$ i and lineages ancestral to i.
- $\mathsf{descS}(i)$ i and descendant lineages of i.
- lenS branch length of part of species tree.
- $\mathrm{uexp}(x)$ function $(1-\exp(-x))/x$

2 Calculation of gene tree prior

2.1 Outline

Time t is measured backwards from present which is at $t=0$. We focus on a single gene tree G. At $t=0$, the gene tree tips are assigned to species tree tips. At each coalescence in G, a parameter (to be estimated) assigns the coalescence to one of the contemporaneous species tree branches. Call the collection of these parameters Λ. Thus, given Λ and the assignments at $t=0$, the probability that a lineage i is in a species tree branch b is known as 0 or 1 at the start and end of every lineage. We denote the start time of i as $\hat{t}(i)$, the end time as $\breve{t}(i)$, the start branch as $\hat{b}(i)$ and the end branch as $\breve{b}(i)$.

The migration during intervals between speciations is modeled by an $s \times s$ rate matrix M, where s is the number of species during the interval, and where M_{bd} is the rate at which a lineage (migrates from species tree branch b to species tree branch d. Migration is regarded as going backwards in time, so this is the rate from b at smaller t to d at larger t.

We decompose the gene tree into coalescences where each coalescence 'owns' the two child lineages. For a coalescence between lineages i and j at time $t = \breve{t}(i) = \breve{t}(j)$ in branch $b = \breve{b}(i) = \breve{b}(j)$, we find the probability that i and j do not coalesce before t, and that both are in branch b at time t. Then we deal with other lineages k that exist at time t, and find the probability that they do not coalesce with i or j, given that both are in branch b at time t. For each pair of lineages considered, we split into several cases based on the number of migrations: there are some cases for very few migrations, and one for more, where an approximation is used. Finally we obtain a density for the coalescence time, by multiplying by θ_b^{-1} where θ_b is the usual population size parameter for branch b (long-term effective population times ploidy times mutation rate).

Let $\mathcal{L}(t)$ be the set of all lineages in G existing at t, and $\mathcal{L}_b(t)$ be the set of lineages in branch b at time t. Let $\mathsf{path}(i)$ be the path in G from the start of i to the root. Let $\mathsf{tmrca}(i, j)$ be the time of the most recent common ancestor of i and j, where $\mathsf{path}(i)$ and $\mathsf{path}(j)$ coalesce.

2.2 Comparisons

Comparison with Palczewski and Beerli (2013). There are two sources of inaccuracy in model of Palczewski and Beerli (2013). The first is the lack of independence between probabilities that different lineages are in a branch b at some time t. Consider two species branches b and c with equal populations sizes, and equal migration rates m each way between them. Assume b and c do not merge for a very long time, and suppose that i is assigned to b and j to c at $t = 0$ and the first coalescence is between i and j. Starting from $t = 0$ the lineages behave independently, but once i and j have coalesced to form k, it is only known that $\Pr(k \in \mathcal{L}_b(t)) = \Pr(k \in \mathcal{L}_c(t)) = 1/2$ and the coalescent intensity between k and any other lineage l is $1/(2\theta)$, and expected time to coalescence equal to 2θ. The true situation is that k and l are either together, with initial intensity $1/\theta$ and the model gives an expected time to coalescence larger than θ, or they are apart, with initial intensity 0, and expected time to coalescence larger than $1/(2m)$ since a migration must happen before the coalescence. Overall the expected time to coalescence is larger than $\theta/2 + 1/(4m)$, and for $m \ll 1/\theta$, this may be much larger than 2θ.

The second problem is that even when the lineages behave independently, as they do until the first coalescence, the method overestimates the coalescent intensity. The problem is especially bad when $m \ll 1/\theta$, for example $m = 1, 1/\theta = 10000$. With i, j, b, and c as above, after a time 0.005, the probability that a migration has happened is about 0.01, and the coalescent intensity in the model is about 100 (and growing), producing an expected time to coalescence of less than 0.015. The true value is over 0.5.

The quality of the approximation is discussed in more detail in Palczewski and Beerli (2013). The model is suited to 'populations' with considerable migration between them, but not suitable for 'species' with small rates of migration. The method here resolves the independence problem by introducing the parameters Λ. This gives a 'fresh start' after each coalescence. The approximation is then improved by considering the cases of 0 or 1 migrations within each pair of coalescing lineages separately, and only using the approximation of Palczewski and Beerli (2013) for the case of at least 2 migrations.

Comparison with Hey et al. (2018) (IMa3). This models migrations explicitly. Exact, but presumably slow with lots of migrations. Allows population size parameters to be integrated out analytically, which is not possible in model proposed here. TODO.

2.3 Decomposition of G

Suppose that i and j are lineages and that for each of them, the time and branch of their start is known. Let be the $T_{ij}(t)$ event that $\mathsf{path}(i)$ and $\mathsf{path}(j)$ do not coalesce before t. Let $T_{ij}(k, t)$ be the event that neither $\mathsf{path}(i)$ and $\mathsf{path}(k)$ nor $\mathsf{path}(j)$ and $\mathsf{path}(k)$ coalesce before t. Denote by $\breve{A}_i$ the event that $\mathsf{path}(i) \in \mathcal{L}_{\breve{b}(i)}(\breve{t}(i))$, that is, that i reaches the right branch at the right time to coalesce, and similarly for $\breve{A}_j$.

First we find the probability of the events $T_{ij}(t)$, $\breve{A}_i$, and $\breve{A}_j$. Note that this does not depend on whether any other lineages coalesce with $\mathsf{path}(i)$ or $\mathsf{path}(j)$ before $\breve{t}(i)$. Secondly we find the probability of the event $T_{ij}(k, t)$, given the coalescence of $\mathsf{path}(i)$ and $\mathsf{path}(j)$. We call these the 'coalescers' and the 'persisters', respectively.

The density for $\breve{t}(i)$ is then found by multiplying by θ_b^{-1}. We can then 'forget' about i and j and continue with the rest of (G, Λ). Note that lineages in $\mathcal{L}(t) \setminus \{i, j\}$ may start at any time in $[0, t)$ and the probability

that they do not coalesce amongst themselves before $\breve{t}(i)$ is calculated when dealing with later coalescences.

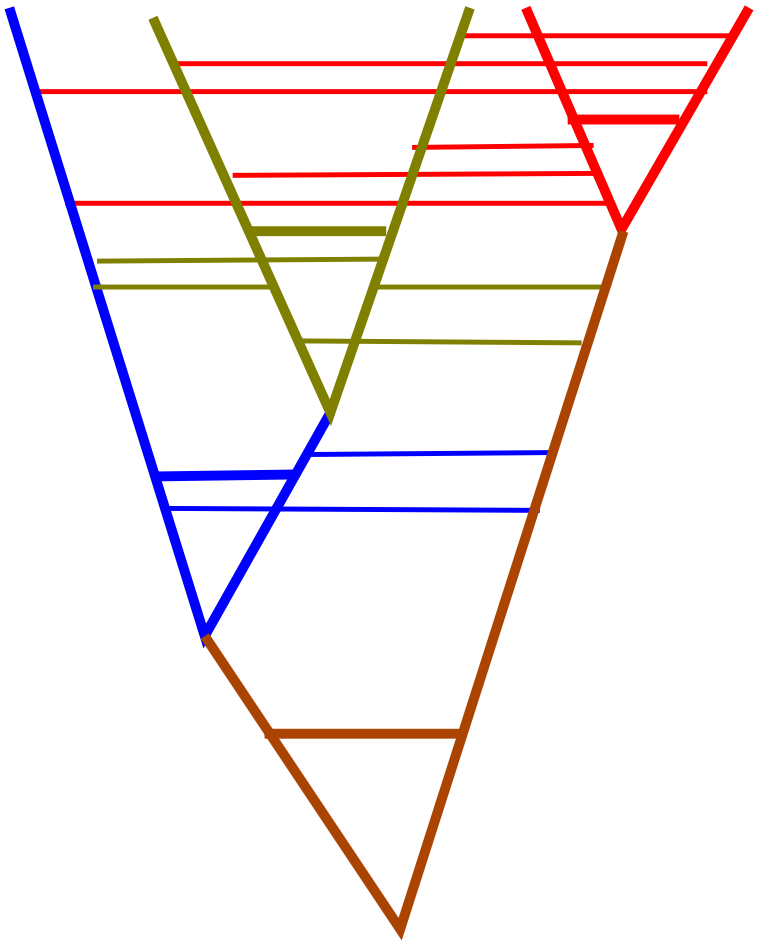

Figure 1: Decomposition of G into 4 coalescences in different colours. At each coalescence, the horizontal lines indicate the pairs of lineages between which potential coalescences are considered.

We focus on one coalescence and omit the dependence on assignments of previous coalescences to branches. Let N_x be the number of migrations that a lineage x contains before t (this is all migrations for i and j, but not the ones after t for k). For the 'coalescers',

$$\begin{aligned}
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j)\right) = \\
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (0,0)\right) + \\
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (0,1)\right) + \\
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (1,0)\right) + \\
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge \mathsf{N}_i + \mathsf{N}_j \geq 2\right).
\end{aligned} \tag{1}$$

Using the fact that i and j are independent,

$$\begin{aligned}
&\Pr\Big(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j)\Big) = \\
&\Pr\Big(T_{ij}(t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (0,0)\Big) \Pr(\breve{A}_i \wedge \mathsf{N}_i = 0) \Pr(\breve{A}_j \wedge \mathsf{N}_j = 0) + \\
&\Pr\Big(T_{ij}(t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (0,1)\Big) \Pr(\breve{A}_i \wedge \mathsf{N}_i = 0) \Pr(\breve{A}_j \wedge \mathsf{N}_j = 1) + \\
&\Pr\Big(T_{ij}(t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (1,0)\Big) \Pr(\breve{A}_i \wedge \mathsf{N}_i = 1) \Pr(\breve{A}_j \wedge \mathsf{N}_j = 0) + \\
&\Pr\Big(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge \mathsf{N}_i + \mathsf{N}_j \geq 2\Big).
\end{aligned} \tag{2}$$

For the 'persisters', we want

$$\Pr\Big(T_{ij}(k,t) \mid \breve{A}_i \wedge \breve{A}_j\Big) \;=\; \Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j\Big) \;/\; \Pr(\breve{A}_i \wedge \breve{A}_j) \tag{3}$$

and the numerator on the right hand side can be expanded as

$$\begin{aligned}
&\Pr\Big(T_{ij}(k,t)) \wedge \breve{A}_i \wedge \breve{A}_j\Big) \;= \\
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,0,0)\Big) + \\
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,0,1)\Big) + \\
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,1,0)\Big) + \\
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (1,0,0)\Big) + \\
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (1,1,0)\Big) + \\
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)\Big)
\end{aligned} \tag{4}$$

The first five terms can be expressed like

$$\begin{aligned}
&\Pr\Big(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (n_i, n_j, n_k)\Big) \;= \\
&\qquad \Pr\Big(T_{ij}(k,t) \;\Big|\; \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (n_i, n_j, n_k)\Big) \;\times \\
&\qquad \Pr\Big(\breve{A}_i \;\wedge \mathsf{N}_i = n_i\Big) \;\Pr\Big(\breve{A}_j \wedge \mathsf{N}_j = n_j\Big) \;\Pr\left(\mathsf{N}_k = n_k\right)
\end{aligned} \tag{5}$$

This is as far as we take the general case. Next we look for an approximation in the case of a simple migration matrix.

3 Special case for M

From now on, we restrict the case where the migration rate is the same between any pair of contemporaneous species tree branches, so $M_{bd} = m$ for every $b \neq d$, and $M_{bb} = -(s-1)m$. Let I be the $s \times s$ identity matrix and U be an $s \times s$ matrix filled with $1/s$. Then $M = smU - smI$ and it is straightforward to show that for any real number x, we have $\exp(Mx) \;=\; U + (I - U)e^{-smx}$.

3.1 Probability that a lineage is in a given branch at a given time

Consider a single lineage i in G. Suppose $t \in [\hat{t}(i), \breve{t}(i)]$ and that b is a branch existing at t. Let $\widehat{P}_{bi}(t) = \Pr(\mathsf{path}(i) \in \mathcal{L}_b(t) | i \in \mathcal{L}_{\hat{b}(i)}(\hat{t}(i)))$. Suppose $\hat{t}(i) \leq v \leq t \leq \breve{t}(i)$, and that no speciations occur during $[v, t]$. Then the behaviour of i is determined by the migration matrix M during $[v, t]$ so

$$\Pr(i \in \mathcal{L}_b(t) | i \in \mathcal{L}_c(v)) = [\exp(M(t-v))]_{bc}. \tag{6}$$

If species tree branches b and c merge to form branch d at time t, we have

$$\widehat{P}_{di}(t) = \widehat{P}_{bi}(t) + \widehat{P}_{ci}(t). \tag{7}$$

Using these equations we can calculate $\widehat{P}_{bi}(t)$ at any time $t \in [\hat{t}(i), \breve{t}(i)]$, starting with the assignment $\hat{b}(i)$ at $\hat{t}(i)$.

Now we focus on the special case for M. Let $[u, v]$ be an interval between speciations, and let s be the number of branches during this interval. We can assume that $\widehat{P}_{ci}(u)$ is known for all c. Using equation (6) in the special case, we have

$$\widehat{P}_{bi}(t) \;=\; s^{-1}(1 - \exp(-sm(t-u))) + \exp(-sm(t-u))\widehat{P}_{bi}(u). \tag{8}$$

Thus we can find $\widehat{P}_{bi}(\breve{t}(i))$ if i ends during $[u, v]$, and $\widehat{P}_{bi}(v)$ if not, ready for the next interval. Finally $\Pr(\breve{A}_i) = \widehat{P}_{\breve{b}(i)i}(\breve{t}(i))$ and $\Pr(\breve{A}_i) = \widehat{P}_{\breve{b}(j)j}(\breve{t}(j))$.

3.2 Probability of migration counts for 'coalescers' case

If it is known that a single migration of a lineage i occurs, and that the destination of the migration is in some part of the species tree, then this destination is uniformly distributed over all the species branches in the part. This observation forms the basis of the calculations in this section.

Let $\mathsf{allS}(i)$ be the part of the species tree that exists between $\hat{t}(i)$ and $\breve{t}(i)$. Let $\mathsf{ancS}(i)$ be the part of the species tree that is ancestral to $\hat{b}(i)$ at $\hat{t}(i)$, between $\hat{t}(i)$ and $\breve{t}(i)$. Let $\mathsf{descS}(i)$ be the part of the species tree that is descendant to $\breve{b}(i)$ at $\breve{t}(i)$, between $\hat{t}(i)$ and $\breve{t}(i)$. Thus $\mathsf{ancS}(i)$ the part that $\mathsf{path}(i)$ can reach without migrating, and $\mathsf{descS}(i)$ is the part of the species tree from which $\breve{b}(i)$ at $\breve{t}(i)$ can be reached without migration. Either $\mathsf{ancS}(i)$ and $\mathsf{descS}(i)$ are disjoint, or $\mathsf{ancS}(i)$ is contained in $\mathsf{descS}(i)$. Let $\mathsf{lenS}(X)$ denote the total branch length of a part X of the species tree.

Suppose that $t \in [\hat{t}(i), \breve{t}(i)]$. The migration intensity at time t is $= (|\mathcal{B}(t)| - 1)m$ which is a step function which changes at speciation times. The total migration intensity $F_{mig}(i)$ can be written as an integral of this function over $[\hat{t}(i), \breve{t}(i)]$. The result is equal to the product of m and the total branch length between $\hat{t}(i)$ and $\breve{t}(i)$, minus $(\breve{t}(i) - \hat{t}(i))$. Alternatively,

$$F_{mig}(i) = m\ \mathsf{lenS}(\mathsf{allS}(i) \setminus \mathsf{ancS}(i)) \tag{9}$$

The distribution of counts follows a Poisson distribution, so we have

$$\Pr(\mathsf{N}_i = 0) = \exp\left[-F_{mig}(i)\right] \tag{10}$$

$$\Pr(\mathsf{N}_i = 1) = F_{mig}(i) \exp\left[-F_{mig}(i)\right] \tag{11}$$

with similar expressions for j. Then $\Pr(\mathsf{N}_i + \mathsf{N}_j \geq 2)$ can be found by subtracting the three other probabilities from from 1.

3.3 Probability of migration counts for 'persisters' case

Let $l \in \{i, j\}$. We have $\Pr(\mathsf{N}_l = 0) = \exp\left[-F_{mig}(l)\right]$ as the last sub-section. If $\mathsf{ancS}(l)$ and $\mathsf{descS}(l)$ are disjoint, $\Pr(\breve{A}_l | \mathsf{N}_l = 0) = 0$. If $\mathsf{ancS}(l)$ is contained in $\mathsf{descS}(l)$, then $\Pr(\breve{A}_l | \mathsf{N}_l = 0) = 1$.

Likewise, we can write $\Pr(\mathsf{N}_l = 1 \wedge \breve{A}_l)$ as $\Pr(\breve{A}_l | \mathsf{N}_l = 1) \Pr(\mathsf{N}_l = 1)$. We have $\Pr(\mathsf{N}_l = 1) = F_{mig}\left(\hat{t}(l), \breve{t}(l)\right) \exp\left[-F_{mig}\left(\hat{t}(l), \breve{t}(l)\right)\right]$ as in the last subsection. Then

$$\Pr(\breve{A}_l | \mathsf{N}_l = 1) = \frac{\mathsf{lenS}(\mathsf{descS}(l) \setminus \mathsf{ancS}(l))}{\mathsf{lenS}(\mathsf{allS}(l) \setminus \mathsf{ancS}(l))} \tag{12}$$

The values for $\Pr(\mathsf{N}_k = 0)$ and $\Pr(\mathsf{N}_k = 1)$ as calculated as in the last sub-section, since these do not depend on what i or j do. Finally $\Pr(\mathsf{N}_l + \mathsf{N}_k \geq 2 \wedge \breve{A}_l)$ can be found by subtracting the three other probabilities from from $\Pr(\breve{A}_l)$.

3.4 The probability of no coalescences: 0 or 1 migration

The 'coalescers' case with $\mathsf{N}_i = 0 \wedge \mathsf{N}_j = 0$, and the 'persisters' case with $\mathsf{N}_i = 0 \wedge \mathsf{N}_k = 0$ are straightforward, since the location of $\mathsf{path}(i)$, $\mathsf{path}(j)$, and $\mathsf{path}(k)$ is known at all t. During the times that they are together in a branch c, the coalescent intensity is θ_c^{-1}. (It's like a bit of a standard multispecies coalescent calculation.)

Suppose that exactly one migration of a lineage x occurs during some interval of length w during which a branch c exists, that lineage y is in c during the interval, and that the migration of x is either from or to c. For convenience we define the function $\mathrm{uexp}(x) = (1 - \exp(-x))/x$. Then the probability that there is no coalescence of x and y during the interval is

$$\frac{1 - \exp(-w\theta_c^{-1})}{w\theta_c^{-1}} = \mathrm{uexp}(w\theta_c^{-1}). \tag{13}$$

This can be shown by using the fact that the migration time is uniformly distributed within the interval, and integrating it out. The result is the same whether the time that x and y spend together is at the start or end of the interval.

The method for a single migration is to split the species tree into intervals, between $\hat{t}(i)$, $\hat{t}(j)$, or $\hat{t}(k)$ as appropriate at the start, and $\breve{t}(i) = \breve{t}(j)$ at the end, and also at any speciation times within the range. Within each interval, we look at each branch segment available for the destination of the migration. For a 'persister' k, this is restricted to $\mathsf{descS}(i) \cup \mathsf{descS}(j)$. This allows us to calculate the probability that the migration went to each branch segment. Then, conditioning on each branch segment, we can find the probability of no coalescence before the segment and the probability of no coalescence after the segment (since we know where both lineages are before and after the segment), and use equation (13) for the segment itself.

In the 'persisters' case, with $(\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (1, 1, 0)$ it is necessary to consider two segments, one for each migration. I've only worked out the spacial case of a fixed number of species, in section 4. TODO

3.5 Approximate Markov model for at least 2 migrations

We use a three state Markov process. The three states are'coalesced' (1) , 'together' (2) (both lineages in same branch, but not coalesced) and 'apart' (3). The approximation does not account for different population sizes in different branches, but instead uses an overall average. The idea is that the lineages

could be 'anywhere' when there are a lot of migrations (although 'at least two' is not really a lot.) During an interval when there are s branches, the mean coalescence rate can be approximated by $\alpha = s^{-1}(\sum_b \theta_b^{-1})$. [I think this is better than taking the reciprocal of the θ_b's.] The process determined by the following stochastic rate matrix for two paths such as $\mathsf{path}(i)$ and $\mathsf{path}(j)$.

$$\Phi = \begin{bmatrix} 0 & 0 & 0 \\ \alpha & -(\alpha + 2(s-1)m) & 2(s-1)m \\ 0 & 2m & -2m \end{bmatrix}. \tag{14}$$

Given the probabilities that $\mathsf{path}(i)$ and $\mathsf{path}(j)$ are together or apart at the start of a period of duration u, then the probability that they are together or apart at the end of the period can be found from $\exp(\Phi u)$. For example $\exp(\Phi t)_{32}$ is the probability that $\mathsf{path}(i)$ and $\mathsf{path}(j)$ are together at time t, given they were apart at time 0, and $\exp(\Phi t)_{32} + \exp(\Phi t)_{33}$ is the probability that they have not coalesced by time t. Let $Y_{ij}(t)$ be the state (1,2, or 3) of lineages i and j at time t.

Suppose $\hat{t}(i) \le \hat{t}(j)$, that is, that i starts first. During the interval $[\hat{t}(i), \hat{t}(j)]$, we can find the probability that i arrives in $\hat{b}(j)$ by $\hat{t}(j)$, using $\hat{P}_{bi}(t)$ from section 3.1.

At a speciation at t, where s branches become $s-1$, the value of $\Pr(Y_{ij}(t) = 2)$ just after the speciation is found from the value of $\Pr(Y_{ij}(t) = 2) + (s(s-1)/2)^{-1} \Pr(Y_{ij}(t) = 3)$ just before the speciation. Likewise the value $\Pr(Y_{ij}(t) = 3)$ just after the speciation is found from the value of $\Pr(Y_{ij}(t) = 2) - (s(s-1)/2)^{-1} \Pr(Y_{ij}(t) = 3)$ just before.

Φ can be exponentiated analytically. This code finds elements of $\exp(\Phi t)$ in two ways

```
library(expm)
s = 5
a = 1
m = 0.05
t = 1
Q = matrix(c(0, a, 0, 0, -(a+2*(s-1)*m), 2*m, 0, 2*(s-1)*m, -2*m), nrow=3, ncol=3)
E = expm(Q*t)
print(E)

x = m * s + a/2
y = 0.5 * sqrt(4*x^2 - 8 *a*m)
Ext = exp(-x*t)
shyt = sinh(y*t)
chyt = cosh(y*t)
Q32 = (2 * m / y) * Ext * (shyt)
Q33 = (1/y) * Ext * ((x - 2*m) *shyt + y * chyt)
Q22 = (1/y) * Ext * ((2*m - x)*shyt + y * chyt)
Q23 = (s-1) * Q32
print(matrix(c(Q22,Q32,Q23,Q33), nrow=2, ncol=2))
```

3.6 The probability of no coalescences: at least 2 migrations

For the 'coalescers' case, we use Φ to find the probability that i and j are together but not coalesced by time t, and multiply by

$$\frac{\widehat{P}_{bi}(t)\widehat{P}_{bj}(t)}{\sum_c \widehat{P}_{ci}(t)\widehat{P}_{cj}(t)} \tag{15}$$

to find the probability that i and j are in b, given that they are together, so

$$\begin{aligned}
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge \mathsf{N}_i + \mathsf{N}_j \geq 2\right) \; = \\
&\Pr\left(T_{ij}(t) \wedge \breve{A}_i \wedge \breve{A}_j \mid \mathsf{N}_i + \mathsf{N}_j \geq 2\right) \; \Pr(\mathsf{N}_i + \mathsf{N}_j \geq 2) \; \simeq \\
&\frac{\widehat{P}_{bi}(t)\widehat{P}_{bj}(t)}{\sum_c \widehat{P}_{ci}(t)\widehat{P}_{cj}(t)} \Pr(Y_{ij}(t) = 2) \; \Pr(\mathsf{N}_i + \mathsf{N}_j \geq 2)
\end{aligned} \tag{16}$$

For the 'persisters' case, we assume approximate independence of $(\breve{A}_i \wedge \breve{A}_j)$ and the other events and estimate

$$\begin{aligned}
&\Pr\left(T_{ij}(k,t) \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)\right) \; \simeq \\
&\Pr\left(T_{ij}(k,t) \wedge (\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)\right) \; \Pr(\breve{A}_i \wedge \breve{A}_j) = \\
&\Pr\left(T_{ij}(k,t) \mid (\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)\right) \; \Pr(\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2) \; \Pr(\breve{A}_i \wedge \breve{A}_j)
\end{aligned} \tag{17}$$

and then assume approximate independence of $(\mathsf{tmrca}(i,k) \geq t)$ and $(\mathsf{tmrca}(j,k) \geq t)$ given the condition on counts, and use Φ to estimate these, so that

$$\begin{aligned}
&\Pr\left(T_{ij}(k,t) \mid (\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)\right) \; \simeq \\
&(\Pr(Y_{ik}(t) = 2) + \Pr(Y_{ik}(t) = 3)) \; (\Pr(Y_{jk}(t) = 2) + \Pr(Y_{jk}(t) = 3))
\end{aligned} \tag{18}$$

4 Details for fixed number of species

Suppose there are s species at all times - a simple island model, no tree. I use $\mathbb{I}[]$ as the indicator function, equal to one if its argument is true, else zero. We find values for equation (2) then equation (4). Let $b = \breve{b}(i) = \breve{b}(j)$, and $t = \breve{t}(i) = \breve{t}(j)$.

4.1 'coalescers'

Let $u_i = \max(\hat{t}(j) - \hat{t}(i), 0)$. Let $u_j = \max(\hat{t}(i) - \hat{t}(j), 0)$. Let $w = \max(\hat{t}(i), \hat{t}(j)) - t$. Then u_i is the time i spends alone, u_j is the time that j spends alone, and w is the time they spend together.

4.1.1 Migration counts

$$\begin{aligned}
\Pr(\mathsf{N}_i = 0) \; &= \; \exp(-(s-1)m(u_i + w)) \\
\Pr(\mathsf{N}_i = 1) \; &= \; (s-1)m(u_i + w)\exp(-(s-1)m(u_i + w))
\end{aligned}$$

Likewise

$$\begin{aligned}
\Pr(\mathsf{N}_j = 0) \; &= \; \exp(-(s-1)m(u_j + w)) \\
\Pr(\mathsf{N}_j = 1) \; &= \; (s-1)m(u_j + w)\exp(-(s-1)m(u_j + w))
\end{aligned}$$

Then

$$\Pr(\mathsf{N}_i + \mathsf{N}_j \geq 2) \; = \; 1 - \Pr(\mathsf{N}_i = 0)\Pr(\mathsf{N}_j = 0) - \Pr(\mathsf{N}_i = 0)\Pr(\mathsf{N}_j = 1) - \Pr(\mathsf{N}_i = 1)\Pr(\mathsf{N}_j = 0)$$

4.1.2 Joint probabilities of arrival and counts

$$\Pr(\breve{A}_i \wedge \mathsf{N}_i = 0) \;=\; \Pr(\breve{A}_i | \mathsf{N}_i = 0) \Pr(\mathsf{N}_i = 0) \;=\; \mathbb{I}[\hat{b}(i) = b] \Pr(\mathsf{N}_i = 0)$$
$$\Pr(\breve{A}_i \wedge \mathsf{N}_i = 1) \;=\; \Pr(\breve{A}_i | \mathsf{N}_i = 1) \Pr(\mathsf{N}_i = 1) \;=\; \mathbb{I}[\hat{b}(i) \neq b] \times (s-1)^{-1} \Pr(\mathsf{N}_i = 1)$$

since i must start elsewhere than branch b, and it has $s-1$ branches to go to, one of which is b. Likewise

$$\Pr(\breve{A}_j \wedge \mathsf{N}_j = 0) \;=\; \mathbb{I}[\hat{b}(j) = b] \Pr(\mathsf{N}_j = 0)$$
$$\Pr(\breve{A}_j \wedge \mathsf{N}_j = 1) \;=\; \mathbb{I}[\hat{b}(j) \neq b] \times (s-1)^{-1} \Pr(\mathsf{N}_j = 1).$$

4.1.3 The $(\mathsf{N}_i, \mathsf{N}_j) = (0,0)$ case

When a coalescence is possible in the (0,0) case (that is, when $\hat{b}(i) = \hat{b}(j) = b$),

$$\Pr\Big(T_{ij}(t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (0,0)\Big) = \exp(-\theta_b^{-1} w)$$

4.1.4 The $(\mathsf{N}_i, \mathsf{N}_j) = (0,1)$ case

When a coalescence is possible in the (0,1) case (that is, when $\hat{b}(i) = b$ and j migrates to b),

$$\Pr\Big(T_{ij}(t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (1,0)\Big) \;=\; (u_j + w)^{-1} u_j \exp(-\theta_b^{-1} w) \;+\; w\mathrm{uexp}(w\theta_b^{-1})$$

4.1.5 The $(\mathsf{N}_i, \mathsf{N}_j) = (1,0)$ case

When a coalescence is possible in the (1,0) case (that is, when $\hat{b}(j) = b$ and i migrates to b), i may migrate to b before $\breve{t}(j)$ or after, and

$$\Pr\Big(T_{ij}(t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j) = (1,0)\Big) \;=\; (u_i + w)^{-1} u_i \exp(-\theta_b^{-1} w) \;+\; w\mathrm{uexp}(w\theta_b^{-1})$$

4.1.6 The $\mathsf{N}_i + \mathsf{N}_j \geq 2$ case

Let $v = |\hat{t}(i) - \hat{t}(j)|$ be the time that only one of i and j exist. Then the probability that i and j are together at $t_{max} = \max(\hat{t}(i), \hat{t}(j))$ is

$$\Pr(Y_{ij}(t_{max}) = 2) = (1/s)(1 - \exp(-smv)) + \exp(-smv)\mathbb{I}[\hat{b}(i) = \hat{b}(j)] \tag{19}$$

Then

$$\Pr(Y_{ij}(t) = 2) = \exp(\Phi w)_{32}(1 - \Pr(Y_{ij}(t_{max}) = 2)) + \exp(\Phi w)_{22} \Pr(Y_{ij}(t_{max}) = 2) \tag{20}$$

and

$$\begin{aligned} &\Pr\Big(\mathsf{tmrca}(i,j) \geq t \wedge \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i + \mathsf{N}_j \geq 2)\Big) \;\simeq \\ &\frac{\widehat{P}_{bi}(t)\widehat{P}_{bj}(t)}{\sum_c \widehat{P}_{ci}(t)\widehat{P}_{cj}(t)} \;\Pr(Y_{ij}(t) = 2) \;\Pr(\mathsf{N}_i + \mathsf{N}_j \geq 2) \end{aligned} \tag{21}$$

4.2 'persisters'

Let u_i be the time that i spends alone, before j or k begin, or zero if j or k start first, so $u_i = \max(0, \min(\hat{t}(j), \hat{t}(k)) - \hat{t}(i))$. Likewise, define u_j and u_k. Let v_{ij} be the time during which i and j exist, but k does not, so $v_{ij} = \max(0, \hat{t}(k) - \max(\hat{t}(i), \hat{t}(j)))$. Likewise define v_{ik} and v_{jk}. Only one of u_i, u_j, u_k and one of v_{ij}, v_{ik}, v_{jk} can be nonzero. Let w be the time during which all three exist, so $w = t - \max(\hat{t}(i), \hat{t}(j), \hat{t}(k))$. we also set $t_i = t - \hat{t}(i)$, $t_j = t - \hat{t}(j)$, $t_k = t - \hat{t}(k)$. These are the duration of i and j, but only part of the duration of k.

4.2.1 Migration counts

$$\begin{aligned}\Pr(\mathsf{N}_k = 0) &= \exp(-(s-1)mt_k)\\ \Pr(\mathsf{N}_k = 1) &= (s-1)mt_k \exp(-(s-1)mt_k)\end{aligned}$$

with similar expressions for N_i and N_j, and

$$\begin{aligned}\Pr(\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2) &= 1 - \Pr(N_i = 0)\Pr(N_j = 0)\Pr(N_k = 0) -\\ &\Pr(N_i = 0)\Pr(N_j = 0)\Pr(N_k = 1) - \Pr(N_i = 0)\Pr(N_j = 1)\Pr(N_k = 0) -\\ &\Pr(N_i = 1)\Pr(N_j = 0)\Pr(N_k = 0) - \Pr(N_i = 1)\Pr(N_j = 1)\Pr(N_k = 0)\end{aligned}$$

4.2.2 Joint probabilities of arrival and counts

We have $\Pr(\breve{A}_i \wedge \mathsf{N}_i = 0)$, $\Pr(\breve{A}_j \wedge \mathsf{N}_j = 0)$, $\Pr(\breve{A}_i \wedge \mathsf{N}_i = 1)$, and $\Pr(\breve{A}_j \wedge \mathsf{N}_j = 1)$ from section 4.1.1. We do not need anything for k.

4.2.3 The $(\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,0,0)$ case

This case is impossible unless $\hat{b}(i) = \hat{b}(j) = b$. Given this, if $\hat{b}(k) \neq b$ no coalescence between k and i or j is possible, and if $\hat{b}(k) = b$, it may coalesce during the intervals that k and i or j co-exist. Thus

$$\Pr\left(T_{ij}(k,t) \;\middle|\; \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,0,0)\right) = \mathbb{I}[\hat{b}(k) \neq b] + \mathbb{I}[\hat{b}(k) = b]\exp(-(v_{ik} + v_{jk} + 2w)\theta_b^{-1})$$

4.2.4 The $(\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,0,1)$ case

Again, this case is impossible unless $\hat{b}(i) = \hat{b}(j) = b$. There is no coalescence unless $\hat{b}(k) \neq b$ and k migrates to b. Given $\hat{b}(k) \neq b$, the probability that k migrates to b is $1/(s-1)$. It may arrive in b before i or j have started, when one exists, buit not the other, or after both have started. The probabilities of these three arrival types are equal to the fraction of t_k during which they can occur. Thus

$$\begin{aligned}\Pr\left(T_{ij}(k,t) \;\middle|\; \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,0,1)\right) &= \mathbb{I}[\hat{b}(k) = b] +\\ \mathbb{I}[\hat{b}(k) \neq b]\ (s-1)^{-1}\ t_k^{-1}\ &\times \Big(u_k \exp(-(v_{ik} + v_{jk} + 2w)\theta_b^{-1}) +\\ (v_{ik} + v_{jk})\mathrm{uexp}(u_i + u_j)\theta_b^{-1})\exp(-2w\theta_b^{-1})\ &+\ w\,\mathrm{uexp}(2w\theta_b^{-1})\Big)\end{aligned}$$

4.2.5 The $(\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0, 1, 0)$ case

This case is impossible unless $\hat{b}(i) = b$ and $\hat{b}(j) \neq b$. If $\hat{b}(k)$ is neither b nor $\hat{b}(j)$, there can be no coalescence.

If $\hat{b}(k) = b$ and $\hat{b}(k) \neq \hat{b}(i)$, then k and i are together for a time $v_{ik} + w$. The lineage j may migrate to b before k starts during an interval of length $u_j + v_{ij}$, or while k exists, during an interval of length $v_{jk} + w$.

If $\hat{b}(k) \neq b$ and $\hat{b}(k) = \hat{b}(j)$, then k cannot coalesce with i, but may coalesce with j before j leaves $\hat{b}(k)$.

Note $\hat{b}(k) = b$ and $\hat{b}(k) = \hat{b}(j)$ cannot happen, since $\hat{b}(j) \neq b$.

$$
\begin{aligned}
&\Pr\Big(T_{ij}(k,t) \;\Big|\; \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (0,1,0)\Big) \;=\; \mathbb{I}[\hat{b}(k) \neq b]\; \mathbb{I}[\hat{b}(k) \neq \hat{b}(j)] \;+ \\
&\mathbb{I}[\hat{b}(k) = b]\; \mathbb{I}[\hat{b}(k) \neq \hat{b}(i)]\; \exp(-(v_{ik} + w)\theta_b^{-1}) \;\times \\
&\qquad t_j^{-1}\Big((u_j + v_{ij})\exp(-(v_{jk} + w)\theta_b^{-1}) + (v_{jk} + w)\mathrm{uexp}\big((v_{jk} + w)\theta_b^{-1}\big)\Big) \;+ \\
&\mathbb{I}[\hat{b}(k) \neq b]\; \mathbb{I}[\hat{b}(k) = \hat{b}(j)]\; t_j^{-1}\Big((u_j + v_{ij}) + (v_{jk} + w)\mathrm{uexp}\big((v_{jk} + w)\theta_{\hat{b}(k)}^{-1}\big)\Big)
\end{aligned}
$$

4.2.6 The $(\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (1, 0, 0)$ case

The same as last subsection 4.2.5 with i and j swapped.

4.2.7 The $(\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (1, 1, 0)$ case

This is impossible unless $\hat{b}(i) \neq b$ and $\hat{b}(j) \neq b$. Lineages i and j behave independently, so we can deal with them one at a time and multiply. If $\hat{b}(k) = b$, then i may migrate to b before k starts during an interval of length $u_i + v_{ij}$, or while k exists, during an interval of length $v_{ik} + w$. Similarly for j.

If $\hat{b}(k) \neq b$, then i may migrate away from $\hat{b}(k)$ before k starts during an interval of length $u_i + v_{ij}$, or while k exists, during an interval of length $v_{ik} + w$. Similarly for j.

$$
\begin{aligned}
&\Pr\Big(T_{ij}(k,t) \;\Big|\; \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i, \mathsf{N}_j, \mathsf{N}_k) = (1,1,0)\Big) \;= \\
&\mathbb{I}[\hat{b}(k) = b] \;\times \\
&t_i^{-1}\;\Big((u_i + v_{ij})\exp(-(v_{ik} + w)\theta_b^{-1}) + (v_{ik} + w)\mathrm{uexp}\big((v_{ik} + w)\theta_b^{-1}\big)\Big) \;\times \\
&t_j^{-1}\Big((u_j + v_{ij})\exp(-(v_{jk} + w)\theta_b^{-1}) + (v_{jk} + w)\mathrm{uexp}\big((v_{jk} + w)\theta_b^{-1}\big)\Big) \quad + \\
&\mathbb{I}[\hat{b}(k) \neq b] \;\times \\
&\Big(\mathbb{I}[\hat{b}(k) \neq \hat{b}(i)] \;+\; \mathbb{I}[\hat{b}(k) \neq \hat{b}(i)]t_i^{-1}\Big(u_i + v_{ij} + (v_{ik} + w)\mathrm{uexp}\big((v_{ik} + w)\theta_{\hat{b}(k)}^{-1}\big)\Big)\Big) \;\times \\
&\Big(\mathbb{I}[\hat{b}(k) \neq \hat{b}(j)] \;+\; \mathbb{I}[\hat{b}(k) \neq \hat{b}(j)]t_j^{-1}\Big(u_j + v_{ij} + (v_{jk} + w)\mathrm{uexp}\big((v_{jk} + w)\theta_{\hat{b}(k)}^{-1}\big)\Big)\Big)
\end{aligned}
$$

4.2.8 The $\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2$ case

We assume approximate independence to obtain

$$\Pr\left(T_{ij}(k,t) \mid \breve{A}_i \wedge \breve{A}_j \wedge (\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)\right) \simeq$$
$$\Pr(\mathsf{tmrca}(i,k) \geq t)\ \Pr(\mathsf{tmrca}(j,k) \geq t)\ \Pr(\breve{A}_i \wedge \breve{A}_j)\ \Pr(\mathsf{N}_i + \mathsf{N}_k \geq 2 \vee \mathsf{N}_j + \mathsf{N}_k \geq 2)$$

and $\Pr(\mathsf{tmrca}(i,k) \geq t)$ and $\Pr(\mathsf{tmrca}(j,k) \geq t)$ can be approximated using Φ.

Let $v = |\hat{t}(i) - \hat{t}(j)| = (u_i + v_{ij} + u_k + v_{jk})$ be the time that only one of i and k exist. The probability that i and k are together at $t_{max} = \max(\hat{t}(i), \hat{t}(k))$ is

$$\Pr(Y_{ik}(t_{max}) = 2) = (1/s)(1 - \exp(-smv) + \exp(-smv)\mathbb{I}[\hat{b}(i) = \hat{b}(k)] \tag{22}$$

Then the probability that i and k have not coalesced by t, namely $\Pr(\mathsf{tmrca}(i,k) \geq t)$, is approximated as

$$\begin{aligned}&\Pr(Y_{ik}(t) = 2)\ +\ \Pr(Y_{ik}(t) = 3)\ = \\ &\exp(\Phi w)_{32}(1 - \Pr(Y_{ik}(t_{max}) = 2)) + \exp(\Phi w)_{22} \Pr(Y_{ik}(t_{max}) = 2)\ + \\ &\exp(\Phi w)_{33}(1 - \Pr(Y_{ik}(t_{max}) = 2)) + \exp(\Phi w)_{23} \Pr(Y_{ik}(t_{max}) = 2)\end{aligned} \tag{23}$$

There is a similar expression for $\Pr(\mathsf{tmrca}(j,k) \geq t)$.

References

Jody Hey, Yujin Chung, Arun Sethuraman, Joseph Lachance, Sarah Tishkoff, Vitor C Sousa, and Yong Wang. Phylogeny Estimation by Integration over Isolation with Migration Models. *Molecular Biology and Evolution*, 35(11):2805–2818, 08 2018. ISSN 0737-4038. doi: 10.1093/molbev/msy162. URL `https://doi.org/10.1093/molbev/msy162`.

Michal Palczewski and Peter Beerli. A continuous method for gene flow. *Genetics*, 194(3):687–696, 2013. ISSN 0016-6731. doi: 10.1534/genetics.113.150904. URL `http://www.genetics.org/content/194/3/687`.