



Solving the migration–recombination equation from a genealogical point of view

F. Alberti¹ · E. Baake¹ · I. Letter² · S. Martínez³

Received: 19 April 2020 / Revised: 8 October 2020 / Accepted: 14 February 2021 /

Published online: 27 March 2021

© The Author(s) 2021

Abstract

We consider the discrete-time migration–recombination equation, a deterministic, nonlinear dynamical system that describes the evolution of the genetic type distribution of a population evolving under migration and recombination in a law of large numbers setting. We relate this dynamics (forward in time) to a Markov chain, namely a labelled partitioning process, backward in time. This way, we obtain a stochastic representation of the solution of the migration–recombination equation. As a consequence, one obtains an explicit solution of the nonlinear dynamics, simply in terms of powers of the transition matrix of the Markov chain. The limiting and quasi-limiting behaviour of the Markov chain are investigated, which gives immediate access to the asymptotic behaviour of the dynamical system. We finally sketch the analogous situation in continuous time.

Keywords Migration–recombination equation · Ancestral recombination graph · Duality · Labelled partitioning process · Quasi-stationarity · Haldane linearisation

Mathematics Subject Classification 60J75 · 92D15 · 60C05 · 05C80 · 37N25

✉ E. Baake
ebaake@techfak.uni-bielefeld.de

F. Alberti
falberti@math.uni-bielefeld.de

I. Letter
restucci@stats.ox.ac.uk

S. Martínez
smartine@dim.uchile.cl

¹ Faculty of Mathematics, Bielefeld University, Postbox 100131, 33501 Bielefeld, Germany

² Statistics Department, University of Oxford, 24-29 St Giles, Oxford OX1 3LB, UK

³ Department of Mathematical Engineering and Center of Mathematical Modeling, UMI 2807 UCHILE-CNRS, Universidad de Chile, Santiago, Chile

1 Introduction

Recombination is a genetic mechanism that ‘mixes’ or ‘reshuffles’ the genetic material of different individuals from generation to generation; it takes place during the reproductive cycle of sexually reproducing organisms. The analysis of models that describe the evolution of populations under recombination together with other processes are among the major challenges in population genetics.

In this contribution, we consider the evolution under the joint action of recombination and migration of individuals between discrete locations (or *demes*); we mainly focus on discrete time, where generations do not overlap. We will be concerned with a deterministic approach here, which assumes that the population is so large that a law of large numbers applies and random fluctuations (‘genetic drift’) may be neglected. The resulting *migration–recombination equation* is a large, nonlinear dynamical system that describes the evolution of the genetic composition of each local population over time, where the genetic composition is identified with a probability distribution (or measure) on a space of sequences of finite length. This model is a variant of the migration–selection–recombination equation formulated by Bürger (2009), who analysed its asymptotic behaviour in the classical dynamical systems setting, forward in time. It is our goal to complement this picture by relating this nonlinear dynamical system to a linear one by embedding the solution into a higher dimensional space, a technique known as *Haldane linearisation* (Hale and Ringwood 1983; Lyubich 1992) in the context of genetic algebras. This extends the approach taken by Baake and Baake (2016) to the case with migration. The resulting linear system has a natural interpretation as a Markov chain on the set of *labelled* partitions of the set of sequence sites. Intuitively, this Markov chain describes how the genetic material of an individual from the current population is partitioned across an increasing number of ancestors, along with their locations, as the lines of descent are traced back into the past. This backward (or *dual*) process combines a variant of the *ancestral recombination graph* (Hudson 1983; Griffiths and Marjoram 1996, 1997; Bhaskar and Song 2012; see also Durrett 2008, Ch. 3.4) with a variant of the *ancestral migration graph* (Notohara 1990; Matsen and Wakeley 2006). It is tractable in the law of large numbers regime considered here, due to the absence of coalescence; this was previously exploited for the recombination equation (without migration) by Baake et al. (2016), Baake and Baake (2016) and Martínez (2017); see Baake and Baake (2020) for a review. For an application of a similar idea in the context of the ancestral selection graph, see Slade and Wakeley (2005).

All this leads to a stochastic representation of the solution of the (nonlinear, deterministic) migration–recombination equation in terms of the labelled partitioning process. As a consequence, one obtains an explicit solution of the nonlinear dynamics, simply in terms of powers of the transition matrix of the Markov chain. In particular, the asymptotic behaviour of the migration–recombination equation emerges without any additional effort, via the (unique) absorbing state of the Markov chain. We also investigate the quasi-limiting behaviour of the Markov chain, based on ideas by Martínez (2017).

The paper is organised as follows. In Sect. 2, we set the scene and introduce the forward-time model. In Sect. 3, we use the notion of (labelled) *recombinators* to

reformulate the forward model in a compact way. A crucial property of the dynamics, namely its consistency under marginalisation, is established in Sect. 4. The core of the paper is Sect. 5, where we solve the forward iteration, together with Sect. 6, which establishes the connection to the labelled partitioning process in terms of a duality, together with a genealogical interpretation. Section 7 is devoted to its limiting and quasi-limiting behaviour, and Sect. 8 sketches how the approach carries over to continuous time.

2 The migration–recombination model

Let us recapitulate the discrete-time *migration–recombination equation* by Bürger (2009). The genetic information of an individual is encoded in terms of a finite sequence of letters, indexed by the set $[n] := \{1, \dots, n\}$ of sequence sites, where $n > 1$ is the fixed length of the sequences. The sites may either be interpreted as nucleotide positions in a DNA sequence, or as gene loci on a chromosome. For each site $i \in [n]$, there is a set (or *alphabet*) \mathcal{A}_i of *letters* (to be interpreted as nucleotides or alleles) that may possibly occur at that site. For the sake of simplicity, we restrict ourselves to *finite* sets \mathcal{A}_i here, but this generalises easily. A *type* is thus identified with a sequence

$$a = (a_1, \dots, a_n) \in \mathcal{A}_1 \times \dots \times \mathcal{A}_n =: \mathcal{A},$$

where \mathcal{A} is called the *type space*. We denote by $\mathcal{P}(\mathcal{A})$ the set of all probability measures on \mathcal{A} . We will also refer to such a probability measure as a *type distribution* or *population*. This implies that we consider *haploid individuals* or *gametes*; it will be sufficient to work at this level, since, in contrast to Bürger (2009), we do not consider selection. Indeed, in the absence of selection, diploid genotypes are independent combinations of haploid gametes at all stages of the life cycle, that is, one has Hardy-Weinberg equilibrium throughout.

It will be crucial for the later analysis to not only consider complete sequences (defined over the full set $[n]$), but also (sub)sequences (‘marginal’ types) that are defined over subsets of $[n]$. Given $U \subseteq [n]$, we set

$$\mathcal{A}_U := \prod_{i \in U} \mathcal{A}_i.$$

Note that $\mathcal{A}_{[n]} = \mathcal{A}$. Furthermore, \mathcal{A}_\emptyset is the empty Cartesian product, which is a set with a single element, namely the empty sequence e . For $V \subseteq U \subseteq [n]$ and $a \in \mathcal{A}_U$, we define the corresponding *marginal type* with respect to V by

$$a^V := (a_i)_{i \in V}.$$

In line with this, for $\nu \in \mathcal{P}(\mathcal{A}_U)$, we define its *marginal distribution* (with respect to V) as the probability distribution on \mathcal{A}_V given by

$$\nu^V(E) := \nu(E \times \mathcal{A}_{U \setminus V})$$

for all $E \subseteq \mathcal{A}_V$. In words, $\nu^V(E)$ is the probability that the marginal with respect to V of a randomly sampled type from ν agrees with some element of E . In somewhat more technical terms, ν^V is the push-forward of ν under the canonical projection from \mathcal{A}_U to \mathcal{A}_V . Clearly, the map $\nu \mapsto \nu^V$ is linear.

In order to discuss migration, we introduce a finite set L of *locations* (or *demes*). The population at location $\alpha \in L$ in generation $t \in \mathbb{N}_0$ is denoted by $\mu_t(\alpha) \in \mathcal{P}(\mathcal{A})$. The collection of all local populations is summarised into the (column) vector $\mu_t = (\mu_t(\alpha))_{\alpha \in L} \in \mathcal{P}(\mathcal{A})^L$; we call μ a *spatially structured population*, or a *metapopulation*. Moreover, for $U \subseteq [n]$, we write $\mu_t^U = (\mu_t^U(\alpha))_{\alpha \in L}$ for the vector of marginal populations. Throughout, indices of vectors and matrices that refer to locations are written as arguments. Unless stated otherwise, vectors are understood as column vectors.

The migration–recombination equation is a discrete-time dynamical system that describes the deterministic evolution of a metapopulation with non-overlapping generations. We assume that, in each generation, this evolution proceeds in two stages. First, individuals migrate between locations; then, random mating takes place among individuals at the same location, followed by reproduction involving recombination.¹ Discrete generations will be indexed by $t \in \mathbb{N}_0$, where a population at time t is understood as the population after the t th round of recombination, but before migration; we will use the corresponding half integers $t + \frac{1}{2}$ to indicate the population after migration, but before recombination.

2.1 Describing migration

We first consider migration, following the presentation by Nagylaki (1992, Chap. 6.2). The most straightforward way to describe migration is via the so-called *forward migration matrix* \tilde{M} . It is a stochastic matrix indexed by L , where the entry $\tilde{M}(\alpha, \beta)$ is the probability that a randomly chosen individual at location α migrates to location β in the next generation. However, it is more convenient to work instead with the *backward migration matrix* M . It is also a stochastic matrix, and $M(\alpha, \beta)$ is the probability that a randomly chosen individual that currently lives at location α has migrated from location β . We assume that the local population sizes $c(\alpha) \in \mathbb{R}_{>0}$ remain constant over time. This is the case if either

$$c(\alpha) = \sum_{\beta \in L} c(\beta) \tilde{M}(\beta, \alpha) \quad (1)$$

for all $\alpha \in L$, or if population regulation takes place after the migration step. In any case, denoting the location of a randomly sampled individual at time $t + \frac{1}{2}$ by $\ell_{t+\frac{1}{2}}$

¹ While recombination, strictly speaking, does *not* occur during reproduction itself, this is not relevant in the simplified setting of our model; simply put, as we are working at the level of gametes, the word ‘reproduction’ refers, in this context, to the formation of new germ cells prior to mating.

and its location in generation t by ℓ_t , we have

$$\begin{aligned}
 M(\alpha, \beta) &= \mathbb{P}(\ell_t = \beta \mid \ell_{t+\frac{1}{2}} = \alpha) = \frac{\mathbb{P}(\ell_t = \beta, \ell_{t+\frac{1}{2}} = \alpha)}{\mathbb{P}(\ell_{t+\frac{1}{2}} = \alpha)} \\
 &= \frac{\mathbb{P}(\ell_{t+\frac{1}{2}} = \alpha \mid \ell_t = \beta)\mathbb{P}(\ell_t = \beta)}{\mathbb{P}(\ell_{t+\frac{1}{2}} = \alpha)} = \frac{c(\beta)}{c(\alpha)} \tilde{M}(\beta, \alpha).
 \end{aligned}$$

Note that M is stochastic by definition, i.e. $\sum_{\beta \in L} M(\alpha, \beta) = 1$ and $M(\alpha, \beta) \geq 0$ for all $\alpha, \beta \in L$.

For additional background, see Nagylaki (1992, Chapter 6.2). In what follows, we will work exclusively with the backward migration matrix. Since we are only interested in relative type frequencies, the population sizes $c(\alpha)$ are irrelevant. After migration (but before recombination), the local population at α is therefore given by

$$\mu_{t+\frac{1}{2}}(\alpha) = \sum_{\beta \in L} M(\alpha, \beta)\mu_t(\beta), \tag{2}$$

and the metapopulation may be written compactly as

$$\mu_{t+\frac{1}{2}} = M\mu_t. \tag{3}$$

2.2 Describing recombination

To describe recombination, we slightly modify the model by Bürger (2009) and follow the notation of Martínez (2017). Here, the partitions of $[n]$ and its subsets will play a central role, see also Baake and Baake (2016) or Baake et al. (2016). For $U \subseteq [n]$, a partition of U is a set δ of mutually disjoint, non-empty subsets of U whose union is U . We will also refer to the elements of a partition as *blocks*. The set of all partitions of U is denoted by $\mathbb{S}(U)$. We say that ε is finer than (is a refinement of) δ ($\varepsilon \preceq \delta$) if every block of ε is contained in some block of δ . The relation \preceq defines a partial order on $\mathbb{S}(U)$. We denote the unique minimal and maximal elements in $\mathbb{S}(U)$ by $\underline{0}_U := \{\{i\} : i \in U\}$ and $\underline{1}_U := \{U\}$; when $U = [n]$, we drop the subscript and write $\underline{0}$ and $\underline{1}$, rather than $\underline{0}_{[n]}$ and $\underline{1}_{[n]}$. By $\delta \wedge \varepsilon$ we denote the coarsest common refinement of δ and ε , that is,

$$\delta \wedge \varepsilon := \{d \cap e : d \cap e \neq \emptyset, d \in \delta, e \in \varepsilon\};$$

it is the coarsest partition finer than both δ and ε .

We say that an offspring of a local population ν is *recombined according to* $\delta = \{d_1, \dots, d_m\} \in \mathbb{S}([n])$ if it has m parents of types $a^{(1)}, \dots, a^{(m)} \in \mathcal{A}^{[n]}$, all sampled independently from ν , and inherits the letters at the sites in d_i from the parent of type $a^{(i)}$. That is, the type of the offspring is $b = (b_1, \dots, b_n)$, where $b_i := a_i^{(j)}$ if $i \in d_j$. The biologically reasonable cases are $m = 1$ (then $\delta = \underline{1}$ and the full offspring

sequence is inherited from a single parent) and $m = 2$ (the offspring sequence is pieced together from two parents). The choice $m > 2$ implies more than two parents, which is not biologically realistic, but we include this case (and, in this way, generalise Bürger (2009)) since it is mathematically interesting and does not require additional effort. It is then clear that the type of an offspring of ν that is recombined according to δ has the distribution

$$\bigotimes_{d \in \delta} \nu^d.$$

That is, recombination according to δ turns ν into the product measure of the marginals with respect to the blocks of δ ; this reflects the random mating, that is, the independence of the parents. Again, we understand this product to respect the ordering of the sites.

We assume that, in each time step, the entire local population is replaced; the proportion of individuals that are replaced by offspring recombined according to $\delta \in \mathbb{S}([n])$ is denoted by $r_\delta \geq 0$, where

$$\sum_{\delta \in \mathbb{S}([n])} r_\delta = 1.$$

The collection $(r_\delta)_{\delta \in \mathbb{S}([n])}$ is known as the *recombination distribution*. Thus, the components of μ_{t+1} are given by

$$\mu_{t+1}(\alpha) = \sum_{\delta \in \mathbb{S}([n])} r_\delta \bigotimes_{d \in \delta} \mu_{t+\frac{1}{2}}^d(\alpha) = \sum_{\delta \in \mathbb{S}([n])} r_\delta \bigotimes_{d \in \delta} \sum_{\beta \in L} M(\alpha, \beta) \mu_t^d(\beta), \quad (4)$$

where we have used (2) and the linearity of marginalisation in the last step. Equation (4) is called the *migration–recombination equation*, or MRE for short.

3 Reformulation of the model

Extending concepts established by Baake et al. (2016), Baake and Baake (2016) and Martínez (2017), we now formulate the MRE (4) in a more compact way. This involves labelling the blocks of a partition by elements of L to keep track of where the letters in the blocks come from.

Definition 3.1 A labelled partition of $U \subseteq [n]$ is a collection $\delta := \{\mathbf{d}_1, \dots, \mathbf{d}_m\}$ for some $m \leq |U|$, where $\mathbf{d}_i = (d_i, \lambda_i)$, $\delta = \{d_1, \dots, d_m\}$ is a partition of U , and $\lambda_i \in L$ for $1 \leq i \leq m$. We call δ the *base* of δ , refer to its elements as the blocks of δ , and interpret λ_i as the *label* of block d_i . We write $\mathbb{LS}(U)$ for the set of all labelled partitions of U . \diamond

In order to rewrite Eq. (4), we now introduce the *labelled recombinator*. It is the labelled analogue of the recombinator used by Baake and Baake (2016) for unlabelled partitions. Since we will later also be interested in the evolution of the distribution of subsequences (compare Sect. 4), we introduce the concept in the required generality right away.

Definition 3.2 Let $U \subseteq [n]$ and $\delta \in \mathbb{LS}(U)$. Then, the labelled recombinator (with respect to δ), namely $\mathcal{R}_\delta^U : \mathcal{P}(\mathcal{A}_U)^L \rightarrow \mathcal{P}(\mathcal{A}_U)$, is defined by

$$\mathcal{R}_\delta^U(v) := \bigotimes_{(d,\lambda) \in \delta} v^d(\lambda);$$

if $U = [n]$, we will drop the superscript and write \mathcal{R}_δ instead of $\mathcal{R}_\delta^{[n]}$. ◊

In words, $\mathcal{R}_\delta(v)$ is the distribution of the type of an offspring individual that is recombined according to δ , where the parent of the labelled block (d, λ) is sampled from the local population $v(\lambda)$. A similar interpretation holds for the *marginal* recombinators; see Theorem 4.3 and Remark 4.5. With this, Eq. (4) can be restated as follows.

Lemma 3.3 *The MRE (4) can be written as*

$$\mu_{t+1} = \sum_{\delta \in \mathbb{LS}([n])} p_\delta \mathcal{R}_\delta(\mu_t) \tag{5}$$

with

$$p_\delta := (p_\delta(\alpha))_{\alpha \in L},$$

where the migration–recombination probabilities are given by

$$p_\delta(\alpha) := r_\delta \prod_{(d,\lambda) \in \delta} M(\alpha, \lambda)$$

and are normalised, i.e. satisfy

$$\sum_{\delta \in \mathbb{LS}([n])} p_\delta(\alpha) = 1 \text{ for all } \alpha \in L.$$

Proof This follows immediately from Definition 3.2 by expanding the measure product in Eq. (4):

$$\begin{aligned} \mu_{t+1}(\alpha) &= \sum_{\delta \in \mathbb{S}([n])} r_\delta \bigotimes_{d \in \delta} \sum_{\lambda \in L} M(\alpha, \lambda) \mu_t^d(\lambda) \\ &= \sum_{\delta \in \mathbb{S}([n])} \sum_{\lambda \in L^\delta} r_\delta \prod_{d \in \delta} M(\alpha, \lambda_d) \bigotimes_{d \in \delta} \mu_t^d(\lambda_d) \\ &= \sum_{\delta \in \mathbb{LS}([n])} p_\delta(\alpha) \bigotimes_{(d,\lambda) \in \delta} \mu_t^d(\lambda) = \sum_{\delta \in \mathbb{LS}([n])} p_\delta(\alpha) \mathcal{R}_\delta(\mu_t), \end{aligned}$$

where in the third step, we identified the double sum over all partitions of $[n]$ and all possible vectors of labels of their blocks with the sum over all labelled partitions. The normalisation is a consequence of $\sum_{\delta \in \mathbb{S}([n])} r_\delta = 1 = \sum_{\beta \in L} M(\alpha, \beta)$. □

We call the probability distribution $p(\alpha) = (p_\delta(\alpha))_{\delta \in \mathbb{LS}([n])}$ the *migration–recombination distribution at α* .

Remark 3.4 Lemma 3.3 has a simple stochastic interpretation. To sample the type of an individual in generation $t + 1$ (say, at location α), we first pick a random labelled partition δ according to $p(\alpha)$ and subsequently sample from $\mathcal{R}_\delta(\mu_t)$. The factorisation of $p_\delta(\alpha)$ in Lemma 3.3 implies that the genome is first partitioned across its parents according to δ , with probability r_δ . Subsequently, the label is reassigned (conditionally) independently for each block, according to $M(\alpha, \cdot)$, as we trace back the origin of each ancestor. Finally, the offspring type is determined by piecing together (fragments of) independent samples of the ancestral sequences at the appropriate locations, in generation t . This leads to the product measure in Definition 3.2. We will further elaborate on this in Sect. 6. \diamond

To continue, we need a few additional concepts around labelled partitions. First, the notion of an *induced (labelled) partition* is required. For $\emptyset \neq V \subseteq U$ and $\delta \in \mathbb{LS}(U)$, we denote by $\delta|_V$ the labelled partition induced by δ on V ; it is given by

$$\delta|_V := \{(d \cap V, \lambda) : d \cap V \neq \emptyset, (d, \lambda) \in \delta\}$$

with base

$$\delta|_V = \{d \cap V : \emptyset \neq d \cap V, d \in \delta\},$$

the partition induced by the (unlabelled) partition δ on V . Simply put, every block inherits the label of the unique block of the original partition that contains it.

Conversely, given a partition δ of U and a family $(\epsilon_d)_{d \in \delta}$ of labelled partitions of its blocks, their union

$$\bigcup_{d \in \delta} \epsilon_d$$

is a labelled partition of U ; its base is the union

$$\bigcup_{d \in \delta} \epsilon_d$$

of the bases ϵ_d .

Finally, given two labelled partitions δ and ϵ , we say that ϵ is *finer than δ* ($\epsilon \preceq \delta$) if $\epsilon \preceq \delta$. The partial order on $\mathbb{S}(U)$ thus carries over to a partial order on $\mathbb{LS}(U)$. For any $\alpha \in L$, there is a unique maximal element; namely, the labelled partition $\mathbb{1}_U^\alpha := \{([n], \alpha)\}$ that consists of a single block with label α . If $U = [n]$, we drop the subscript.

Remark 3.5 It is not difficult to see that $\epsilon \preceq \delta$ if and only if

$$\epsilon = \bigcup_{d \in \delta} \epsilon|_d.$$

For a fixed $\delta \in \mathbb{S}([n])$, this implies the following bijection between the labelled partitions ϵ with $\epsilon \preceq \delta$ and collections $(\epsilon_d)_{d \in \delta}$ of labelled partitions of the individual

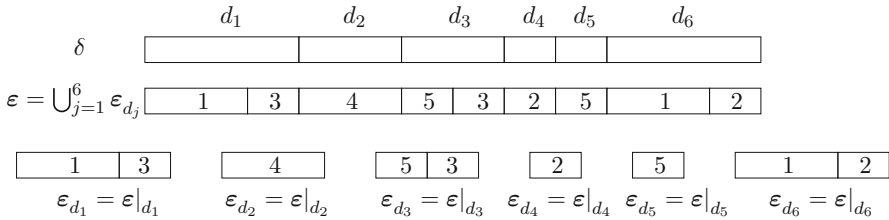


Fig. 1 At the top, an unlabelled partition of $[n]$. In the middle, a labelled refinement of δ , which gives rise to labelled partitions of the blocks of δ (bottom). Conversely, one can start with the collection of labelled partitions at the bottom and join them to obtain a labelled refinement of δ

blocks of δ . Given ϵ with $\epsilon \preceq \delta$, we obtain the collection $(\epsilon|_d)_{d \in \delta}$ of labelled partitions induced by ϵ on the blocks of δ . Conversely, given a collection $(\epsilon_d)_{d \in \delta}$ of labelled partitions of the blocks of δ , we set $\epsilon := \bigcup_{d \in \delta} \epsilon_d$; note that $\epsilon \preceq \delta$ and $\epsilon|_d = \epsilon_d$. See also Fig. 1. \diamond

We will now see that the recombinator for a union of labelled partitions of disjoint subsets is the product of the recombinators for the individual labelled partitions; compare also Baake and Baake (2016, Lemma 2).

Lemma 3.6 *Let $\delta \in \mathbb{S}([n])$ and $\epsilon_d \in \mathbb{L}\mathbb{S}(d)$ for all $d \in \delta$. Then, for all $v \in \mathcal{P}(X)^L$,*

$$\mathcal{R}_{\bigcup_{d \in \delta} \epsilon_d}(v) = \bigotimes_{d \in \delta} \mathcal{R}_{\epsilon_d}^d(v^d).$$

In particular, for $\epsilon \in \mathbb{L}\mathbb{S}([n])$ with $\epsilon \preceq \delta$, we have

$$\mathcal{R}_\epsilon(v) = \bigotimes_{d \in \delta} \mathcal{R}_{\epsilon|_d}^d(v^d).$$

Proof For the first claim, we write out the labelled recombinators and see that

$$\bigotimes_{d \in \delta} \mathcal{R}_{\epsilon_d}^d(v^d) = \bigotimes_{d \in \delta} \bigotimes_{(e, \lambda) \in \epsilon_d} v^e(\lambda) = \bigotimes_{(e, \lambda) \in \bigcup_{d \in \delta} \epsilon_d} v^e(\lambda) = \mathcal{R}_{\bigcup_{d \in \delta} \epsilon_d}(v).$$

For the second claim, see Remark 3.5. \square

We now turn to the marginalisation consistency of the MRE (4), a property that will turn out as the key to its solution.

4 Marginalisation consistency

Just as in the continuous-time case for pure recombination treated by Baake and Baake (2016), the marginalisation consistency of the model is a crucial ingredient. We therefore now turn to the dynamics that the MRE (4) induces on the marginal type distributions. As a warm-up, we prove the following elementary, but useful, result.

Lemma 4.1 *Let $U, V \subseteq [n], U \cap V = \emptyset$, and let ν_U, ν_V be probability measures on \mathcal{A}_U and \mathcal{A}_V , respectively. Then, we have for any $W \subseteq U \cup V$*

$$(\nu_U \otimes \nu_V)^W = \nu_U^{U \cap W} \otimes \nu_V^{V \cap W}.$$

Proof Note that $\mathcal{A}_W = \mathcal{A}_{U \cap W} \times \mathcal{A}_{V \cap W}$. Let us fix $E_{U \cap W} \subseteq \mathcal{A}_{U \cap W}$ and $E_{V \cap W} \subseteq \mathcal{A}_{V \cap W}$. Then, for any $W \subseteq U \cup V$,

$$\begin{aligned} (\nu_U \otimes \nu_V)^W(E_{U \cap W} \times E_{V \cap W}) &= (\nu_U \otimes \nu_V)(E_{U \cap W} \times E_{V \cap W} \times \mathcal{A}_{(U \cup V) \setminus W}) \\ &= (\nu_U \otimes \nu_V)((E_{U \cap W} \times \mathcal{A}_{U \setminus W}) \times (E_{V \cap W} \times \mathcal{A}_{V \setminus W})) \\ &= \nu_U(E_{U \cap W} \times \mathcal{A}_{U \setminus W}) \nu_V(E_{V \cap W} \times \mathcal{A}_{V \setminus W}) \\ &= \nu_U^{U \cap W}(E_{U \cap W}) \nu_V^{V \cap W}(E_{V \cap W}). \end{aligned}$$

□

Remark 4.2 It is important to note that Lemma 4.1 remains true if $U \cap W = \emptyset$ or $V \cap W = \emptyset$. Assume, for instance, that $U \cap W = \emptyset$. Recalling that the empty Cartesian product \mathcal{A}_\emptyset is the singleton $\{e\}$ (recall that e is the empty sequence), $\nu_U^{U \cap W}$ is the unique probability measure on $\{e\}$ and can be treated as the scalar 1, in the sense that

$$\nu_U^{U \cap W} \otimes \nu_V^{V \cap W} = \nu_V^{V \cap W} \otimes \nu_U^{U \cap W} = \nu_V^{V \cap W}.$$

◇

We now prove the main result of this section, which shows that the MRE is consistent under marginalisation.

Theorem 4.3 *Let $(\mu_t)_{t \in \mathbb{N}_0}$ be a solution of the MRE (4) and U a nonempty subset of $[n]$. Then, $(\mu_t^U)_{t \in \mathbb{N}_0}$ satisfies the marginal MRE*

$$\mu_{t+1}^U = \sum_{\delta \in \mathbb{LS}(U)} p_\delta^U \mathcal{R}_\delta^U(\mu_t^U),$$

where p_δ^U is given by

$$p_\delta^U := \sum_{\substack{\delta' \in \mathbb{LS}([n]) \\ \delta' \upharpoonright_U = \delta}} p_{\delta'} \text{ for } \delta \in \mathbb{LS}(U).$$

Proof By Lemma 3.3 and the linearity of marginalisation, we have

$$\mu_{t+1}^U = \left(\sum_{\delta' \in \mathbb{LS}([n])} p_{\delta'} \mathcal{R}_{\delta'}(\mu_t) \right)^U = \sum_{\delta' \in \mathbb{LS}([n])} p_{\delta'} (\mathcal{R}_{\delta'}(\mu_t))^U.$$

Using Lemma 4.1, we obtain for all $\delta' \in \mathbb{LS}([n])$

$$\begin{aligned} (\mathcal{R}_{\delta'}(\mu_t))^{U} &= \left(\bigotimes_{(d', \lambda') \in \delta'} \mu_t^{d'}(\lambda') \right)^U = \bigotimes_{\substack{(d', \lambda') \in \delta' \\ d' \cap U \neq \emptyset}} \mu_t^{d' \cap U}(\lambda') \\ &= \bigotimes_{(d, \lambda) \in \delta' |_U} \mu_t^d(\lambda) = \mathcal{R}_{\delta' |_U}^U(\mu_t^U), \end{aligned}$$

where in the second step, we ignored the factors corresponding to d' with $d' \cap U = \emptyset$ (compare Remark 4.2). Thus,

$$\mu_{t+1}^U = \sum_{\delta' \in \mathbb{LS}([n])} p_{\delta'} \mathcal{R}_{\delta' |_U}^U(\mu_t^U) = \sum_{\delta \in \mathbb{LS}(U)} p_{\delta}^U \mathcal{R}_{\delta}^U(\mu_t^U),$$

which is what we wanted to show. □

The $p_{\delta}^U(\alpha)$ are the *marginal migration–recombination probabilities* (at α), and, accordingly, $p^U(\alpha) = (p_{\delta}^U(\alpha))_{\delta \in \mathbb{LS}([n])}$ is called the *marginal migration–recombination distribution* (at α). We will now see that the marginal migration–recombination probabilities have a product structure analogous to that of the migration–recombination probabilities in Lemma 3.3.

Lemma 4.4 *The marginal migration–recombination probabilities $p_{\delta}^U(\alpha)$ from Theorem 4.3 can be written as*

$$p_{\delta}^U(\alpha) = \left(\sum_{\substack{\delta' \in \mathbb{S}([n]) \\ \delta' |_U = \delta}} r_{\delta'} \right) \prod_{(d, \lambda) \in \delta} M(\alpha, \lambda).$$

Proof We write the (given) labelled partition δ as

$$\delta = \{(d_1, \lambda_1), \dots, (d_k, \lambda_k)\}.$$

Next, we split the conditional sum over the labelled partitions into the sums over the appropriate partitions and their labels. Thus,

$$\begin{aligned} p_{\delta}^U(\alpha) &= \sum_{\substack{\delta' \in \mathbb{LS}([n]) \\ \delta' |_U = \delta}} p_{\delta'}(\alpha) \\ &= \sum_{\substack{\delta' = \{d'_1, \dots, d'_m\} \in \mathbb{S}([n]) \\ \{d'_1, \dots, d'_m\} |_U = \delta}} r_{\delta'} \sum_{\lambda'_1, \dots, \lambda'_m \in L} \prod_{j=1}^k \mathbb{1}_{\lambda'_j = \lambda_j} \prod_{j=1}^m M(\alpha, \lambda'_j), \end{aligned} \tag{6}$$

where the blocks are indexed so that $d'_j \cap U = d_j$ for all $1 \leq j \leq k$ and $d'_j \cap U = \emptyset$ for $k + 1 \leq j \leq m$. Clearly,

$$\begin{aligned} & \sum_{\lambda'_1, \dots, \lambda'_m \in L} \prod_{j=1}^k \mathbb{1}_{\lambda'_j = \lambda_j} \prod_{j=1}^m M(\alpha, \lambda'_j) \\ &= \left(\sum_{\lambda'_1, \dots, \lambda'_k \in L} \prod_{j=1}^k \mathbb{1}_{\lambda'_j = \lambda_j} \prod_{j=1}^k M(\alpha, \lambda'_j) \right) \left(\sum_{\lambda'_{k+1}, \dots, \lambda'_m \in L} \prod_{j=k+1}^m M(\alpha, \lambda'_j) \right), \end{aligned}$$

with the usual convention that the empty product is 1. Now, we can use the indicator in the first bracket to eliminate the summation, yielding

$$\sum_{\lambda'_1, \dots, \lambda'_k \in L} \prod_{j=1}^k \mathbb{1}_{\lambda'_j = \lambda_j} \prod_{j=1}^k M(\alpha, \lambda'_j) = \prod_{j=1}^k M(\alpha, \lambda_j).$$

The second bracket is equal to one, by the stochasticity of M :

$$\sum_{\lambda'_{k+1}, \dots, \lambda'_m \in L} \prod_{j=k+1}^m M(\alpha, \lambda'_j) = \prod_{j=k+1}^m \sum_{\lambda' \in L} M(\alpha, \lambda') = 1.$$

Inserting this back into (6) finishes the proof. □

Remark 4.5 The same stochastic interpretation as for Eq. (5) (see Remark 3.4) holds also for the marginalised system. With probability

$$r_\delta^U := \sum_{\substack{\delta' \in \mathbb{S}(\{n\}) \\ \delta' \upharpoonright U = \delta}} r_\delta, \tag{7}$$

the subsequence with respect to U of a sampled individual is partitioned across its ancestors according to δ . Then, the labels are reassigned independently according to M , reflecting their independent migration. ◇

5 Solution of the forward iteration

Next, we use the marginalisation consistency established in the previous section to tame the MRE (4). As discussed by Baake and Baake (2016) for pure recombination, the main idea is to consider the time evolution of the (column) vector $\mathcal{R}(\mu_t) = (\mathcal{R}_\delta(\mu_t))_{\delta \in \mathbb{L}\mathbb{S}(\{n\})}$, rather than μ_t alone; note that we recover $\mu_t(\alpha)$ as the $\mathbf{1}^\alpha$ -component of $\mathcal{R}(\mu_t)$.

Theorem 5.1 *Let T be the matrix, indexed by the labelled partitions $\mathbb{LS}([n])$, with entries*

$$T_{\delta\epsilon} = \begin{cases} 0, & \text{if } \epsilon \not\preceq \delta, \\ \prod_{(d,\lambda)\in\delta} p_{\epsilon|d}^d(\lambda), & \text{if } \epsilon \preceq \delta, \end{cases}$$

where the $p_{\epsilon|d}^d(\lambda)$ are as in Lemma 4.4. Then, T is a stochastic matrix. Assume that $(\mu_t)_{t\in\mathbb{N}_0}$ satisfies the MRE (4). Then, $\mathcal{R}(\mu_t)$ satisfies the linear recursion

$$\mathcal{R}(\mu_{t+1}) = T\mathcal{R}(\mu_t).$$

In particular,

$$\mathcal{R}(\mu_t) = T^t\mathcal{R}(\mu_0),$$

where T^t denotes the t th power of T .

Proof By Definition 3.2 and Theorem 4.3,

$$\begin{aligned} \overline{\mathcal{R}}_{\delta}(\mu_{t+1}) &= \bigotimes_{(d,\lambda)\in\delta} \sum_{\epsilon_d\in\mathbb{LS}(d)} p_{\epsilon_d}^d(\lambda)\mathcal{R}_{\epsilon_d}^d(\mu_t^d) \\ &= \sum_{\substack{\epsilon_d\in\mathbb{LS}(d) \\ \forall d\in\delta}} \left(\prod_{(d,\lambda)\in\delta} p_{\epsilon_d}^d(\lambda) \right) \bigotimes_{d\in\delta} \mathcal{R}_{\epsilon_d}^d(\mu_t^d). \end{aligned}$$

By Remark 3.5, the right-hand side is equal to

$$\sum_{\epsilon\preceq\delta} \left(\prod_{(d,\lambda)\in\delta} p_{\epsilon|d}^d(\lambda) \right) \bigotimes_{d\in\delta} \mathcal{R}_{\epsilon|d}^d(\mu_t^d) = \sum_{\epsilon\preceq\delta} T_{\delta\epsilon}\mathcal{R}_{\epsilon}(\mu_t),$$

where we used Lemma 3.6 and the underdot indicates the summation variable. That T is a stochastic matrix is a straightforward consequence of $p^d(\alpha)$ being a probability distribution on $\mathbb{LS}(d)$ for all $d \subseteq [n]$ and all $\alpha \in L$. □

We have just witnessed how the solution of a nonlinear system, embedded in a higher dimensional space, turns into the solution of a linear system and may thus be given explicitly, simply via matrix powers. This is an extension of a technique called *Haldane linearisation* (Hale and Ringwood 1983; Baake and Baake 2016, 2020) to the case with migration. The underlying mechanism can be found in the genealogical structure, which is discussed next.

6 Stochastic interpretation, genealogical content, and duality

Let us now turn to the probabilistic content of Theorem 5.1. We will see that the appearance of the stochastic matrix T is no coincidence; rather, it has a natural interpretation as the transition matrix of a Markov chain describing the random genealogy of a single individual.

Definition 6.1 The *labelled partitioning process* (LPP) is a discrete-time Markov chain $\Sigma := (\Sigma_t)_{t \in \mathbb{N}_0}$ with values in $\mathbb{LS}([n])$ and transition matrix T , that is,

$$\mathbb{P}(\Sigma_{t+1} = \epsilon \mid \Sigma_t = \delta) = T_{\delta\epsilon}$$

for all $\delta, \epsilon \in \mathbb{LS}([n])$. ◇

In words, Σ_{t+1} is constructed from Σ_t by independently replacing each labelled block $(d, \lambda) \in \Sigma_t$, with probability $p_{\epsilon_d}^d(\lambda)$, by the (labelled) blocks of ϵ_d ; see also Fig. 2.

The genealogical interpretation of Σ , started in $\mathbf{1}^\alpha$, is as follows. Each labelled block (d, λ) of Σ_t corresponds to a different ancestor of the individual at present, sampled at location α , who lived at location λ , at t generations before the present. The elements of d are the sequence sites that are inherited from this ancestor. As we look one generation further into the past, d is replaced by the blocks of a labelled partition $\epsilon_d \in \mathbb{LS}(d)$, which describes how the type of that ancestor is, in turn, pieced together from its parents, alive $t + 1$ generations before the present. Note that, now, the labelled partitions of d are relevant rather than those of $[n]$. This is because we already know that this ancestor only contributes sites contained in d , whence we only need to trace back the ancestry of these sites. (This reflects the marginalisation consistency of the model, compare Remark 4.5). Furthermore, the various blocks split independently as the population, in the law of large numbers regime assumed here, is so large that two given individuals never share a common ancestor; thus, their lineages are conditionally independent.

The connection between the solution of the MRE (4) and the genealogical process is formalised in the following theorem, which is a probabilistic restatement of Theorem 5.1 and draws on the notion of duality for Markov processes (Liggett 2010; Jansen and Kurt 2014); in particular, we think about the solution of the forward-time equation as a Markov chain with deterministic transitions.

Theorem 6.2 *The LPP and the solution of the MRE (4) are dual with respect to the duality function*

$$(\delta, \mu) \mapsto \mathcal{R}_\delta(\mu).$$

That is, for all $\delta \in \mathbb{LS}([n])$ and all $\mu_0 \in \mathcal{P}(\mathcal{A})^L$, we have

$$\mathbb{E}[\mathcal{R}_{\Sigma_t}(\mu_0) \mid \Sigma_0 = \delta] = \mathcal{R}_\delta(\mu_t).$$

In particular, this entails the stochastic representation

$$\mu_t(\alpha) = \mathbb{E}[\mathcal{R}_{\Sigma_t}(\mu_0) \mid \Sigma_0 = \mathbf{1}^\alpha]$$

for the solution of the MRE (4).

Proof We prove the theorem by induction over t . For $t = 0$, there is nothing to show. Assuming now that

$$\mathbb{E}[\mathcal{R}_{\Sigma_t}(\mu_0) \mid \Sigma_0 = \delta] = \mathcal{R}_\delta(\mu_t)$$

for any $t > 0$, we compute, using Theorem 5.1 in the first step, the induction hypothesis in the second, time-homogeneity in the third, and the Markov property in the last:

$$\begin{aligned} \mathcal{R}_\delta(\mu_{t+1}) &= \sum_{\varepsilon \preceq \delta} T_{\delta\varepsilon} \mathcal{R}_\varepsilon(\mu_t) \\ &= \sum_{\varepsilon \preceq \delta} \mathbb{P}[\Sigma_1 = \varepsilon \mid \Sigma_0 = \delta] \mathbb{E}[\mathcal{R}_{\Sigma_t}(\mu_0) \mid \Sigma_0 = \varepsilon] \\ &= \sum_{\varepsilon \preceq \delta} \mathbb{P}[\Sigma_1 = \varepsilon \mid \Sigma_0 = \delta] \mathbb{E}[\mathcal{R}_{\Sigma_{t+1}}(\mu_0) \mid \Sigma_1 = \varepsilon] \\ &= \sum_{\varepsilon \preceq \delta} \mathbb{E}[\mathcal{R}_{\Sigma_{t+1}}(\mu_0) \mid \Sigma_0 = \delta]. \end{aligned}$$

This proves the statement for $t + 1$. □

Note that the duality function used here is vector valued. This is a slight extension of the standard notion, since the duality function is usually assumed to take values in \mathbb{R} ; see the references above for a thorough exposition.

To get a better feel for this probabilistic way of thinking, we take advantage of the stochastic representation from Theorem 6.2 to construct an explicit solution formula in the case of two sites. When evaluating the expectation, we distinguish two cases. Either, the two sites have not been separated until generation t , which happens with probability r_1^t . In this case, both sites have the same ancestor who comes, with probability $(M^t)(\alpha, \beta)$, from location β . Hence, in this case, $\mu_t(\alpha) = (M^t \mu_0)(\alpha)$. If, on the other hand, the sites *have* been separated, we denote by σ the smallest t such that $|\Sigma_t| = 2$. In this case, the letters come from two different parents. Their origins are determined by performing independent random walks on L for the remaining time $t - \sigma + 1$. Summing over all possible values for σ and the label of the block at the time of splitting (which is γ with probability $(M^{\sigma-1})(\alpha, \gamma)$), we see that

$$\begin{aligned} \mu_t(\alpha) &= r_1^t (M^t \mu_0)(\alpha) \\ &+ \sum_{\gamma \in L} \sum_{\sigma=1}^t r_1^{\sigma-1} r_0 (M^{\sigma-1})(\alpha, \gamma) (M^{t-\sigma+1} \mu_0)^{\{1\}}(\gamma) \otimes (M^{t-\sigma+1} \mu_0)^{\{2\}}(\gamma). \end{aligned} \tag{8}$$

In the case without migration (i.e. when ignoring the labels), this genealogical process is a variant of the *ancestral recombination graph* (Hudson 1983; Griffiths and Marjoram 1996, 1997; Bhaskar and Song 2012), which was used by Baake and Baake (2016) to solve the recombination equation; see also Baake and Baake (2020). More precisely, the *unlabelled* partitioning process Σ is simply the base of Σ . Likewise, the transition matrix T^{ul} of Σ is obtained from T by marginalising over the labels. Thus, T^{ul} has the entries

$$T_{\delta\varepsilon}^{\text{ul}} = \begin{cases} 0, & \text{if } \varepsilon \not\preceq \delta, \\ \prod_{d \in \delta} r_{\varepsilon|d}^d, & \text{if } \varepsilon \preceq \delta, \end{cases} \tag{9}$$

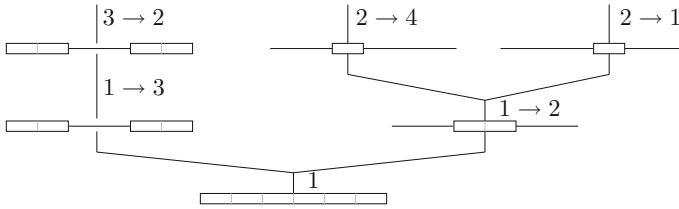


Fig. 2 An illustration of the LPP starting from $\mathbf{1}^1$, the trivial partition consisting of a single block with label 1; the set of locations is $L = \{1, 2, 3, 4\}$. Backward time runs from bottom to top. In each generation, the blocks of the partition are first subject to individual splitting and we trace back the ancestral lines that belong to each fragment; compare Remarks 4.5 and 3.4. The fragments provided by each ancestor are labelled with their locations; we write $\alpha \rightarrow \beta$ to indicate migration from α to β . Recall that in the forward-time model, recombination occurs after migration. Thus, when looking backward in time, splitting due to recombination occurs before the reassignment of the label due to migration. In particular, the first event in this example is a splitting of our sequence located in deme 1

and the transition rates for the LPP factorise as

$$T_{\delta \varepsilon} = T_{\delta \varepsilon}^{\text{ul}} \prod_{(d, \lambda) \in \delta} \prod_{(e, \gamma) \in \varepsilon | d} M(\lambda, \gamma),$$

compare Lemma 4.4. Note that Σ is a process of progressive refinement, which never returns to a state $\not\prec$ the current state. This is due to the absence of coalescence events in the law of large numbers regime, which means that the ancestral recombination graph is actually a tree.

Remark 6.3 The LPP can be interpreted as a multitype branching random walk (BRW) on L , with the types given by the subsets of $[n]$. The particles move according to the transition kernel M , and, as evident from the product structure of the transitions in Eq. (9), undergo independent branching that is the same at every location; each individual of type d branches with probability r_ε^d into $|\varepsilon|$ individuals of types $e_1, \dots, e_{|\varepsilon|}$, where $\varepsilon = \{e_1, \dots, e_{|\varepsilon|}\}$.

7 Limiting and quasi-limiting behaviour of the LPP

We assume now that M is primitive (that is, irreducible and aperiodic), which guarantees the existence of and convergence to a unique stable stationary distribution $q = (q(\alpha))_{\alpha \in L} \in \mathbb{R}^L$ such that

$$q^\top = q^\top M, \tag{10}$$

where \top denotes transpose.

We also assume that

$$\bigwedge \{\delta \in \mathbb{S}([n]) : r_\delta > 0\} = \underline{0}. \tag{11}$$

That is, the coarsest common refinement of all partitions with positive recombination probability is the minimal partition $\underline{0}$ of $[n]$ into singletons. This is only a matter of technical convenience; otherwise, we could simply consider as a single site any set of

sites that are not separated by any partition δ with $r_\delta > 0$. Note that Eq. (11) implies that \underline{Q} is the unique absorbing state of the unlabelled partitioning process. We can now explicitly state the asymptotic behaviour of the MRE (4).

Theorem 7.1 *Under the above assumptions, one has*

$$\lim_{t \rightarrow \infty} \mu_t = \mu_\infty = (\mu_\infty(\alpha))_{\alpha \in L},$$

where

$$\mu_\infty(\alpha) = \bigotimes_{i=1}^n \mu_\infty^{\{i\}}(\alpha) \tag{12}$$

and

$$\mu_\infty^{\{i\}}(\alpha) := \sum_{\beta \in L} q(\beta) \mu_0^{\{i\}}(\beta) \tag{13}$$

for all $\alpha \in L$. The convergence is geometric, i.e. there is a $\gamma \in (0, 1)$ such that

$$\mu_t = \mu_\infty + \mathcal{O}(\gamma^t)$$

as $t \rightarrow \infty$, uniformly in μ_0 .

This is in line with Bürger (2009, Theorem 3.1), which states that the solution of the MRE (4) approaches (at a uniform geometric rate) the submanifold defined by spatial stationarity and linkage equilibrium. Spatial stationarity means that

$$\mu(\alpha) = \sum_{\beta \in L} q(\beta) \mu(\beta)$$

with q of (10); and, under the assumption (11), linkage equilibrium means that $\mu(\alpha)$ is the product of its one-dimensional marginals, as in Eq. (12). However, like the explicit time evolution in Theorem 5.1, the explicit expression in Eq. (13) seems to be new.

In view of Theorem 6.2, this result is highly plausible: almost surely (at a uniform geometric rate), the partitioning process will enter its unique absorbing state where all blocks are singletons. In the sequel, independent migration processes will, for each block, converge to the unique stationary distribution q , again at a geometric rate and uniformly in the initial distribution. This behaviour is also clear in terms of the BRW picture. At some point, the type of each particle is a singleton, whence the particles stop branching and just keep performing independent random walks; see Remark 6.3.

For the formal proof, note that the uniform convergence of the migration processes follows directly from the primitivity of M via standard theory (Karlin and Taylor 1975, Appendix, Thm. 2.3). That the partitioning process enters its absorbing state at a uniform geometric rate is the content of the following lemma.

Lemma 7.2 *Let*

$$\eta := \max_{\delta \in \mathbb{S}(\{n\}) \setminus \{\emptyset\}} T_{\delta\delta}^{\text{ul}} < 1$$

be the maximal sojourn probability of the unlabelled partitioning process and let

$$\tau := \min\{t \in \mathbb{N}_0 : \Sigma_t = \mathbb{0}\}$$

be its time to absorption. Then, uniformly in the initial distribution,

$$\mathbb{P}(\tau > t) = \mathcal{O}((\eta + \varepsilon)^t)$$

for any $\varepsilon > 0$ as $t \rightarrow \infty$.

Proof Since the state space is finite and the partitioning process never returns to a state \neq the current state, this Markov chain may jump at most a finite number of times, say m times, before it is absorbed in $\mathbb{0}$. Thus, for any fixed $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}(\tau > t) &\leq \mathbb{P}(\text{the chain has performed at most } m \text{ jumps up to time } t) \\ &\leq \sum_{j=0}^m \binom{t}{j} (1 - \eta)^j \eta^{t-j} \\ &\leq \sum_{j=0}^m \left(\frac{1 - \eta}{\eta}\right)^j t^m \eta^t = C' t^m \eta^t \leq C \eta^t \left(\frac{\eta + \varepsilon}{\eta}\right)^t = C(\eta + \varepsilon)^t, \end{aligned}$$

where $C' = \sum_{j=0}^m \left(\frac{1-\eta}{\eta}\right)^j$ and C is sufficiently large. □

Next, we investigate the asymptotic behaviour of the LPP.

Proposition 7.3 *There exists a $\gamma \in (0, 1)$ such that*

$$\mathbb{P}(\Sigma_t = \{(\{1\}, \alpha_1), \dots, (\{n\}, \alpha_n)\}) = \prod_{i=1}^n q(\alpha_i) + \mathcal{O}(\gamma^t)$$

as $t \rightarrow \infty$, uniformly in $\alpha_1, \dots, \alpha_n \in L$ and the initial distribution of the LPP. For $\delta \in \mathbb{LS}(\{n\})$ with $\delta \neq \mathbb{0}$,

$$\mathbb{P}(\Sigma_t = \delta) = \mathcal{O}((\eta + \varepsilon)^t),$$

for all $\varepsilon > 0$, again uniformly in the initial distribution.

Proof Let τ be as in Lemma 7.2. The second statement follows immediately from Lemma 7.2 by noting that

$$\mathbb{P}(\Sigma_t = \delta) \leq \mathbb{P}(\tau > t).$$

Now, assume that δ is of the form

$$\delta = \{(\{1\}, \alpha_1), \dots, (\{n\}, \alpha_n)\}.$$

Then, for all $\gamma_1 > \eta$,

$$\begin{aligned} \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\delta}) &= \mathbb{P}\left(\boldsymbol{\Sigma}_t = \boldsymbol{\delta} \mid \tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right) \mathbb{P}\left(\tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right) \\ &\quad + \mathbb{P}\left(\boldsymbol{\Sigma}_t = \boldsymbol{\delta} \mid \tau > \left\lfloor \frac{t}{2} \right\rfloor\right) \mathbb{P}\left(\tau > \left\lfloor \frac{t}{2} \right\rfloor\right) \\ &= \mathbb{P}\left(\boldsymbol{\Sigma}_t = \boldsymbol{\delta} \mid \tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right) + \mathcal{O}(\gamma_1^t) \end{aligned} \tag{14}$$

as $t \rightarrow \infty$, where the last step follows by an application of Lemma 7.2. Furthermore,

$$\begin{aligned} \mathbb{P}\left(\boldsymbol{\Sigma}_t = \boldsymbol{\delta} \mid \tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right) &= \mathbb{P}\left(\Lambda_t^{(i)} = \alpha_i \text{ for all } 1 \leq i \leq n \mid \tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right) \\ &= \prod_{i=1}^n \mathbb{P}\left(\Lambda_t^{(i)} = \alpha_i \mid \tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right). \end{aligned} \tag{15}$$

Here, the $(\Lambda_t^{(i)})_{t \in \mathbb{N}_{\geq \tau}}$ for $i \in L$ are the labels of the (singleton) blocks from time τ onwards; they are independent L -valued Markov chains with transition matrix M . By standard theory, we can be sure that, regardless of the initial value, there is a $\gamma_2 \in (0, 1)$ such that

$$\mathbb{P}\left(\Lambda_t^{(i)} = \alpha_i \mid \tau \leq \left\lfloor \frac{t}{2} \right\rfloor\right) = q(\alpha_i) + \mathcal{O}(\gamma_2^t),$$

uniformly in α_i . Combining this with Eqs. (14) and (15) proves the theorem. □

Proof of Theorem 7.1 By Theorem 6.2, Proposition 7.3, and Definition 3.2, we have for some $\gamma \in (0, 1)$, independent of μ_0 ,

$$\begin{aligned} \mu_t(\alpha) &= \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \mathbf{1}^\alpha] \\ &= \mathcal{O}(\gamma^t) + \sum_{\beta_1, \dots, \beta_n \in L} \left(\prod_{i=1}^n q(\beta_i) \right) \\ &\quad \times \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \mathbf{1}^\alpha, \boldsymbol{\Sigma}_t = \{(\{1\}, \beta_1), \dots, (\{n\}, \beta_n)\}] \\ &= \sum_{\beta_1, \dots, \beta_n \in L} \bigotimes_{i=1}^n q(\beta_i) \mu_0^{[i]}(\beta_i) + \mathcal{O}(\gamma^t) \\ &= \bigotimes_{i=1}^n \sum_{\beta \in L} q(\beta) \mu_0^{[i]}(\beta) + \mathcal{O}(\gamma^t) \\ &= \bigotimes_{i=1}^n \mu_\infty^{[i]}(\alpha) + \mathcal{O}(\gamma^t) = \mu_\infty(\alpha) + \mathcal{O}(\gamma^t) \end{aligned}$$

□

Since the asymptotic behaviour of the LPP is so simple, we now go one step further and inquire about its *quasi*-limiting behaviour; that is, its asymptotic behaviour, conditioned on non-absorption of its base. Generally speaking, quasi-limiting distributions describe the first-order approximation of the deviation from the stationary behaviour. Recall that the partitioning process (labelled or unlabelled) is a process of progressive refinement, and never returns to a state coarser than the current state. This is very different from the situation considered by Collet et al. (2013), where the focus is on irreducible chains.

Unlike the limiting distribution, the quasi-limiting distribution will generally depend on the initial distribution. For convenience of notation, we let the LPP start from a maximal labelled partition $\mathbf{1}^\alpha$, consisting of a single block with label α . However, the following discussion can easily be adapted to the more general setting. In what follows, we will exclude the pathological case of $r_0 = 1$, where the probability of non-absorption is zero, and the conditional distribution we are interested in is not well defined.

We start by recalling the quasi-limiting behaviour of Σ , which was already investigated by Martínez (2017). We posit throughout that $\Sigma_0 = \mathbf{1}$. To state the result, we need some additional notation. First, we define the set of states

$$\mathbb{S}^\downarrow([n]) := \{\delta \in \mathbb{S}([n]) : \exists \ell \in \mathbb{N} \text{ s.t. } ((T^{\text{ul}})^\ell)_{\mathbf{1}_\delta} > 0\}$$

that are *reachable* by Σ when starting in $\mathbf{1}$. As before, η denotes the maximal sojourn probability of Σ (compare Lemma 7.2). We will also need the set

$$\mathcal{F} := \{\delta \in \mathbb{S}^\downarrow([n]) : T_{\delta\delta}^{\text{ul}} = \eta\}$$

of reachable states with maximal sojourn probability. Note that our assumption $r_0 \neq 1$ guarantees that $\eta > 0$. Finally, we define the *first hitting time* of any given $\delta \in \mathbb{S}([n])$,

$$\tau_\delta := \min\{t \in \mathbb{N}_0 : \Sigma_t = \delta\},$$

we write $\tau_{\mathcal{F}} := \min_{\delta \in \mathcal{F}} \tau_\delta$ for the first hitting time of \mathcal{F} , and, as before, $\tau = \tau_0$ for the time to absorption. The following result is known; see Martínez (2017, Theorem 5.5).

Theorem 7.4 *For all $\delta \in \mathcal{F}$, one has*

$$0 < \mathbb{E}[\eta^{-\tau_\delta}; \tau_\delta < \infty] \leq \mathbb{E}[\eta^{-\tau_{\mathcal{F}}}; \tau_{\mathcal{F}} < \infty] < \infty.$$

For all $\delta \in \mathbb{S}([n])$, the limit

$$\mathbb{P}_{\text{qlim}}^\Sigma(\delta) := \lim_{t \rightarrow \infty} \mathbb{P}(\Sigma_t = \delta \mid \tau > t)$$

exists and is given by

$$\mathbb{P}_{\text{qlim}}^\Sigma(\delta) = \frac{\mathbb{E}[\eta^{-\tau_\delta}; \tau_\delta < \infty]}{\mathbb{E}[\eta^{-\tau_{\mathcal{F}}}; \tau_{\mathcal{F}} < \infty]} \mathbb{1}_{\delta \in \mathcal{F}}.$$

Thus defined, $\mathbb{P}_{\text{qlim}}^\Sigma$ is a probability measure on $\mathbb{S}([n])$, called the quasi-limiting distribution of Σ (starting from $\underline{1}$).

Recall that the labels of the different blocks evolve conditionally independently. Thus, we expect the quasi-limiting distribution of the LPP to be similar to the quasi-limiting distribution from Theorem 7.4, garnished with the stationary distribution q of the migration process. To be more explicit, we will prove the following result.

Theorem 7.5 For all $\delta \in \mathbb{LS}([n])$,

$$\lim_{t \rightarrow \infty} \mathbb{P}(\Sigma_t = \delta \mid \tau > t) = \left(\prod_{(d,\lambda) \in \delta} q(\lambda) \right) \mathbb{P}_{\text{qlim}}^\Sigma(\delta),$$

where q is the unique stationary distribution (10) of the migration process.

Remark 7.6 In Theorem 7.1, we have approximated the solution of the MRE (4) by using Proposition 7.3 to approximate the distribution of the labelled partitioning process by its limiting distribution. We can try to improve on this rather coarse estimate by also taking into account the quasi-limiting distribution; at least in principle, the disintegration

$$\mathbb{P}(\Sigma_t = \delta) = \mathbb{P}(\Sigma_t = \delta \mid \tau \leq t) \mathbb{P}(\tau \leq t) + \mathbb{P}(\Sigma_t = \delta \mid \tau > t) \mathbb{P}(\tau > t)$$

allows us to express the error term in Theorem 7.1 via the quasi-limiting distribution, at least when migration is strong compared to recombination. Acquiring precise asymptotics, however, would require more detailed knowledge about the probability $\mathbb{P}(\tau > t)$ and the rate of convergence of the conditional distribution $\mathbb{P}(\Sigma_t = \delta \mid \tau > t)$ to the quasi-limiting distribution.

At the heart of the proof is the observation that further refinement of any $\delta \in \mathcal{F}$ immediately leads to absorption; this was also one of the crucial ingredients in the proof of Theorem 7.4, see Martínez (2017, Theorem 5.5) for the original reference.²

Lemma 7.7 For all $\delta \in \mathcal{F}$, we have

$$T_{\delta\delta}^{\text{ul}} + T_{\delta\underline{0}}^{\text{ul}} = 1. \tag{16}$$

Proof We show that, for all $\delta \in \mathbb{S}^\downarrow([n])$ with $T_{\delta\delta}^{\text{ul}} + T_{\delta\underline{0}}^{\text{ul}} \neq 1$, one has $\delta \notin \mathcal{F}$. Indeed, for any such δ , there is an $\varepsilon \notin \{\underline{0}, \delta\}$ with $T_{\delta\varepsilon}^{\text{ul}} > 0$. For any such ε , there is at least one block $e \in \varepsilon$ with $|e| > 1$. For any such e , the partition

$$\varepsilon' := \{e\} \cup \{\{i\} : i \in [n] \setminus e\} \prec \delta$$

² This result is also implicit in Lemma 5.1 of the corresponding corrigendum; for completeness, we present its proof.

is reachable by Assumption (11) (with $[n]$ replaced by individual blocks of δ). We then have

$$T_{\varepsilon'\varepsilon'}^{\text{ul}} = r_{\{e\}}^e > r_{\{\tilde{d}\}}^{\tilde{d}} \prod_{\substack{\tilde{d} \in \delta \\ d \neq \tilde{d}, |d| > 1}} r_{\{d\}}^d = \prod_{d \in \delta} r_{\{d\}}^d = T_{\delta\delta}^{\text{ul}},$$

where \tilde{d} is the block in δ that contains e . The inequality is true since $\varepsilon' < \delta$ implies that either $|e| < |\tilde{d}|$, in which case $r_{\{e\}}^e > r_{\{\tilde{d}\}}^{\tilde{d}}$; or $|\{d \in \delta : |d| > 1\}| > 1$, which entails that the constrained product is not empty (note that $r_{\{d\}}^d < 1$ for d with $|d| > 1$). We have thus proved that $\delta \notin \mathcal{F}$. \square

Remark 7.8 One might be tempted to assume that the sojourn probability is nondecreasing along every path

$$\mathbb{1} \succcurlyeq \delta_1 \succcurlyeq \delta_2 \succcurlyeq \dots \succcurlyeq 0$$

from the maximal partition to the absorbing state. To illustrate that this is not true in general, consider the following setup. Let $n = 4$ and assume the recombination distribution given by $r_0 = \frac{1}{2}$, $r_{\{\{1,2\},\{3,4\}\}} = \frac{1}{10}$, $r_{\mathbb{1}} = \frac{2}{5}$ and $r_\delta = 0$ otherwise. Then, the sojourn probability of the state $\mathbb{1}$ is $r_{\mathbb{1}} = \frac{2}{5}$, while the (finer) state $\{\{1, 2\}, \{3, 4\}\}$ has the smaller sojourn probability

$$r_{\{\{1,2\},\{3,4\}\}} = (1 - r_0)^2 = \frac{1}{4}.$$

\diamond

The idea of the proof of Theorem 7.5 is simple. First, notice that Lemma 7.7 implies that conditional on non-absorption, Σ remains constant after $\tau_{\mathcal{F}}$. From then on, the labels keep on evolving independently according to M , and their distributions converge to q . To make this rigorous, we just need to make sure that $t - \tau_{\mathcal{F}}$ is large enough (conditional on non-absorption). This is the content of the next Lemma.

- Lemma 7.9** (a) *There exists $c > 0$ such that $\mathbb{P}(\tau > t) \geq c\eta^t$ for all $t \in \mathbb{N}$.*
 (b) *Let $\eta' := \max_{\delta \in \mathbb{S}([n]) \setminus (\mathcal{F} \cup \{0\})} T_{\delta\delta}^{\text{ul}}$. Then, for all $\eta'' > \eta'$, there exists $C > 0$ such that $\mathbb{P}(\tau_{\mathcal{F}} \wedge \tau > t) \leq C(\eta'')^t$ for all $t \in \mathbb{N}$.*
 (c) *There is a $\gamma \in (0, 1)$ such that $\lim_{t \rightarrow \infty} \mathbb{P}(\tau_{\mathcal{F}} > \gamma t \mid \tau > t) = 0$.*

Proof First, we show (a). By definition, $\mathcal{F} \subseteq \mathbb{S}^\downarrow(I)$. Thus, there exists a $t_0 \in \mathbb{N}$ such that $\mathbb{P}(\tau_{\mathcal{F}} = t_0) > 0$. Then, we have for all $t \geq t_0$ that

$$\begin{aligned} \mathbb{P}(\tau > t) &\geq \mathbb{P}(\tau > t, \tau_{\mathcal{F}} = t_0) \\ &= \mathbb{P}(\tau > t \mid \tau_{\mathcal{F}} = t_0) \mathbb{P}(\tau_{\mathcal{F}} = t_0) = c' \eta^{t-t_0} = (c' \eta^{-t_0}) \eta^t \end{aligned}$$

with $c' = \mathbb{P}(\tau_{\mathcal{F}} = t_0)$. Note that we used Lemma 7.7 in the second-last step. Now, simply choose

$$c := \min \left\{ \frac{\mathbb{P}(\tau > t)}{\eta^t} : 0 \leq t \leq t_0 \right\} \cup \{c' \eta^{-t_0}\}.$$

For the proof of (b), we couple $(\Sigma_t)_{t \in \mathbb{N}_0}$ to another process $(N_t)_{t \in \mathbb{N}_0}$ with values in $\mathbb{N}_0 \cup \{\infty\}$ and $N_0 = 0$. It evolves as follows. When $\Sigma_{t+1} = \Sigma_t$, then $N_{t+1} := N_t$ and when $\Sigma_{t+1} \in \mathcal{F} \cup \{0\}$, we set $N_{t+1} := \infty$. In all other cases, we perform a Bernoulli experiment with success probability

$$\frac{1 - \eta'}{1 - T_{\Sigma_t \Sigma_t}^{\text{ul}}}.$$

Upon success, we set $N_{t+1} := N_t + 1$; otherwise, $N_{t+1} := N_t$. Note that the marginal $(N_t)_{t \in \mathbb{N}_0}$ of the coupling $(\Sigma_t, N_t)_{t \in \mathbb{N}_0}$ stochastically dominates a process that has independent Bernoulli increments with parameter $1 - \eta'$.

As we have argued before, the partitioning process can only jump a finite number of times before hitting either 0 or \mathcal{F} . Thus, there is a positive integer m such that, for all $t \in \mathbb{N}$, $\tau \wedge \tau_{\mathcal{F}} > t$ implies $N_t \leq m$. Thus,

$$\mathbb{P}(\tau \wedge \tau_{\mathcal{F}} > t) \leq \mathbb{P}(N_t \leq m) \leq \sum_{k=0}^m \binom{t}{k} (1 - \eta')^k (\eta')^{t-k} = P(t)(\eta')^t < C(\eta'')^t,$$

where $P(t)$ is a polynomial with degree $\leq m$, and C and η'' are as stated.

Finally, (c) is a straightforward consequence of (a) and (b); first, fix any $\eta'' \in (\eta', \eta)$. Then, choose γ such that $(\eta'')^\gamma < \eta$. □

After these preparations, the proof of Theorem 7.5 is not difficult.

Proof of Theorem 7.5 Choose γ as in (c) of Lemma 7.9. We split

$$\mathbb{P}(\Sigma_t = \delta \mid \tau > t) = \mathbb{P}(\Sigma_t = \delta, \tau_{\mathcal{F}} > \gamma t \mid \tau > t) + \mathbb{P}(\Sigma_t = \delta, \tau_{\mathcal{F}} \leq \gamma t \mid \tau > t).$$

The first probability tends to zero as $t \rightarrow \infty$, due to our choice of γ . The second can be rewritten as

$$\mathbb{P}(\Sigma_t = \delta \mid \tau > t, \tau_{\delta} \leq \gamma t) \mathbb{P}(\Sigma_t = \delta, \tau_{\mathcal{F}} \leq \gamma t \mid \tau > t),$$

where we have used that Lemma 7.7 implies

$$\{\tau > t, \tau_{\mathcal{F}} \leq \gamma t, \Sigma_t = \delta\} = \{\tau > t, \tau_{\delta} \leq \gamma t\}.$$

Here, the second factor converges to $\mathbb{P}_{\text{qlim}}^{\Sigma}(\delta)$ by the choice of γ and Lemma 7.9(c).

Now consider the first factor. Together with $\tau > t$ and Lemma 7.7, $\tau_{\delta} \leq \gamma t$ implies that $\Sigma_s = \delta$ for all s between γt and t . During this period, the labels of the blocks of δ evolve independently, and by the uniform convergence to the stationary distribution q , we obtain

$$\lim_{t \rightarrow \infty} \mathbb{P}(\Sigma_t = \delta \mid \tau > t, \tau_{\delta} \leq \gamma t) = \prod_{(d, \lambda) \in \delta} q(\lambda),$$

which completes the argument. For additional details, see also the proof of Proposition 7.3. □

8 Recombination and migration in continuous time

Let us close by briefly discussing how our results carry over from the discrete-time to the continuous-time setting. To distinguish the notation from the discrete-time setting, we write $\omega = (\omega_t)_{t \geq 0}$ instead of $\mu = (\mu_t)_{t \in \mathbb{N}_0}$. The recombination distribution r is replaced by a collection $\varrho = (\varrho_\delta)_{\delta \in \mathbb{S}([n])}$, respectively, of non-negative recombination rates; for each $\delta \in \mathbb{S}([n])$, each individual is between time t and $t + dt$ replaced by a new offspring that is recombined according to δ , with probability $\varrho_\delta dt$.

Instead of the stochastic backward migration matrix, we use a Markov generator N on $[n]$; between time t and $t + dt$ and for $\alpha \neq \beta$, an individual at location α is, with probability $N(\alpha, \beta) dt$, replaced by an individual from location β ; we assume that this happens independently of recombination.

Putting this together means for the type distribution that we replace $\omega_t(\alpha)$ by the convex combination

$$\left(1 - \sum_{\delta \in \mathbb{S}([n])} \varrho_\delta dt - \sum_{\beta \neq \alpha} N(\alpha, \beta) dt\right) \omega_t + \sum_{\delta \in \mathbb{S}(I)} \varrho_\delta \mathcal{R}_\delta(\omega_t) dt + \sum_{\beta \neq \alpha} N(\alpha, \beta) \omega_t(\beta) dt.$$

In other words,

$$\dot{\omega}_t(\alpha) = \sum_{\beta \in L} N(\alpha, \beta) \omega_t(\beta) + \sum_{\delta \in \mathbb{S}([n])} \varrho_\delta (\mathcal{R}_\delta - \text{Id}) \omega_t(\alpha); \tag{17}$$

note that we have used that $N(\alpha, \alpha) = -\sum_{\beta \neq \alpha} N(\alpha, \beta)$ since N is a Markov generator. The backward view can be easily adapted as follows. Again, we have an LPP (this time in continuous time) $\Sigma^c = (\Sigma_t^c)_{t \geq 0}$. It evolves as follows. At rate ϱ_ε for all ε , each labelled block (d, λ) of Σ_t^c is split into the blocks of the induced partition $\varepsilon|_d$; each of these fragments inherits the label λ . In addition and independently, for every $\alpha \in L$, each block with label α is relabelled β at rate $N(\alpha, \beta)$. Somewhat more formally, Σ^c is a Markov chain in continuous time with generator \mathcal{Q} defined by its nondiagonal elements

$$\mathcal{Q}_{\delta\varepsilon} = \begin{cases} \varrho_{\varepsilon|_d}^d, & \text{if } \varepsilon = (\delta \setminus \{(d, \lambda)\}) \cup \varepsilon|_d \times \{\lambda\} \text{ for some } d \in \delta, \\ N(\alpha, \beta), & \text{if } \varepsilon = (\delta \setminus \{(d, \alpha)\}) \cup \{(d, \beta)\} \text{ for some } d \in \delta, \\ 0, & \text{otherwise,} \end{cases}$$

where the marginal recombination rates are defined in analogy with the marginal recombination probabilities (compare Eq. (7)):

$$\varrho_\varepsilon^d = \sum_{\substack{\varepsilon' \in \mathbb{S}([n]) \\ \varepsilon'|_d = \varepsilon}} \varrho_{\varepsilon'}.$$

Note that, in the case without migration and with recombination restricted to single crossovers, that is, to partitions of the form $\{[1 : i], [i + 1, n]\}$ for some $1 \leq i < n$, the

continuous-time backward dynamics (and thus, by duality, the forward dynamics; see Eq. (18)) has a simple explicit solution, which is due to the fact that crossover events “rain down” on sequences in an independent Poissonian fashion (Baake and Baake 2020). See also Lambert et al. (2020) for the (much more involved) extension to the case with (a small amount of) coalescence in the infinite-sequence limit.

But let us return to the full Eq. (17). As before (compare Theorem 6.2), one can prove the duality relation

$$\mathcal{R}_\delta(\omega_t) = \mathbb{E}[\mathcal{R}_{\Sigma_t^c}(\omega_0) \mid \Sigma_0^c = \delta], \quad (18)$$

whence we obtain the solution

$$\omega_t(\alpha) = \sum_{\delta \in \text{LS}(\{n\})} (e^{tQ})_{1^\alpha \delta} \mathcal{R}_\delta(\omega_0) \quad (19)$$

by solving the associated (linear) Kolmogorov backward equation, in perfect analogy to Theorem 5.1. The duality relation can be proved by a straightforward adaptation of the techniques of Baake and Baake (2016). Indeed, it was shown there that the recombination part of Eq. (17) is dual to the splitting (or branching) part of Σ^c . Showing that the migration part is dual to the random walk defined by N is a standard exercise.

Because Eq. (19) is not very concrete, let us derive a more explicit solution formula for the special case $n = 2$. We give a probabilistic argument, analogous to Eq. (8). First, note that with probability $e^{-\varrho_0 t}$, the sites are not separated until time t , that is, $\Sigma_t^c = 1$; the single block has performed a random walk with transition kernel N for the entire duration t . Hence, in this case, $\omega_t(\alpha) = (e^{tN} \omega_0)(\alpha)$. On the other hand, if the blocks have been split at time $\sigma \in [0, t]$, then both sites have performed independent random walks, starting at time σ at the location γ where the split took place, and the solution is given by $(e^{(t-\sigma)N} \omega_0)^{\{1\}}(\gamma) \otimes (e^{(t-\sigma)N} \omega_0)^{\{2\}}(\gamma)$. Integrating over all possible values for σ (keeping in mind that σ is exponentially distributed with mean $1/\varrho_0$) and γ (keeping in mind that, at the moment of splitting, the block has label γ with probability $(e^{\sigma N})(\alpha, \gamma)$), we obtain

$$\begin{aligned} \omega_t(\alpha) &= e^{-\varrho_0 t} (e^{tN} \omega_0)(\alpha) \\ &+ \varrho_0 \sum_{\gamma \in L} \int_0^t e^{-\varrho_0 \sigma} (e^{\sigma N})(\alpha, \gamma) (e^{(t-\sigma)N} \omega_0)^{\{1\}}(\gamma) \otimes (e^{(t-\sigma)N} \omega_0)^{\{2\}}(\gamma) d\sigma. \end{aligned} \quad (20)$$

For more than two loci, one can proceed in a similar fashion, disintegrating the solution conditional on the waiting time(s) between splitting events. However, this becomes cumbersome very quickly as one has to keep track of various different contributions, corresponding to different realisations of the jump chain of the (unlabelled) partitioning process. In particular, the form of these contributions changes, depending on the tree topology; see Fig. 3.

Remark 8.1 It is straightforward to adapt the partitioning process in the diffusion limit to the setting with migration, both for finite, discrete sequences as treated by Baake

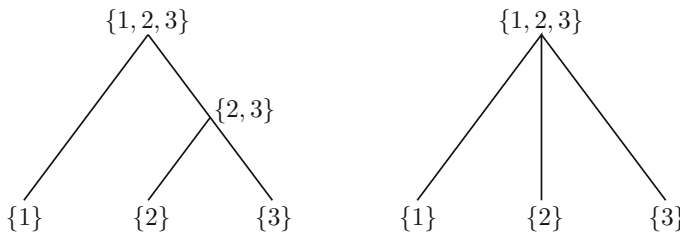


Fig. 3 Two realisations of the jump chain of the unlabelled partitioning process. Contributions to the solution corresponding to the left panel consist of two iterated integrals, while the topology on the right gives rise to a single integral as in Eq. (20). Note also that by permutation of the labels in the left panel, there are 2 additional realisations of the jump chain with the same topology

et al. (2016) as well as continuous sequences; see Lambert et al. (2020). The new feature of this LPP with coalescence is that two blocks can only coalesce if they share the same label. However, an exhaustive treatment would go beyond the scope of this work.

Acknowledgements It is our pleasure to thank Reinhard Bürger for enlightening discussions and two anonymous referees for insightful comments. This work was supported by the German Research Foundation (DFG), within the SPP 1590 (FA) and the CRC 1283, Project C1 (EB); by Grant CMM Basal CONICYT, Project AFB 170001 (SM); and by ANID/Doctorado en el extranjero doctoral scholarship, Grant No. 2018-72190055 (IL). Thanks for hospitality go from IL to the Bielefeld group and from EB to the Santiago group.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baake E, Baake M (2003) An exactly solved model for mutation, recombination and selection. *Can J Math* 55: 3–41 and Erratum 60: 264–265 (2008)
- Baake E, Baake M (2016) Haldane linearisation done right: solving the nonlinear recombination equation the easy way. *Discrete Contin Dyn Syst A* 36:6645–6656
- Baake E, Esser M, Probst S (2016) Partitioning, duality, and linkage disequilibria in the Moran model with recombination. *J Math Biol* 73:161–197
- Baake E, Baake M (2020) Ancestral lines under recombination. In: Baake E, Wakolbinger A (eds) *Probabilistic structures in evolution*. EMS Press, Berlin (**in press**). [arXiv:2002.08658](https://arxiv.org/abs/2002.08658)
- Baake E, Baake M, Salamat M (2016) The general recombination equation in continuous time and its solution. *Discrete Contin Dyn Syst A* 36: 63–95 and Erratum and addendum 36:2365–2366 (2016)
- Baake E, von Wangenheim U (2014) Single-crossover recombination and ancestral recombination trees. *J Math Biol* 68:1371–1402

- Bhaskar A, Song YS (2012) Closed-form asymptotic sampling distributions under the coalescent with recombination for an arbitrary number of loci. *Adv Appl Probab* 44:391–407
- Bürger R (2009) Multilocus selection in subdivided populations I. Convergence properties for weak or strong migration. *J Math Biol* 58:939–978
- Christiansen FB (1999) Population genetics of multiple loci. Wiley, Chichester
- Collet P, Martínez S, San Martín J (2013) Quasi-stationary distributions. Markov chains, diffusions and dynamical systems. Springer, Berlin
- Darroch JN, Seneta E (1965) On quasi-stationary distribution in absorbing discrete-time finite Markov chains. *J Appl Probab* 2:88–100
- Durrett R (2008) Probability models for DNA sequence evolution, 2nd edn. Springer, New York
- Griffiths RC, Marjoram P (1996) Ancestral inference from samples of DNA sequences with recombination. *J Comput Biol* 3:479–502
- Griffiths RC, Marjoram P (1997) An ancestral recombination graph. In: Donnelly P, Tavaré S (eds) *Progress in population genetics and human evolution*. Springer, New York, pp 257–270
- McHale D, Ringwood GA (1983) Haldane linearisation of baric algebras. *J Lond Math Soc* 28:17–26
- Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol* 23:183–201
- Jansen S, Kurt N (2014) On the notion(s) of duality for Markov processes. *Probab Surv* 11:59–120
- Karlin S, Taylor HM (1975) A first course in stochastic processes, 2nd edn. Academic Press, San Diego
- Lambert A, Miró Pina V, Schertzer E (2020) Chromosome painting: how recombination mixes ancestral colors. *Ann Appl Probab* (online first)
- Liggett TM (2010) Continuous time Markov processes: an introduction. American Mathematical Society, Providence
- Lyubich YI (1992) Mathematical structures in population genetics. Springer, Berlin
- Martínez S (2017) A probabilistic analysis of a discrete-time evolution in recombination. *Adv Appl Math* 91:115–136; and Corrigendum 110:403–411 (2019)
- Matsen FA, Wakeley J (2006) Convergence to the island-model coalescent process in populations with restricted migration. *Genetics* 172:701–708
- Nagylaki T (1992) Introduction to theoretical population genetics. Springer, Berlin
- Nagylaki T, Hofbauer J, Brunovský P (1999) Convergence of multilocus systems under weak epistasis or weak selection. *J Math Biol* 38:103–133
- Notohara M (1990) The coalescent and the genealogical process in geographically structured populations. *J Math Biol* 29:59–75
- Slade PF, Wakeley J (2005) The structured ancestral selection graph and the many-demes limit. *Genetics* 169:1117–1131
- von Wangenheim U, Baake E, Baake M (2010) Single-crossover recombination in discrete time. *J Math Biol* 60:727–760

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.