J. Math. Biol. (2016) 73:161–197
DOI 10.1007/s00285-015-0936-6

Mathematical Biology

Partitioning, duality, and linkage disequilibria in the Moran model with recombination

Mareike Esser[1] · **Sebastian Probst**[1] · **Ellen Baake**[1]

Received: 11 February 2015 / Revised: 22 June 2015 / Published online: 6 November 2015

Abstract The multilocus Moran model with recombination is considered, which describes the evolution of the genetic composition of a population under recombination and resampling. We investigate a marginal ancestral recombination process, where each site is sampled only in one individual and we do not make any scaling assumptions in the first place. Following the ancestry of these loci backward in time yields a partition-valued Markov process, which experiences splitting and coalescence. In the diffusion limit, this process turns into a marginalised version of the multilocus ancestral recombination graph. With the help of an inclusion–exclusion principle and so-called recombinators we show that the type distribution corresponding to a given partition may be represented in a systematic way by a sampling function. The same is true of correlation functions (known as linkage disequilibria in genetics) of all orders. We prove that the partitioning process (backward in time) is dual to the Moran population process (forward in time), where the sampling function plays the role of the duality function. This sheds new light on the work of Bobrowski et al. (J Math Biol 61:455–473, 2010). The result also leads to a closed system of ordinary differential equations for the expectations of the sampling functions, which can be translated into expected type distributions and expected linkage disequilibria.

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

Mareike Esser
messer@techfak.uni-bielefeld.de

Sebastian Probst
sprobst@techfak.uni-bielefeld.de

[1] Faculty of Technology, Bielefeld University, Box 100131, 33501 Bielefeld, Germany

Keywords Moran model with recombination · Ancestral recombination process · Linkage disequilibria · Möbius inversion · Duality

Mathematics Subject Classification 92D10 · 60J28

1 Introduction

Models that describe the evolution of finite populations under recombination are among the major challenges in population genetics. This article is devoted to the *Moran model with recombination* (in continuous time), which is briefly described as follows (see Durrett 2008; Bobrowski et al. 2010). A chromosome is identified with a linear arrangement (or *sequence*) of n discrete positions called *sites*, which are collected in the set $S = \{1, \ldots, n\}$. A site may be understood as a nucleotide site or a gene locus. We will throughout consider chromosomes as (haploid) *individuals*, that is, we think at the level of gametes (rather than that of diploid individuals that carry two copies of the genetic information). Site i is occupied by letter $x_i \in \mathbb{X}_i$, where $\mathbb{X}_i$ is a finite set, $1 \leqslant i \leqslant n$. If sites are nucleotide sites, a natural choice for each $\mathbb{X}_i$ is the nucleotide alphabet {A, G, C, T}; if sites are gene loci, $\mathbb{X}_i$ is the set of alleles that can occur at locus i. The genetic type of each individual is thus described by the sequence $x = (x_1, x_2, \ldots, x_n) \in \mathbb{X}_1 \times \cdots \times \mathbb{X}_n =: \mathbb{X}$, where $\mathbb{X}$ is the type space. Recombination means that a new individual is formed as a 'mixture' of an (ordered) pair of parents, say x and y. We will restrict ourselves to *single-crossover recombination*, that is, the offspring inherits the leading segment (up to site i, for some $1 \leqslant i < n$) from the first and the trailing segment (after site i) from the second parent. The recombined type thus is $(x_{\leqslant i}, y_{>i}) := (x_1, \ldots, x_i, y_{i+1}, \ldots, y_n)$; we say that a crossover has happened between sites i and $i+1$. The sites that come from the paternal and the maternal sequence, respectively, define a *partition* $\mathcal{A}$ of S into two parts (we need not keep track of which part was 'maternal' and which was 'paternal'). All partitions of S into two ordered (or contiguous) parts ($\mathcal{A} = \{\{1, 2, \ldots, i\}, \{i+1, \ldots, n\}\}$, $i \in S\backslash\{n\}$) can be realised, via a single crossover event. Altogether, whenever an offspring is created, its sites are partitioned between parents according to $\mathcal{A}$ with probability $r_{\mathcal{A}}$, where $r_{\mathcal{A}} \geqslant 0$, $\sum_{\mathcal{A}\in\mathbb{O}_2(S)} r_{\mathcal{A}} \leqslant 1$, and $\mathbb{O}_2(S)$ is the set of all ordered partitions of S into two parts. Let us note that, due to the one-to-one correspondence between elements of $S\backslash\{n\}$ and those of $\mathbb{O}_2(S)$, the specification of the $r_{\mathcal{A}}$ simply means that a crossover probability is associated with each site in $S\backslash\{n\}$. The sum $\sum_{\mathcal{A}\in\mathbb{O}_2(S)} r_{\mathcal{A}}$ is the probability that some recombination event takes place during reproduction. With probability $r_{\{S\}} = 1 - \sum_{\mathcal{A}\in\mathbb{O}_2(S)} r_{\mathcal{A}}$, there is no recombination, in which case the offspring is the full copy of a single parent. We write $\mathbb{O}_{\leqslant 2}(S) := \mathbb{O}_2(S) \cup \{S\}$ for the set of ordered partitions into at most two parts. The collection $\{r_{\mathcal{A}}\}_{\mathcal{A}\in\mathbb{O}_{\leqslant 2}(S)}$ is known as the *recombination distribution* (Bürger 2000, p. 55).

Consider now a *population* of a constant number of N haploid individuals (that is, gametes), which evolves as follows (see Fig. 1). Each individual has an exponential lifespan with parameter 1 (this choice of the parameter is without loss of generality; it simply sets the time scale). When an individual dies, it is replaced by a new one as follows. First draw a partition $\mathcal{A}$ according to the recombination distribution. Then

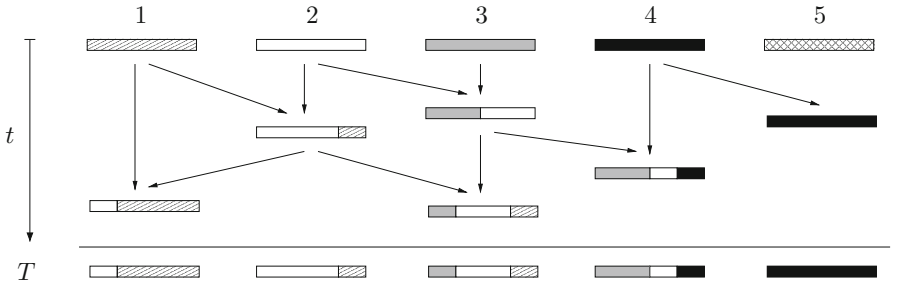

Fig. 1 Snapshot of a Moran model realisation with $N = 5$ individuals. For example, in the first event, individual 3 dies and is replaced by a recombined copy of individuals 2 and 3. The *last line* shows the composition of the population at the final timepoint, T

draw $|\mathcal{A}|$ parents from the population (the parents may include the individual that is about to die), uniformly and with replacement, where $|\mathcal{A}|$ is the number of parts in $\mathcal{A}$. If $|\mathcal{A}| = 2$, the offspring inherits the leading segment of $\mathcal{A}$ from the first and the trailing segment from the second parent, as described above. If $|\mathcal{A}| = 1$ (and thus $\mathcal{A} = \{S\}$), the offspring is a full copy of a single parent (again chosen uniformly from among all individuals); this is called a *(pure) resampling* event. All events are independent of each other.

Note that it may seem biologically more realistic to draw two parents *without* replacement. However, assuming sampling *with* replacement entails significant simplifications, and yields the same process as sampling without replacement with a slight change in the recombination distribution. More precisely, since drawing the same individual twice means that the offspring is a full copy of this single parent, our process agrees (in distribution) with the analogous process without replacement if $r_{\mathcal{A}}$ is replaced by $r_{\mathcal{A}}(N-1)/N$ for all $\mathcal{A} \in \mathbb{O}_2(S)$ (and $r_{\{S\}}$ is set accordingly).

The model will be described more formally later. For now, let us summarise the two main lines of research in this context. On the one hand, there has been considerable interest in how the composition of the population evolves over time, and, in particular, how the correlations between sites (known as linkage disequilibria) will develop; see the overviews in Hein et al. (2005, Chap. 5.4), Durrett (2008, Chap. 3.3 and 8.2), or Wakeley (2009, Chap. 7.2.4). Since there is no mutation, a single type will go to fixation in the long run, that is, the entire population will ultimately consist of this single type. In the absence of recombination, this will be one of the types initially present, and it is well known that the fixation probability for a given type equals its initial frequency. If there is recombination, the type that ultimately wins can also be a newly-composed type, but little is known about the fixation probabilities of the many possible types. The explicit development over time is even more challenging, due to an intricate interplay of resampling and recombination. It is usually approached forward in time, e.g., Ohta and Kimura (1969), Polanska and Kimmel (1999, 2005), Song and Song (2007), Baake and Herms (2008), Durrett (2008, Chap. 8.2), or Bobrowski et al. (2010). In the deterministic limit, which emerges when $N \to \infty$ without rescaling of the $r_{\mathcal{A}}$ or of time, the population is described by a system of ordinary differential

equations, again forward in time. This system has an explicit solution, both for the type distribution and for correlation functions of all orders, for an arbitrary number of sites (Baake and Baake 2003; Baake 2005). This also provides a decent approximation for large but finite populations (Baake and Herms 2008), but dealing appropriately with the stochasticity of finite populations remains a major challenge.

The second line of research is concerned with genealogical aspects and sampling formulae. Here, one starts from a sample taken from the present population and traces back the ancestry of the various segments the individuals are composed of. A major challenge lies in the calculation of the probabilities for the type distribution of a random sample, that is, one aims at so-called *sampling formulae*, see Durrett (2008, Chap. 3.6). These questions are naturally approached backward in time. Usually, one employs the *diffusion (or weak recombination) limit*, that is, time is sped up by a factor of N, followed by $N \to \infty$ such that $Nr_{\mathcal{A}} \to \varrho_{\mathcal{A}}$, $\varrho_{\mathcal{A}}$ a constant, $\mathcal{A} \in \mathbb{O}_2(S)$. Obtaining sampling formulae is tied to the situation in which the population has reached a stationary state; even this case is very hard to treat, and coping with time dependence seems to be hopeless.

The aim of this article is to build a bridge between these two lines of research. We will explore the type distribution and the correlations over time, in the stochastic setting. A starting point will be a recent paper by Bobrowski et al. (2010), who approach this question. Their setting is entirely forward in time, which effectively hides some of the underlying structure. In contrast, we will proceed backward in time and provide a genealogocial approach for the analysis of correlations. The crucial notion in this context will be that of *duality* between the original Moran model forward in time and a suitable ancestral process that follows back the ancestry of selected segments from today's population. This will also shed new light on the results of Bobrowski et al. (2010). In order to keep the approach as general as possible, we will throughout adhere to the original (finite N) model, without taking any limit, but will discuss the various scalings and limits where appropriate.

The paper is organised as follows. In Sect. 2, we start by collecting some important facts about partitions and Möbius functions. We then (Sect. 3) introduce the model more formally and motivate our genealogical approach, which may be considered a marginal version of the usual ancestral process with recombination. In Sect. 4, we describe our ancestral process, which is a partitioning process that keeps track of how the ancestral material is partitioned between individuals. In Sect. 5, we introduce a systematic description via recombinators, which describe the action of recombination on a population and have proved very useful in the deterministic setting. We complement them here by *sampling functions*, which are additionally required for finite populations. The collection of sampling functions will be crucial since it will also serve as *duality function* in Sect. 6, where the duality between the Moran model forward in time and the partitioning process backward in time is proved. This proof, at the same time, yields a differential equation system for the expectations of the sampling functions, which are the building blocks for the linkage disequilibria. In Sect. 7, we apply our results to the cases of two and three sites. We will see that the expected linkage disequilibria (of second and third order) decay exponentially even in the presence of resampling, and identify further linear combinations of expected sampling functions that decay exponentially. For two sites, we also obtain the explicit time course

for the expected composition of the population, and, at the same time, the fixation probabilities of the various types.

2 Preliminaries: partitions and Möbius functions

Working with partitions will be essential to our approach, and we will rely throughout on the powerful concept of *Möbius functions* and *Möbius inversion*. Let us briefly collect the basic notions and standard results; more background material as well as the proofs may be found in Rota (1964), Berge (1971, Chap. 3.2), Aigner (1979, Chap. I, II, IV) and Stanley (1986, Chap. 3).

2.1 Partitions

Let W be a finite, nonempty, totally ordered set, such as a finite subset of $\mathbb{N}$; later, W will be S or a subset thereof. Let $\mathbb{P} = \mathbb{P}(W)$ be the set of partitions of W. We write such a partition as $\mathcal{A} = \{A_1, \ldots, A_m\}$, where $A_j \neq \varnothing$ for all j and $A_j \cap A_k = \varnothing$ for all $j \neq k$ together with $A_1 \cup \cdots \cup A_m = W$. We call A_j a *block (or part)* of $\mathcal{A}$ and $m = |\mathcal{A}|$ is the number of blocks in $\mathcal{A}$.

We say that a partition $\mathcal{A} = \{A_1, \ldots, A_m\}$ of W is *ordered* (or *contiguous*, or an *interval partition*) if every A_j is ordered in W, that is, $A_j = \{x \in W \mid \min A_j \leqslant x \leqslant \max A_j\}$. For example, if $W = \{1, 2, 5, 7, 9\}$, then $\{\{1, 2, 5\}, \{7, 9\}\}$ is ordered, but $\{\{1, 2, 7\}, \{5, 9\}\}$ is not. The set of all ordered partitions of W is denoted by $\mathbb{O}(W)$, the set of all ordered partitions of W *into (exactly) two parts* is $\mathbb{O}_2(W)$, and the set of all ordered partitions of W *into at most two parts* is $\mathbb{O}_{\leqslant 2}(W)$.

For a given partition $\mathcal{A} = \{A_1, \ldots, A_m\}$ of W, let $M := \{1, 2, \ldots, m\} = M(\mathcal{A})$ and, for $J \subseteq M$, we define $\mathcal{A}_J := \{A_j\}_{j \in J}$ and $A_J := \cup_{j \in J} A_j$. $\mathcal{A}_J$ is a partition of A_J. In particular, $\mathcal{A}_M = \mathcal{A}$, $A_M = W$, $\mathcal{A}_{\{j\}} = \{A_j\}$, and $\mathcal{A}_{M \setminus \{j\}} = \mathcal{A} \setminus \{A_j\}$, for any $j \in M$. Note that M depends on $\mathcal{A}$, but we suppress this dependence when there is no risk of confusion. We will throughout abbreviate $J \setminus j := J \setminus \{j\}$ and $J \cup k := J \cup \{k\}$.

The natural ordering relation on $\mathbb{P}(W)$ is denoted by $\preccurlyeq$, where $\mathcal{A} \preccurlyeq \mathcal{B}$ means that $\mathcal{A}$ is a *refinement* of $\mathcal{B}$, that is, every block of $\mathcal{A}$ is a subset of a block of $\mathcal{B}$; equivalently, $\mathcal{B}$ is a *coarsening* of $\mathcal{A}$. $\mathcal{A} \prec \mathcal{B}$ means that $\mathcal{A} \preccurlyeq \mathcal{B}$ and $\mathcal{A} \neq \mathcal{B}$. Together with the resulting partial order, $\mathbb{P}(W)$ is a *poset* and, in particular, a *finite lattice*. $\mathbb{P}(W)$ has a unique *minimal* or *finest* partition, which is denoted as $\mathbf{0} = \{\{x\} \mid x \in W\}$; likewise, there is a unique *maximal* or *coarsest* one, namely $\mathbf{1} = \{W\}$.

When U and V are *disjoint* (finite) sets, two partitions $\mathcal{A} \in \mathbb{P}(U)$ and $\mathcal{B} \in \mathbb{P}(V)$ can be joined into $\mathcal{A} \cup \mathcal{B}$ to form an element of $\mathbb{P}(U \dot\cup V)$. Furthermore, if $U \subseteq W$, a partition $\mathcal{A} \in \mathbb{P}(W)$, with $\mathcal{A} = \{A_1, \ldots, A_m\}$ say, defines a unique partition of U by restriction. The latter is denoted by $\mathcal{A}|_U$, and its parts are precisely all non-empty sets of the form $A_i \cap U$ with $1 \leqslant i \leqslant m$. In particular, $\mathbf{1}|_U$ is the coarsest element in $\mathbb{P}(U)$. For two partitions $\mathcal{A}$ and $\mathcal{B}$, the *least upper bound* will be denoted by $\mathcal{A} \vee \mathcal{B}$, namely the finest partition $\mathcal{C}$ for which $\mathcal{A} \preccurlyeq \mathcal{C}$ and $\mathcal{B} \preccurlyeq \mathcal{C}$. Analogously define the *greatest lower bound* of $\mathcal{A}$ and $\mathcal{B}$ by $\mathcal{A} \wedge \mathcal{B}$.

Example 1 Consider the two partitions $\mathcal{A} = \{\{1, 3, 4\}, \{2, 5\}\}$ and $\mathcal{B} = \{\{1, 4\}, \{2, 3\}, \{5\}\}$ of $W = \{1, \ldots, 5\}$ together with a subset $U = \{1, 2, 4\}$ of W. Then $\mathcal{A} \wedge \mathcal{B} = \{\{1, 4\}, \{2\}, \{3\}, \{5\}\}$, $\mathcal{A} \vee \mathcal{B} = \{\{1, \ldots, 5\}\}$, and $\mathcal{A}|_U = \{\{1, 4\}, \{2\}\}$.

2.2 Möbius functions on the poset of partitions and Möbius inversion

The *Möbius function of a poset* is a general and powerful tool in discrete mathematics. It may be considered as a systematic way of implementing the inclusion–exclusion principle. We rely on it in two contexts in this article: First, we use it to turn sampling without replacement into sampling with replacement, and vice versa. Second, we need it to turn type frequencies into linkage disequilibra.

Refering to Aigner (1979, Prop. 4.6), let us only summarise here that the Möbius function μ is defined for all $\mathcal{A} \preccurlyeq \mathcal{C} \in \mathbb{P}(W)$ via

$$\sum_{\mathcal{A} \preccurlyeq \dot{\mathcal{B}} \preccurlyeq \mathcal{C}} \mu(\mathcal{A}, \mathcal{B}) = \begin{cases} 1, & \mathcal{A} = \mathcal{C}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where the underdot indicates the summation variable. Let $\mathcal{A} \preccurlyeq \mathcal{B} \in \mathbb{P}(W)$, with $m = |\mathcal{B}|$ the number of blocks in $\mathcal{B}$, and n_j the number of blocks of $\mathcal{A}$ within block B_j of $\mathcal{B}$, that is, n_j is the number of blocks in $\mathcal{A}|_{B_j}$, $1 \leqslant j \leqslant m$. The Möbius function of the pair $(\mathcal{A}, \mathcal{B})$ is then given by

$$\mu(\mathcal{A}, \mathcal{B}) = \prod_{j=1}^{m} \mu\big(\mathcal{A}|_{B_j}, \mathbf{1}|_{B_j}\big) = \prod_{j=1}^{m} (-1)^{n_j - 1}(n_j - 1)!\,, \tag{2}$$

see Rota (1964, Sect. 7, Ex. 1) or Berge (1971, Chap. 3.2, Ex. 4). We can now state the fundamental Möbius inversion principle as in Aigner (1979, Prop. 4.18). Let f and g be mappings from $\mathbb{P}(W)$ to $\mathbb{C}$ which are, for all $\mathcal{A} \in \mathbb{P}(W)$, related via

$$g(\mathcal{A}) = \sum_{\dot{\mathcal{B}} \succcurlyeq \mathcal{A}} f(\mathcal{B}). \tag{3}$$

Then, this can be solved for f via the inversion formula

$$f(\mathcal{A}) = \sum_{\dot{\mathcal{B}} \succcurlyeq \mathcal{A}} \mu(\mathcal{A}, \mathcal{B})\, g(\mathcal{B}). \tag{4}$$

More precisely, this is *inversion from above*. An analogous formula applies for *inversion from below*; this relies on refinements rather than coarsenings, with '$\succcurlyeq$' replaced by '$\preccurlyeq$' in (3) and (4). It is important to note that Möbius inversion is not restricted to functions; it also applies to bounded operators.

3 The model and the genealogical approach

In this section, we define the model formally and motivate our genealogical approach.

3.1 The Moran model with single-crossover recombination

We identify the population at time t by a (random) counting measure Z_t on $\mathbb{X}$. Namely, $Z_t(\{x\})$ denotes the number of individuals of type $x \in \mathbb{X}$ at time t, and $Z_t(\mathbb{A}) := \sum_{x\in\mathbb{A}} Z_t(\{x\})$ for $\mathbb{A} \subseteq \mathbb{X}$; we abbreviate $Z_t(\{x\})$ as $Z_t(x)$. If we define δ_x as the point measure on x (i.e., $\delta_x(y) = \delta_{x,y}$ for $x, y \in \mathbb{X}$), we can also write $Z_t = \sum_{x\in X} Z_t(x)\,\delta_x$. Since our Moran population has constant size N, we have $\|Z_t\| = N$ for all times, where $\|Z_t\| := \sum_{x\in\mathbb{X}} Z_t(x) = Z_t(\mathbb{X})$ is the norm (or total variation) of Z_t.

So, $\{Z_t\}_{t\geqslant 0}$ is a Markov process in continuous time with values in

$$E := \big\{z \in \{0,\dots,N\}^{|\mathbb{X}|} \mid \|z\| = N\big\}, \tag{5}$$

where $|\mathbb{X}|$ is the number of elements in $\mathbb{X}$. We will describe the action of recombination on (positive) measures with the help of so-called *recombinators* as introduced by Baake and Baake (2003); see also Baake and Herms (2008) for a pedestrian introduction. Let $\mathcal{M}_+(\mathbb{X})$ be the set of all positive, finite measures on $\mathbb{X}$ and we understand $\mathcal{M}_+(\mathbb{X})$ to include the zero measure. Define the canonical projection $\pi_I : \mathbb{X} \to \bigtimes_{i\in I} \mathbb{X}_i =: \mathbb{X}_I$, for $I \subseteq S = \{1,\dots,n\}$, by $\pi_I(x) = (x_i)_{i\in I}$ as usual. For $\omega \in \mathcal{M}_+(\mathbb{X})$, the shorthand $\pi_I.\omega := \omega \circ \pi_I^{-1}$ indicates the marginal measure with respect to the sites in $I \subseteq S$, where π_I^{-1} is the preimage of π_I. The operation . (where the dot is on the line and should not be confused with a multiplication sign) is known as the *pushforward* of ω w.r.t. π_I. In terms of coordinates, the definition may be spelled out as

$$\big(\pi_I.\omega\big)\big(x_I\big) = \omega \circ \pi_I^{-1}\big(x_I\big) = \omega\big(\{x \in \mathbb{X} \mid \pi_I(x) = x_I\}\big), \quad x_I \in \mathbb{X}_I.$$

Note that $\pi_\varnothing.\omega = \|\omega\|$ and $\pi_S.\omega = \omega$.

Consider now $\mathcal{A} = \{\{1,2,\dots,i\},\{i+1,\dots,n\}\} \in \mathbb{O}_2(S)$ and $\omega \in \mathcal{M}_+(\mathbb{X})\backslash 0$, and define the *projective recombinator* as

$$R^p_{\mathcal{A}}(\omega) := \frac{1}{\|\omega\|^2}\,\big(\pi_{\{1,\dots,i\}}.\omega\big) \otimes \big(\pi_{\{i+1,\dots,n\}}.\omega\big), \tag{6}$$

where $\otimes$ indicates the tensor product (or product measure). Moreover, we set $R^p_{\mathbf{1}}(\omega) := \omega/\|\omega\|$. $R^p_{\mathcal{A}}(\omega)$ is a probability measure for *all* $\omega \in \mathcal{M}_+(\mathbb{X})\backslash 0$, where the zero measure is excluded to make it well-defined. In words, $R^p_{\mathcal{A}}$ turns ω into the (normalised) product measure of its marginals with respect to the blocks in $\mathcal{A}$. Writing out (6) in terms of coordinates gives

$$\begin{aligned}\big(R^p_{\mathcal{A}}(\omega)\big)(x) &= \frac{1}{\|\omega\|^2}\,\big(\pi_{\{1,\dots,i\}}.\omega\big)\big(x_{\{1,\dots,i\}}\big)\,\big(\pi_{\{i+1,\dots,n\}}.\omega\big)\big(x_{\{i+1,\dots,n\}}\big)\\ &= \frac{1}{\|\omega\|^2}\,\omega(x_1,\dots,x_i,*,\dots,*)\,\omega(*,\dots,*,x_{i+1},\dots,x_n),\end{aligned}$$

where $*$ means marginalisation. If $\omega = z$ is the current population, then $R^p_{\mathcal{A}}(z)$ is the type distribution that results when a new individual is created by drawing a leading and (possibly) a trailing segment [as encoded by $\mathcal{A} \in \mathbb{O}_{\leqslant 2}(S)$] from the current population, uniformly and with replacement.

Remark 1 $R^p_{\mathcal{A}}$ is a projective version of the recombinator defined by Baake and Baake (2003); it differs from the latter by a factor of $1/\|\omega\|$. Clearly, both versions agree on the set of probability measures. As we shall see, the projective version is more suitable in the stochastic setting, while the original recombinators are better adapted to the deterministic situation. Since recombinators will only appear in the projective version in this article, we will drop the superscript and the specification 'projective' and call $R_{\mathcal{A}} := R^p_{\mathcal{A}}$ a *recombinator* by slight abuse of language.

In Sect. 5, we will generalise the recombinators and learn more about their probabilistic meaning and mathematical properties. For the moment, let us use them to reformulate the Moran model with recombination in a compact way. Namely, since all individuals die at rate 1, the population loses type-y individuals at rate $Z_t(y)$. Each loss is replaced by a new individual, which is sampled uniformly from $R_{\mathcal{A}}(Z_t)$ with probability $r_{\mathcal{A}}$, $\mathcal{A} \in \mathbb{O}_{\leqslant 2}(S)$. Therefore, when $Z_t = z$, the transition to $z + \delta_x - \delta_y$ occurs with rate

$$\lambda(z;\, y,\, x) := \sum_{\mathcal{A}\in\mathbb{O}_{\leqslant 2}(S)} r_{\mathcal{A}}\big(R_{\mathcal{A}}(z)\big)(x)\, z(y). \tag{7}$$

The summand for $\mathcal{A} = \mathbf{1}$ corresponds to pure resampling, whereas all other summands include recombination. Note that λ includes 'silent transitions' ($x = y$).

Remark 2 We would like to mention that the model may alternatively be formulated in terms of reproducing individuals rather than dying individuals, as follows. Each individual reproduces at rate 1 and picks an $\mathcal{A} \in \mathbb{O}_{\leqslant 2}(S)$ according to the recombination distribution. If $\mathcal{A} \in \mathbb{O}_2$, the reproducing individual contributes the sites in one of the blocks in $\mathcal{A}$ and picks a random partner that contributes the sites in the other block to the offspring. If $\mathcal{A} = \mathbf{1}$, the reproducing individual contributes all sites. The offspring pieced together in this way replaces a uniformly chosen individual from the population. In this formulation, which is closer in spirit to the *deterministic* single-crossover model, offspring of type x are created at rate $N r_{\mathcal{A}}(R_{\mathcal{A}}(Z_t))(x)$ and replace an individual of type y with probability $Z_t(y)/N$. This explains the different normalisation of the original recombinator, whereas the additional factor of $N = \|Z_t\|$ is absorbed in its definition in Baake and Baake (2003). The resulting transition rates, however, are again those in (7). Here, we stay with the formulation that led to (7) in the first place, since it seems more natural for finite populations.

Let us summarise our model as follows:

Definition 1 (*Moran model with single crossovers*) The Moran model with single crossovers is the Markov chain in continuous time $\{Z_t\}_{t\geqslant 0}$ with state space E of (5) and generator matrix Λ with nondiagonal elements

$$\Lambda(z,\, z+w) = \sum_{\substack{x,y\in\mathbb{X}:\\ \delta_x-\delta_y=w}} \lambda(z;\, y,\, x), \quad w \neq 0,$$

for $z \in E$, $w \in E - z$ (where $E - z := \{v \mid z + v \in E\}$) and $\Lambda(z, z) = -\sum_{\substack{v\in E-z:\\ v\neq 0}} \Lambda(z, z + v)$.

3.1.1 Limits of the forward model

Consider now the family of processes $\{Z_t^{(N)}\}_{t\geqslant 0}$, $N = 1, 2, \ldots$, where we add the upper index to indicate dependence on population size. Also consider the normalised version $\{Z_t^{(N)}/N\}_{t\geqslant 0}$; $Z_t^{(N)}/N$ is a random probability measure on $\mathbb{X}$. For $N \to \infty$ and without any rescaling of the $r_{\mathcal{A}}$ or of time, the sequence $\{Z_t^{(N)}\}_{t\geqslant 0}$ converges to the solution of the *deterministic single-crossover equation*

$$\dot{\omega}_t = \sum_{\mathcal{A}\in\mathbb{O}_2(S)} r_{\mathcal{A}}\big(R_{\mathcal{A}}(\omega_t) - \omega_t\big) \tag{8}$$

with initial value ω_0, ω_0 a probability measure, and we assume that $\lim_{N\to\infty} Z_0^{(N)}/N = \omega_0$. This is a *dynamical law of large numbers* and due to Ethier and Kurtz (1986, Thm. 11.2.1). The precise statement as well as the proof are perfectly analogous to Prop. 1 in Baake and Herms (2008), which assumes a slightly different sampling scheme for recombination. We therefore leave out the details here. The deterministic single-crossover equation (8) was investigated by Baake and Baake (2003) and Baake (2005). For comparison, note that, in view of Remark 2, the probability $r_{\mathcal{A}}$ in (8) is multiplied by the unit rate at which each individual reproduces, and this way turns into a recombination *rate*.

The Moran model with recombination also has a well-known diffusion limit, which emerges when $N \to \infty$ under $Nr_{\mathcal{A}} \to \varrho_{\mathcal{A}}$, $\varrho_{\mathcal{A}}$ a constant, $\mathcal{A} \in \mathbb{O}_2(S)$, after a speedup of time by a factor of N. In the case of two loci and two alleles, this goes back to Ohta and Kimura (1969); see also Durrett (2008, Chap. 8.2) for a modern exposition. Two loci with an arbitrary (but finite) number of alleles are treated by Jenkins et al. (2015). This should readily generalise to the case of a finite number of loci with a finite number of alleles, but we do not spell it out here, since we will not draw on the diffusion limit of the forward process later.

3.2 The ancestral recombination process (ARP) and its marginal version

In line with standard population-genetic thinking, we employ a genealogical approach by tracing back the ancestry of (parts of) the genetic material from a population at present that evolved according to the Moran model with single-crossover recombination. The standard genealogical approach for models with recombination is the ancestral recombination graph (ARG) first formulated by Hudson (1983). Today, many different notions of 'ARG' are in use. We stick to the usual convention here that the ARG assumes the diffusion limit. Hudson's original version was for two loci, but multilocus generalisations (Griffiths and Marjoram 1996; Bhaskar and Song 2012) and continuous sequence versions ($n \to \infty$, see, e.g., Durrett 2008, Chap. 3.4) are immediate.

The ARG starts from a sample of individuals from the present population and follows their ancestry backward. When a sequence (or a part of a sequence) experiences a recombination event, it branches into a leading and a trailing segment; when two (parts of) sequences go back to a common ancestor, there is a coalescence event. For overviews see Hein et al. (2005, Chap. 5), Durrett (2008, Chap. 3.4), or Wakeley (2009, Chap. 7.2). Mutation can be independently superimposed on the ARG, but will not be considered in this article. One is then interested in the full information on the sample, namely, the probabilities for all possible type distributions of the sample. The stationary state of the ARG may be characterised by a collection of so-called *sampling recursions*; they may be solved analytically for tiny samples (leading to explicit *sampling formulae*), or numerically for larger ones, see Golding (1984), or Durrett (2008, Chap. 3.6). But feasibility is limited due to the enormous state space, even for small samples. Alternatively, one resorts to computationally intensive Monte-Carlo or importance-sampling methods to simulate the ARG (Griffiths and Marjoram 1996; Wang and Rannala 2008; Jenkins and Griffiths 2011). Recently, Song and coworkers discovered structural properties of the ARG that allow for an efficient combination of analytical and simulation techniques in the regime of *strong recombination* (Jenkins and Song 2010); more precisely, they work in terms of expansions in $1/\varrho$ as $\varrho \to \infty$,

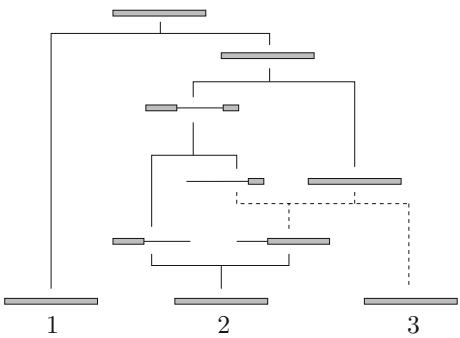

Fig. 2 A realisation of the full ancestral recombination process, starting from $m = 3$ individuals; ancestral material is *shaded*, non-ancestral material is indicated by *thin horizontal lines*. The mixed recombination-coalescence event indicated by *dashed lines* can only appear in the finite population recombination process (ARP). In the diffusion limit, and thus in the ARG, recombination and coalescence act in isolation

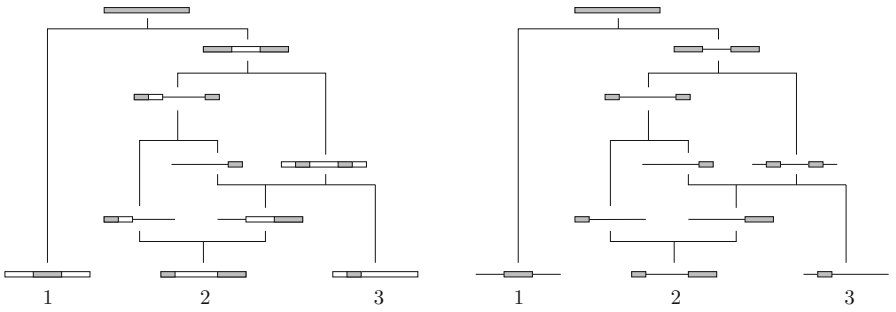

Fig. 3 The marginalised version corresponding to the ARP in Fig. 2, in which we only follow blocks of the partition (*shaded*), that is, block A_i is sampled in individual number i, $1 \leqslant i \leqslant m$. Material that is ancestral to the sampled individuals, but not to the blocks considered, is shown as *open rectangles* (*left*). But since this is not traced back, it can be treated in the same way as material non-ancestral to the sampled individuals (*right*). Consequently, the sample will finally consist of the blocks of the partition only

where all $\varrho_{\mathcal{A}}$, with $\mathcal{A} \in \mathbb{O}_2(S)$, are assumed to scale linearly with the common factor ϱ.

In contrast, we will work in the setting of both *finite n* and *finite N*. The corresponding *ancestral recombination process (ARP)*, which is illustrated in Fig. 2, is a finite-population version of the multilocus ARG. We then simplify matters by only aiming at reduced information. Namely, we consider a partition $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ of S (with $m \leqslant N$). Now sample m individuals from the present population and follow back the ancestry of the sites in A_1 in the first individual, in A_2 in the second individual, ..., in A_m in the m'th individual, without considering any other sites and any other individuals, as in Fig. 3. That is, each locus is considered in one individual only. The result may be viewed as a *marginalised* version of the ancestral recombination process, and, in the diffusion limit, turns into a marginal version of the multilocus ARG starting from a sample of size m. We will see that this information is sufficient to characterise the time evolution of the expected linkage disequilibria of all orders. We will not employ any scaling or limit, in order to allow for arbitrary strengths of recombination. It will turn out that the approach of Bobrowski et al. (2010) actually corresponds to this marginal ancestral recombination process, although this is not apparent from their formulation forward in time.

More precisely, the letters at the loci considered at present, together with their ancestry, can be constructed by a three-step procedure (see Fig. 4). First we run a partitioning process $\{\Sigma_t\}_{t \geqslant 0}$ on $\mathbb{P}(S)$, backward in time, starting at a given initial partition Σ_0 with $|\Sigma_0| = m$. Σ_t describes the partitioning of sites into parental individuals at time t; sites in the same block (in different blocks) belong to the same (to different) individuals. Clearly, $|\Sigma_t|$ is the number of ancestral individuals at time t. The process $\{\Sigma_t\}_{t \geqslant 0}$ is independent of the types and will be described in detail in the next section. In the second step, a letter is assigned to each site of S at time t (i.e. in the past) in the following way. For every part of Σ_t, pick an individual from the initial population (without replacement) and copy its letters to the sites in the block considered. For illustration, also assign a colour to each block, thus indicating different parental individuals. In the last step, the letters and colours are propagated

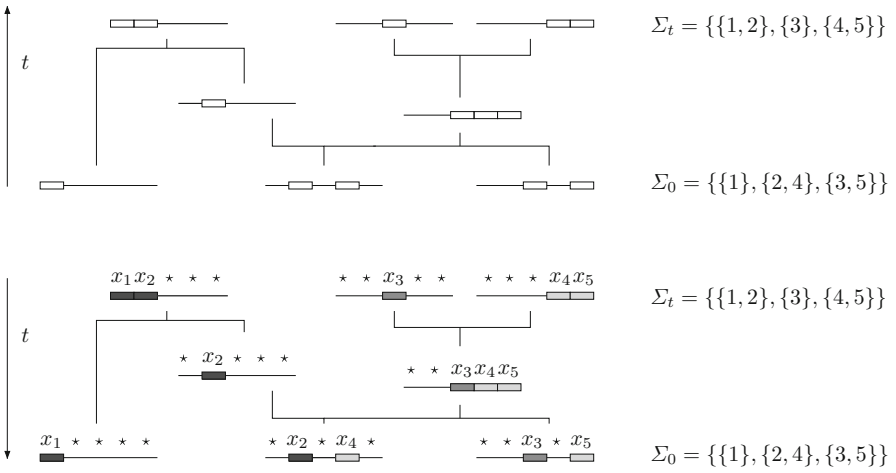

Fig. 4 Construction of one possible ancestry of a collection of sites that correspond to the initial partition $\Sigma_0 = \{\{1\}, \{2, 4\}, \{3, 5\}\}$. The *upper panel* shows the partitioning process (backward in time). In the *lower panel*, *letters* and *colours* are assigned to each block of Σ_t and propagated downward (forward in time)

downward (i.e. forward in time) according to the realisation of $\{\Sigma_t\}_{t\geqslant 0}$ laid down in the first step. A similar construction was used in the ancestral process by Baake and von Wangenheim (2014), but restricted to a sample of size 1 (i.e. start with $\Sigma_0 = \mathbf{1}$), and in discrete time in the deterministic limit. Let us now describe the partitioning process in detail.

4 The partitioning process

The partitioning process $\{\Sigma_t\}_{t\geqslant 0}$ is a Markov process on $\mathbb{P}(S)$, which describes how the sites are partitioned into different individuals backward in time. Since there is a one-to-one relationship between the individuals and the blocks of the partition, we may identify individuals with the ancestral material they carry.

The process $\{\Sigma_t\}_{t\geqslant 0}$ consists of a mixture of splitting (S) and coalescence (C) events. It can be constructed independently of the types. In this section, we describe the process by arguing on the grounds of the underlying Moran model; in Sect. 6, we will formally prove that this is indeed the correct dual process for it.

Since we trace back sites in subsets $U \subseteq S$ (rather than complete sequences), we need the corresponding *marginal recombination probabilities*

$$r_{\mathcal{B}}^{U} := \sum_{\substack{\mathcal{A}\in\mathbb{O}_{\leqslant 2}(S)\\ \mathcal{A}|_U=\mathcal{B}}} r_{\mathcal{A}}^{S} \tag{9}$$

for any $\mathcal{B} \in \mathbb{O}_{\leqslant 2}(U)$, where $r_{\mathcal{A}}^{S} = r_{\mathcal{A}}$. Note that, for $|U| = 1$, the only recombination parameter is $r_{\mathbf{1}}^{U} = 1$. If U is ordered in S (i.e. $U = \{x \in S : \min(U) \leqslant x \leqslant \max(U)\}$)

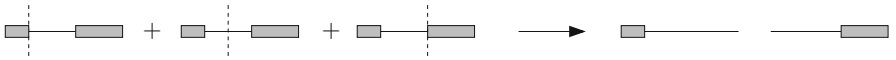

Fig. 5 Let $S = \{1, \ldots, 5\}$ and $U = \{1, 4, 5\} \subset S$. For the partition $\mathcal{B} = \{\{1\}, \{4, 5\}\}$, there are three recombination events that partition U into $\mathcal{B}$, thus $r_{\mathcal{B}}^{U} = r_{\{\{1\},\{2,3,4,5\}\}} + r_{\{\{1,2\},\{3,4,5\}\}} + r_{\{\{1,2,3\},\{4,5\}\}}$

and $\mathcal{B} \neq \mathbf{1}|_U$, then $r_{\mathcal{B}}^{U}$ is simply the probability of crossover after the (unique) site that leads to partition $\mathcal{B}$. If U is not ordered in S, then $r_{\mathcal{B}}^{U}$ is the sum of the probabilities of all crossovers that lead to partition $\mathcal{B}$, as illustrated in Fig. 5.

Assume now that U is an unordered block of Σ_t. This means there is so-called *trapped material*, that is, non-ancestral sites enclosed between ancestral regions. All crossover events within a given trapped segment contribute to the separation of the adjacent ancestral segments—in contrast to crossovers in flanking non-ancestral regions to the left or the right of U, which do not affect the genealogy. Note finally that the upper index in $r_{\mathcal{B}}^{U}$ can, in principle, be omitted since $U = \cup_{i=1}^{|\mathcal{B}|} B_i$, and we will do so when appropriate.

Now start with the initial partition Σ_0. Suppose that the current state is $\Sigma_t = \mathcal{A} = \{A_1, \ldots, A_m\}$ and denote by Δ the waiting time to the next event. Δ is exponentially distributed with parameter m, since each block corresponds to an individual, and each individual is independently affected at rate 1. When the event happens, choose a block uniformly. If A_j is picked, then $\Sigma_{t+\Delta}$ is obtained as follows (see Fig. 6 for an example).

In the splitting step, block A_j turns into an intermediate state $\mathcal{J}$ with probability $r_{\mathcal{J}}^{A_j}$, $\mathcal{J} \in \mathbb{O}_{\leqslant 2}(A_j)$. In detail:

- (S_1) With probability $r_{\mathbf{1}}^{A_j}$, the block A_j remains unchanged. The resulting intermediate state (of this block) is $\mathcal{J} = \mathbf{1}|_{A_j}$.
- (S_2) With probability $r_{\mathcal{J}}^{A_j}$, $\mathcal{J} \in \mathbb{O}_2(A_j)$, block A_j splits into two parts, $\mathcal{J} = \{A_{j_1}, A_{j_2}\}$, which are ordered in A_j, but not necessarily in S. Recall that, via (9), $r_{\mathcal{J}}^{A_j}$ takes into account *all* recombination probabilities that lead to $\mathcal{J}$, including those within trapped material.

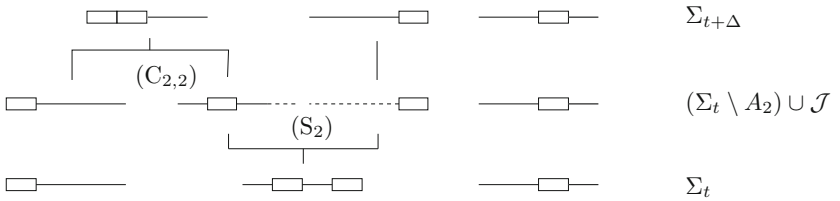

Fig. 6 One step of the partitioning process with current state $\Sigma_t = \{A_1, A_2, A_3\} = \{\{1\}, \{2, 4\}, \{3\}\}$. In this example, A_2 is chosen and splits into $\mathcal{J} = \{\{2\}, \{4\}\}$. In the following step $(C_{2,2})$, the leading part coalesces with A_1, whereas the trailing part remains separate, so that we end up in $\Sigma_{t+\Delta} = \{\{1, 2\}, \{3\}, \{4\}\}$

Now, each block of $\mathcal{J}$ chooses out of N parents, uniformly and with replacement. Among these, there are $m-1$ parents that carry one block of $\mathcal{A}_{M\setminus j} = \mathcal{A}\setminus A_j$ each; the remaining $N-(m-1)$ parents are *empty*, that is, they do not carry ancestral material available for coalescence. Coalescence happens if the choosing block picks a parent that carries ancestral material; otherwise, the choosing block becomes an ancestral block of its own, which is available for coalescence from then onwards. The possible outcomes are certain coarsenings of $\mathcal{A}_{M\setminus j} \cup \mathcal{J}$, namely:
If $\mathcal{J} = \{A_j\}$ [case (S_1)], then either

($C_{1,1}$) With probability $(N-(m-1))/N$, block A_j does not coalesce with any block of $\mathcal{A}_{M\setminus j}$. As a result, $\Sigma_{t+\Delta} = \Sigma_t = \mathcal{A}$.
($C_{1,2}$) With probability $1/N$, block A_j coalesces with block A_k, $k \in M\setminus j$. This results in $\Sigma_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}}$.

If $\mathcal{J} = \{A_{j_1}, A_{j_2}\}$ [case (S_2)], we get the following possibilities:

($C_{2,1}$) With probability $(N-(m-1))(N-m)/N^2$, no block of $\mathcal{J}$ coalesces with a block of $\mathcal{A}_{M\setminus j}$, so $\Sigma_{t+\Delta} = \mathcal{A}_{M\setminus j} \cup \mathcal{J}$.
($C_{2,2}$) With probability $(N-(m-1))/N^2$, one block of $\mathcal{J}$ coalesces with block A_k, $k \in M\setminus j$, while the other block of $\mathcal{J}$ chooses an empty individual. This ends up in the state $\Sigma_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup \{A_{\{j_1,k\}}, A_{j_2}\}$ or $\Sigma_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup \{A_{\{j_2,k\}}, A_{j_1}\}$. That is, in going from Σ_t to $\Sigma_{t+\Delta}$, either block A_{j_1} or A_{j_2} is moved from A_j to A_k.
($C_{2,3}$) With probability $(N-(m-1))/N^2$, the blocks A_{j_1} and A_{j_2} coalesce with each other, but choose an empty individual, which gives $\Sigma_{t+\Delta} = \mathcal{A}$.
($C_{2,4}$) With probability $1/N^2$, the block A_{j_1} coalesces with A_k and A_{j_2} coalesces with A_ℓ, $k, \ell \in M\setminus j$. This yields either $\Sigma_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k,\ell\}} \cup \{A_{\{j_1,k\}}, A_{\{j_2,\ell\}}\}$ if $k \neq \ell$, or $\Sigma_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}}$ if $k = \ell$.

Summarising, we see that a transition from $\mathcal{A}$ to $\mathcal{B}$, via partitioning of block A_j into $\mathcal{J}$, $j \in M$, $\mathcal{J} \in \mathbb{O}_{\leqslant 2}(A_j)$, is possible whenever $\mathcal{B} \succcurlyeq \mathcal{A}_{M\setminus j} \cup \mathcal{J}$ and $\mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}$, or, equivalently, whenever

$$\mathcal{B}|_{A_j} \succcurlyeq \mathcal{J} \quad \text{and} \quad \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}.$$

Each block of $\mathcal{J}$ coalesces into every block currently available with probability $1/N$, and remains separate with probability $(N-k)/N$ if there are currently k blocks available; in the latter case, the block considered becomes number $k+1$. We can therefore summarise the rate of the said transition as

$$\vartheta_{j,\mathcal{J};\mathcal{A},\mathcal{B}} = \begin{cases} r_{\mathcal{J}}^{A_j} \frac{1}{N^{|\mathcal{J}|}} \frac{(N-(m-1))!}{(N-|\mathcal{B}|)!}, & \text{if } \mathcal{B}|_{A_j} \succcurlyeq \mathcal{J},\ \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Note that this includes silent events where $\mathcal{B} = \mathcal{A}$. Thus, the partitioning process $\{\Sigma_t\}_{t\geqslant 0}$ is a continuous-time Markov chain on $\mathbb{P}(S)$ characterised by the generator $\Theta := (\Theta_{\mathcal{A}\mathcal{B}})_{\mathcal{A},\mathcal{B}\in\mathbb{P}(S)}$ with nondiagonal elements

$$\Theta_{\mathcal{A}\mathcal{B}} = \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{O}_{\leqslant 2}(A_j)} \vartheta_{j,\mathcal{J};\mathcal{A},\mathcal{B}}$$

$$= \begin{cases} r_{\mathcal{J}}^{A_j} \frac{1}{N^2} \frac{(N-(m-1))!}{(N-|\mathcal{B}|)!}, & \text{if } \mathcal{B}|_{A_j} = \mathcal{J}, \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}, \\ & \quad \text{for some } j \in M,\ \mathcal{J} \in \mathbb{O}_2(A_j), \\ \frac{2}{N^2} + \frac{N-1}{N^2}\left(r_{\mathbf{1}}^{A_j} + r_{\mathbf{1}}^{A_k}\right), & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}} \\ & \quad \text{for some } j \neq k \in M, \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}, \end{cases} \tag{11}$$

and $\Theta_{\mathcal{A}\mathcal{A}} = -\sum_{\mathcal{B}\in\mathbb{P}(S)\setminus\mathcal{A}} \Theta_{\mathcal{A}\mathcal{B}}$. Note that, for $\mathcal{J} \in \mathbb{O}_2(A_j)$ we have distinguished between $\mathcal{B}|_{A_j} = \mathcal{J}$ and $\mathcal{B}|_{A_j} = \mathbf{1}|_{A_j} \succ \mathcal{J}$. The latter corresponds to $k = \ell$ in ($C_{2,4}$) and leads to the same transition as a *pure coalescence event* in ($C_{1,2}$). More precisely, the total coalescence rate of j and k is

$$\frac{1}{N}\left(r_{\mathbf{1}}^{A_j} + r_{\mathbf{1}}^{A_k}\right) + \frac{1}{N^2}\left(\sum_{\mathcal{J}\in\mathbb{O}_2(A_j)} r_{\mathcal{J}}^{A_j} + \sum_{\mathcal{K}\in\mathbb{O}_2(A_k)} r_{\mathcal{K}}^{A_k}\right) = \frac{2}{N^2} + \frac{N-1}{N^2}\left(r_{\mathbf{1}}^{A_j} + r_{\mathbf{1}}^{A_k}\right)$$

as stated, since $\sum_{\mathcal{J}\in\mathbb{O}_2(U)} r_{\mathcal{J}}^{U} = 1 - r_{\mathbf{1}}^{U}$, $U \subseteq S$. Note also that transitions to partitions $\mathcal{B}$ with $|\mathcal{B}| > N$ are impossible, as it must be.

Remark 3 In fact, this generator Θ coincides with the generator Θ worked out by Bobrowski and Kimmel (2003) and Bobrowski et al. (2010) with a very different approach, forward in time. For $n \leqslant 3$, they state the generator matrices explicitly, and the identity with (11) is easily checked by elementwise comparison. For $n > 3$, they provide an algorithm, which runs through all individuals and all sites and builds up the matrix Θ incrementally, in the following manner. For every given individual, leading and trailing segments (for the split after site i, for all $i \in S\setminus n$) are taken into account, irrespective of whether the segments contain ancestral material. This way, the algorithm does not distinguish between transitions induced by recombination events within ancestral (or trapped) material and recombination events that are invisible in the genealogical perspective, that is, events that are effectively pure coalescence events. Instead, a case distinction is performed that is based on whether or not one or both segments coalesce with individuals that do or do not carry ancestral material. A detailed investigation of this approach, which involves expanding the cases into 11 subcases and rearranging these according to the emerging partitions of the complete ancestral material, leads precisely to our cases ($C_{2,1}$) to ($C_{2,4}$) (here, both emerging segments contain ancestral material) and ($C_{1,1}$) and ($C_{1,2}$) (here one segment is empty). Since this approach somehow disguises or mixes the various partitions of ancestral material that may arise due to a transition, it does not lead to a closed expression for Θ. In contrast, our approach yields the matrix elements explicitly for arbitrary n, and gives them a natural and plausible meaning in terms of the partitioning process in backward time.

4.1 Limits of the partitioning process

We now examine how the partitioning process behaves in the two limiting cases mentioned in Sect. 3, namely, the deterministic limit and the diffusion limit. Recall that, in the deterministic limit, we let $N \to \infty$ without rescaling the recombination probabilities or time. Consider, therefore, the family of processes $\{\Sigma_t^{(N)}\}_{t\geqslant 0}$, $N = 1, 2, \ldots$, generated by $\Theta^{(N)}$, where we again make the dependence on population size explicit through the upper index. In the limit, only the pure splitting events ($C_{2,1}$) survive, more precisely:

Proposition 1 (Deterministic limit) *In the deterministic limit, the sequence of partitioning processes $\{\Sigma_t^{(N)}\}_{t\geqslant 0}$ with initial states $\Sigma_0^{(N)} \equiv \sigma$ converges in distribution to the process $\{\Sigma'_t\}_{t\geqslant 0}$ with initial state $\Sigma'_0 = \sigma$ and generator Θ' defined by its nondiagonal elements*

$$\Theta'_{\mathcal{A}\mathcal{B}} = \begin{cases} r_{\mathcal{J}}^{A_j}, & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus j} \cup \mathcal{J} \text{ for some } j \in M \text{ and } \mathcal{J} \in \mathbb{O}_2(A_j), \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}. \end{cases}$$

Hence, $\{\Sigma'_t\}_{t\geqslant 0}$ is a process of progressive refinements, that is, $\Sigma'_\tau \preccurlyeq \Sigma'_t$ for all $\tau > t$. In particular, if $\Sigma'_0 \in \mathbb{O}(S)$, then $\Sigma'_t \in \mathbb{O}(S)$ for all times.

Proof Inspecting the N-dependence of the elements of $\Theta = \Theta^{(N)}$ in (11) gives the following order of magnitude for the nondiagonal elements:

$$\Theta^{(N)}_{\mathcal{A}\mathcal{B}} = \begin{cases} \frac{1}{N^{m+1-|\mathcal{B}|}}\, r_{\mathcal{J}}^{A_j}\big(1 + \mathcal{O}\big(\frac{1}{N}\big)\big), & \text{if } \mathcal{B}|_{A_j} = \mathcal{J},\ \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j} \\ & \quad \text{for } j \in M\,, \mathcal{J} \in \mathbb{O}_2(A_j), \\ \frac{1}{N}\big(r_{\mathbf{1}}^{A_j} + r_{\mathbf{1}}^{A_k}\big) + \mathcal{O}\big(\frac{1}{N^2}\big), & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}} \\ & \quad \text{for some } j \neq k \in M, \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}. \end{cases} \tag{12}$$

Obviously, $\Theta^{(N)} = \Theta' + \mathcal{O}(1/N)$, which proves convergence of the sequence of generators of $\{\Sigma_t^{(N)}\}_{t\leqslant 0}$ to that of $\{\Sigma'_t\}_{t\geqslant 0}$. This entails convergence of the corresponding sequence of semigroups. With the help of Thms. 4.2.11 and 4.9.10 of Ethier and Kurtz (1986), this guarantees convergence of $\{\Sigma_t^{(N)}\}_{t\geqslant 0}$ to $\{\Sigma'_t\}_{t\geqslant 0}$ in distribution.

The remainder of the statement is obvious since, under Θ', the only transitions are those that involve the refinement of a single block, say A_j, into two blocks ordered in A_j. If Σ'_0 is ordered in S, then all its blocks are ordered in S, and all blocks of Σ'_t will be ordered in S for all times. □

Remark 4 Obviously, in the limit, ancestral material that has been separate will never come together again in one individual, such that there are no coalescence events. When starting with $\Sigma'_0 = \{S\}$, the genealogy may be represented by a binary tree, which

successively branches into smaller segments; for other initial conditions, one gets a corresponding collection (i.e. a forest) of binary trees. We call these trees *ancestral recombination trees* or *ART*s; a discrete-time analogue was studied by Baake and von Wangenheim (2014).

We now turn to the diffusion limit and use the factor N rather than (the more common) $2N$ since our N is the *haploid* population size. Here, one considers a sequence of processes in which time is sped up by a factor of N and the recombination probabilities $r_{\mathcal{A}}$ are rescaled such that $\lim_{N\to\infty} N r_{\mathcal{A}} \to \varrho_{\mathcal{A}}$, $\varrho_{\mathcal{A}}$ a constant, for $\mathcal{A} \in \mathbb{O}_2(S)$; consequently, $r_{\mathbf{1}} \to 1$ as $N \to \infty$. Note that the $\varrho_{\mathcal{A}}$ are *rates* rather than probabilities. The corresponding ARG is the obvious generalisation of Hudson's original ARG to n loci, which we formulate here in our framework for the sake of completeness, as follows. Every ordered pair of lines coalesces at rate 1; every line splits into two at rate $\varrho_{\mathcal{A}}$ for every $\mathcal{A} \in \mathbb{O}_2(S)$, and the ancestral material is distributed between the new lines according to $\mathcal{A}$.

In this formulation, however, certain silent events are included, namely those events that happen in non-ancestral material flanking the ancestral parts. These events do not affect the partitioning of ancestral material and may be removed by working with the marginalised recombination rates instead. That is, if a sequence currently carries a set U of ancestral sites, then the relevant recombination rates (in the diffusion limit) are $\varrho^U_{\mathcal{B}}$, with $\mathcal{B} \in \mathbb{O}_2(U)$, which are defined as in (9) but with r replaced by ϱ. Analogous modifications where the recombination rates depend on the (continuous) region spanned by ancestral material have been investigated by Wiuf and Hein (1997) as well as McVean and Cardin (2005).

If we now restrict attention to the ancestry of n loci partitioned between m individuals, we obtain the *marginal version* of the ARG, which may be formulated as follows.

Definition 2 (*Marginalised n-locus ARG*) Start with the set of n sites distributed across $m \leqslant n$ individuals (or lines) according to a partition Σ''_0 with m parts. Throughout the process, every line is identified with the ancestral material it carries. If it currently carries ancestral sites $U \subseteq S$, it splits into $\mathcal{J} \in \mathbb{O}_2(U)$ at rate $\varrho^U_{\mathcal{J}}$. Every ordered pair of lines coalesces at rate 1, and so do the ancestral sites they carry. That is, the marginalised ARG is the partitioning process $\{\Sigma''_t\}_{t\geqslant 0}$ defined by the generator Θ'' with nondiagonal elements

$$\Theta''_{\mathcal{A}\mathcal{B}} = \begin{cases} \varrho^{A_j}_{\mathcal{J}}, & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus j} \cup \mathcal{J} \text{ for some } j \in M, \mathcal{J} \in \mathbb{O}_2(A_j), \\ 2, & \text{if } \mathcal{B} \succ \mathcal{A} \text{ and } |\mathcal{B}| = |\mathcal{A}| - 1, \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}. \end{cases}$$

Proposition 2 (Diffusion limit of the partitioning process) *In the diffusion limit, the sequence of partitioning processes* $\{\Sigma^{(N)}_{Nt}\}_{t\geqslant 0}$ *with initial states* $\Sigma^{(N)}_0 \equiv \sigma$ *converges in distribution to the process* $\{\Sigma''_t\}_{t\geqslant 0}$ *with initial state* $\Sigma''_0 = \sigma$ *and generator* Θ''.

Proof Due to the rescaling of time, the generator of $\{\Sigma^{(N)}_{Nt}\}_{t\geqslant 0}$ has nondiagonal elements $N\Theta^{(N)}_{\mathcal{A}\mathcal{B}}$. Referring back to (12), they converge to $\lim_{N\to\infty} N\Theta^{(N)}_{\mathcal{A}\mathcal{B}} = \Theta''_{\mathcal{A}\mathcal{B}}$,

since we have $r_{\mathbf{1}}^U \to 1$ and $N r_{\mathcal{J}}^U \to \varrho_{\mathcal{J}}^U$ for $\mathcal{J} \in \mathbb{O}_2(U)$. With the same argument as in the proof of Proposition 1, one obtains convergence in distribution as claimed. □

Remark 5 As was to be expected, only pure splitting events and pure coalescence events survive in the diffusion limit. The 'mixed transitions', which involve both splitting and coalescence (i.e. the dashed lines in Fig. 2) vanish under the rescaling; see also Hein et al. (2005, Fig. 5.11). Let us note that several other variants of the recombination process lead to the same diffusion limit. For example, this is true of the simpler (but biologically less realistic) versions of the continous-time Moran model with recombination where recombination is a parallel process that happens independently of reproduction (rather than coupled to reproduction as assumed here), see Baake and Herms (2008). Even the discrete-time Wright-Fisher model with recombination lies in the domain of attraction of the diffusion limit.

5 Recombinators and sampling functions

In this section, we will have a closer look at three operators associated with recombination and how they are related to each other. We start by generalising our recombinators, then introduce closely related sampling functions and finally multilocus correlation functions, known as linkage disequilibria.

5.1 Recombinators

We have already met $R_{\mathcal{A}}$ for $\mathcal{A} \in \mathbb{O}_{\leqslant 2}(S)$; we now need the generalisation to arbitrary $\mathcal{A} \in \mathbb{P}(S)$. For $\omega \in \mathcal{M}_+(\mathbb{X})\backslash 0$, we first define the *non-normalised recombinator* via

$$\bar{R}_{\mathcal{A}}(\omega) = \bigl(\pi_{A_1}.\omega\bigr) \otimes \cdots \otimes \bigl(\pi_{A_m}.\omega\bigr), \tag{13}$$

where it is implied that the product measure refers to the ordering of the sites as specified by the set S. In words, $\bar{R}_{\mathcal{A}}$ turns ω into the product of its marginals with respect to the blocks in $\mathcal{A}$. We will throughout denote non-normalised mappings by an overbar. Clearly, $\bar{R}_{\varnothing}(\omega) = \|\omega\|$, $\bar{R}_{\mathbf{1}}(\omega) = \omega$ and $\|\bar{R}_{\mathcal{A}}(\omega)\| = \|\omega\|^{|\mathcal{A}|}$. The corresponding normalised version is

$$R_{\mathcal{A}}(\omega) := \frac{\bar{R}_{\mathcal{A}}(\omega)}{\bigl\|\bar{R}_{\mathcal{A}}(\omega)\bigr\|}, \tag{14}$$

which is well-defined since $\omega \neq 0$. Obviously, $R_{\mathcal{A}}(\omega) = \bar{R}_{\mathcal{A}}(\omega/\|\omega\|)$ and $R_{\mathcal{A}}(\omega)$ is a probability measure on $\mathbb{X}$, which coincides with (6) for $\mathcal{A} \in \mathbb{O}_{\leqslant 2}(S)$.

Let us now give a probabilistic interpretation for the case that a recombinator $R_{\mathcal{A}}$ acts on a certain population described by a counting measure $z \in E$. For the moment, attach labels $1, 2, \ldots, N$ to the N individuals in the population, and let these individuals have (random) types $X_t^1, X_t^2, \ldots, X_t^N \in \mathbb{X}$ at time t. The type distribution then is $Z_t = \sum_{k=1}^N \delta_{X_t^k}$. For $U \subseteq S$ and $k \in \{1, \ldots, N\}$, let $X_{t,U}^k := \pi_U(X_t^k)$, and consider the following procedure. Let a partition $\mathcal{A} = \{A_1, \ldots, A_m\}$ of S together

with a collection of labels $\ell = (\ell_1, \ldots, \ell_m) \in \{1, \ldots, N\}^m$ associated with the blocks be given, i.e., $(\mathcal{A}, \ell) := \{(A_1, \ell_1), \ldots, (A_m, \ell_m)\}$. Then, piece together a sequence by taking the sites in A_1 from individual ℓ_1, the sites in A_2 from individual ℓ_2, ... the sites in A_m from individual ℓ_m. The resulting sequence is $X^{\ell}_{t,\mathcal{A}} := (X^{\ell_1}_{t,A_1}, \ldots, X^{\ell_m}_{t,A_m})$. We are now interested in the event

$$\big\{X_{t,\mathcal{A}} = x\big\} := \dot{\bigcup_{\ell \in \{1,\ldots,N\}^m}} \big\{X^{\ell}_{t,\mathcal{A}} = x\big\} \tag{15}$$

and the corresponding counting measure $|\{X_{t,\mathcal{A}} = x\}|$. Eq. (15) is also taken as the definition of the random variable $X_{t,\mathcal{A}}$. Clearly, this counts how often one obtains sequence x when performing the above procedure on a population Z_t and all combinations of individuals are included. Let us emphasise that individuals are combined with replacement, that is, two or more blocks may come from the same individual. Therefore, the event $\{X_{t,\mathcal{A}} = x\}$ may also be understood as the union of the independent events $\{X_{t,A_j} = x_{A_j}\}$, $j \in M$, where

$$\big\{X_{t,A_j} = x_{A_j}\big\} := \dot{\bigcup_{\ell_j \in \{1,\ldots,N\}}} \big\{X^{\ell_j}_{t,A_j} = x_{A_j}\big\}. \tag{16}$$

Therefore,

$$\big|\big\{X_{t,\mathcal{A}} = x\big\}\big| = \prod_{j \in M} \big|\big\{X_{t,A_j} = x_{A_j}\big\}\big| = \prod_{j \in M} \big(\pi_{A_j}.Z_t\big)(x_{A_j}) = \big(\bar{R}_{\mathcal{A}}(Z_t)\big)(x). \tag{17}$$

Clearly, $R_{\mathcal{A}}(Z_t)$, the corresponding normalised version, is the type distribution that results when a sequence is created by taking the letters for the blocks in $\mathcal{A}$ from individuals drawn uniformly and with replacement from the population Z_t. So

$$\big(R_{\mathcal{A}}(z)\big)(x) = \mathbf{P}\big[X_{t,\mathcal{A}} = x \mid Z_t = z\big],$$

where $\mathbf{P}$ denotes probability. Note that the left-hand side depends on time only through the value z of Z_t.

5.2 Sampling function

For $\mathcal{A} \in \mathbb{P}(S)$ and $\omega \in \mathcal{M}_+(\mathbb{X})\backslash 0$, we now define our *sampling function*

$$\bar{H}_{\mathcal{A}}(\omega) := \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \mu(\mathcal{A}, \mathcal{B})\ \bar{R}_{\mathcal{B}}(\omega), \tag{18}$$

where μ is the Möbius function in (2). $\bar{H}_{\mathcal{A}}(\omega)$ is not a positive measure in general; but it will turn out as positive for the important case where $\omega \in E$ with $\|\omega\| \geqslant |\mathcal{A}|$,

see Lemma 1. We will therefore postpone the normalisation step. In any case, Möbius inversion [compare (3) and (4)] immediately yields the inverse of (18):

Fact 1 *For every* $\mathcal{A} \in \mathbb{P}(S)$,

$$\bar{R}_{\mathcal{A}}(\omega) = \sum_{\mathcal{B} \dot{\succcurlyeq} \mathcal{A}} \bar{H}_{\mathcal{B}}(\omega).$$

We can now give $\bar{H}_{\mathcal{A}}$ a meaning by reconsidering the procedure that led to (17) but, this time, individuals are not replaced. That is, for $|\mathcal{A}| \leqslant N$, we now look at the events

$$\{\widetilde{X}_{t,\mathcal{A}} = x\} := \dot{\bigcup_{\substack{\ell \in \{1,\dots,N\}^m \\ \ell_i \neq \ell_j \, \forall \, i \neq j}}} \{X^{\ell}_{t,\mathcal{A}} = x\} \tag{19}$$

and the corresponding counting measure $|\{\widetilde{X}_{t,\mathcal{A}} = x\}|$. Since individuals are not replaced, the events $\{\widetilde{X}_{t,A_j} = x_{A_j}\}$, $j \in M$ [defined as in (16) with X replaced by $\widetilde{X}$] are now *dependent*; an expression for $|\{\widetilde{X}_{t,\mathcal{A}} = x\}|$ analogous to (17) is therefore not immediate. Instead, we resort to an inclusion–exclusion argument and prove

Proposition 3 *For* $\mathcal{A} \in \mathbb{P}(S)$ *with* $|\mathcal{A}| \leqslant N$ *and* $Z_t \in E$, *we have*

$$\big|\{\widetilde{X}_{t,\mathcal{A}} = x\}\big| = \big(\bar{H}_{\mathcal{A}}(Z_t)\big)(x).$$

Proof Fix a given partition $\mathcal{A} \in \mathbb{P}(S)$ with $|\mathcal{A}| = m \leqslant N$. For every $\ell \in \{1, 2, \dots, N\}^m$, the pair $(\mathcal{A}, \ell)$ uniquely defines a pair $(\mathcal{B}, \tilde{\ell})$, where $\tilde{\ell} \in \{\ell \in \{1, 2, \dots, N\}^{|\mathcal{B}|} : \ell_j \neq \ell_k \, \forall \, j \neq k\}$ and $\mathcal{B} \succcurlyeq \mathcal{A}$, as follows. Join all blocks of $\mathcal{A}$ that have the same label, and attach that label to the new block. The result is $(\mathcal{B}, \tilde{\ell})$. The other way round, every $(\mathcal{B}, \tilde{\ell})$ with $\mathcal{B} \succcurlyeq \mathcal{A}$ and $\tilde{\ell} \in \{\ell \in \{1, 2, \dots, N\}^{|\mathcal{B}|} : \ell_j \neq \ell_k \, \forall \, j \neq k\}$ uniquely defines the labelling ℓ of the blocks of $\mathcal{A}$ (keep in mind that $\mathcal{A}$ is fixed): block $A_k \in \mathcal{A}$ receives the label of that block $B_j \in \mathcal{B}$ in which it is contained. We can therefore identify the set $\{(\mathcal{A}, \ell) : \ell \in \{1, 2, \dots, N\}^m\}$ with the set $\bigcup_{\mathcal{B} \dot{\succcurlyeq} \mathcal{A}} \{(\mathcal{B}, \tilde{\ell}) : \tilde{\ell} \in \{\ell \in \{1, 2, \dots, N\}^{|\mathcal{B}|} : \ell_j \neq \ell_k \, \forall \, j \neq k\}\}$. With (17) and (19) in mind, we can therefore *decompose* the event $\{X_{t,\mathcal{A}} = x\} = \dot{\bigcup}_{\mathcal{B} \dot{\succcurlyeq} \mathcal{A}} \{\widetilde{X}_{t,\mathcal{B}} = x\}$, which entails

$$\big|\{X_{t,\mathcal{A}} = x\}\big| = \sum_{\mathcal{B} \dot{\succcurlyeq} \mathcal{A}} \big|\{\widetilde{X}_{t,\mathcal{B}} = x\}\big|.$$

By (17), the left-hand side equals $(\bar{R}_{\mathcal{A}}(Z_t))(x)$. Due to the Möbius inversion principle (applied backward), $|\{\widetilde{X}_{t,\mathcal{B}} = x\}|$ on the right-hand side must equal $(\bar{H}_{\mathcal{B}}(Z_t))(x)$, as claimed. □

Lemma 1 *For $\mathcal{A} \in \mathbb{P}(S)$ with $|\mathcal{A}| = m \leqslant N$ and $z \in E$, $\overline{H}_{\mathcal{A}}(z)$ is a positive measure with*

$$\| \overline{H}_{\mathcal{A}}(z) \| = N\,(N-1)\cdots(N-m+1) > 0.$$

Proof Since, under the given assumptions, $(\overline{H}_{\mathcal{A}}(z))(x) = |\{\widetilde{X}_{t,\mathcal{A}} = x \mid Z_t = z\}| \geqslant 0$ for all x by Proposition 3, it is a positive measure, and its norm can be evaluated via

$$\| \overline{H}_{\mathcal{A}}(z) \| = \sum_{x \in \mathbb{X}} \bigl|\{\widetilde{X}_{t,\mathcal{A}} = x \mid Z_t = z\}\bigr|.$$

By means of (19), this equals the number of possibilities of how to choose m labelled individuals out of N ones *without* replacement, where the order is respected; this is $N\,(N-1)\cdots(N-m+1)$, which is positive since $m \leqslant N$. □

Under the assumptions of Proposition 3, we can therefore define the normalised version of $\overline{H}_{\mathcal{A}}(z)$:

$$H_{\mathcal{A}}(z) := \frac{\overline{H}_{\mathcal{A}}(z)}{\| \overline{H}_{\mathcal{A}}(z) \|} = \frac{(N-m)!}{N!}\; \overline{H}_{\mathcal{A}}(z). \tag{20}$$

$H_{\mathcal{A}}(z)$ is the type distribution that results when a sequence is created by taking the letters for the blocks as encoded by $\mathcal{A}$ from individuals drawn uniformly and *without replacement* from the population z, hence

$$\bigl(H_{\mathcal{A}}(z)\bigr)(x) = \mathbf{P}\bigl[\widetilde{X}_{t,\mathcal{A}} = x \mid Z_t = z\bigr].$$

$H_{\mathcal{A}}(z)$ will later serve as duality function. The situation described here is exactly what happens when a sample is taken in our marginal ancestral recombination process: either the initial sample (according to Σ_0, from the present population Z_t) or the ancestral one (according to Σ_t, from the initial population Z_0)—hence our name *sampling function*. In this light, Fact 1 expresses counting with replacement in terms of counting without replacement, provided ω is a counting measure.

It is also instructive to express the normalised sampling functions in terms of the normalised recombinators. For $z \in E$ and $|\mathcal{A}| \leqslant N$, this gives, via (14),

$$H_{\mathcal{A}}(z) = \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \frac{(N-|\mathcal{A}|)!\, N^{|\mathcal{B}|}}{N!}\, \mu(\mathcal{A}, \mathcal{B})\, R_{\mathcal{B}}(z).$$

Note that $(N-|\mathcal{A}|)!\, N^{|\mathcal{B}|}/N! = \mathcal{O}(N^{|\mathcal{B}|-|\mathcal{A}|})$. This illustrates how the inclusion of coarser partitions yields higher-order correction terms. The other way round, using (14), Fact 1, and (20), one gets

$$R_{\mathcal{A}}(z) = \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \frac{N!}{N^{|\mathcal{A}|}(N-|\mathcal{B}|)!}\, H_{\mathcal{B}}(z). \tag{21}$$

5.3 Restriction to subsystems

Recall that we write the restriction of a measure $\omega \in \mathcal{M}_+(\mathbb{X})$ to a subspace $\mathbb{X}_U := \bigtimes_{i\in U}\mathbb{X}_i$ of $\mathbb{X}$ as $\pi_U.\omega := \omega \circ \pi_U^{-1}$, which corresponds to marginalisation. Clearly we can also define recombinators for any non-empty subset $U \subseteq S$ and any partition $\mathcal{A} = \{A_1, \ldots, A_m\} \in \mathbb{P}(U)$ as $\bar{R}^U_{\mathcal{A}}(\pi_U.\omega)$, in perfect analogy with $\bar{R}^S_{\mathcal{A}}(\omega)$ for $\mathcal{A} \in \mathbb{P}(S)$, which is $\bar{R}_{\mathcal{A}}(\omega)$; and likewise for $R^U_{\mathcal{A}}$, $\bar{H}^U_{\mathcal{A}}$, and $H^U_{\mathcal{A}}$ (if $\omega \neq 0$). For clarity, we sometimes denote the subsystem by a superscript. However, as in the case of the marginal recombination probabilities, it can be dispensed with since $U = \cup_{j=1}^{|\mathcal{A}|} A_j$ if $\mathcal{A} \in \mathbb{P}(U)$. The interpretation in terms of sampling, as well as Fact 1, carry over.

Let us collect some basic properties of recombinators:

Fact 2 *For $\mathcal{A}, \mathcal{B} \in \mathbb{P}(S)$ and $U, V \subseteq S$ with $S = U \dot\cup V$ one has*

(A) $R_{\mathcal{A}} R_{\mathcal{B}} = R_{\mathcal{A}\wedge\mathcal{B}}$.

(B) $\pi_U.R^S_{\mathcal{A}}(\omega) = R^U_{\mathcal{A}|_U}(\pi_U.\omega)$.

(C) *If in addition $\mathcal{A} \preccurlyeq \{U, V\}$, then $\bar{R}^S_{\mathcal{A}} = \bar{R}^U_{\mathcal{A}|_U} \otimes \bar{R}^V_{\mathcal{A}|_V}$. Explicitly, this reads*

$$\bar{R}^S_{\mathcal{A}}(\omega) = \big(\bar{R}^U_{\mathcal{A}|_U} \otimes \bar{R}^V_{\mathcal{A}|_V}\big)(\omega) = \big(\bar{R}^U_{\mathcal{A}|_U}(\pi_U.\omega)\big) \otimes \big(\bar{R}^V_{\mathcal{A}|_V}(\pi_V.\omega)\big).$$

Here and in what follows, we may omit the argument when the meaning is clear.

Proof of Fact 2 Property (A) is Proposition 2 and property (B) is Lemma 1 of Baake et al. (2016) (they both remain true in our normalisation). Property (C) is an obvious generalisation of Proposition 2 of von Wangenheim et al. (2010). It is easily seen by using first property (A), then (13), then (B) and finally (13) once more to give

$$\begin{aligned}\bar{R}^S_{\mathcal{A}}(\omega) &= \bar{R}^S_{\{U,V\}}\big(\bar{R}^S_{\mathcal{A}}(\omega)\big) = \big((\pi_U.\,\bar{R}^S_{\mathcal{A}}) \otimes (\pi_V.\,\bar{R}^S_{\mathcal{A}})\big)(\omega)\\ &= \big(\bar{R}^U_{\mathcal{A}|_U}(\pi_U.\omega)\big) \otimes \big(\bar{R}^V_{\mathcal{A}|_V}(\pi_V.\omega)\big) = \big(\bar{R}^U_{\mathcal{A}|_U} \otimes \bar{R}^V_{\mathcal{A}|_V}\big)(\omega).\end{aligned}$$

□

Let us note a connection between recombination and sampling that will be important in what follows.

Lemma 2 *Let $S = U \dot\cup V$ for two nonempty subsets $U, V \subseteq S$. For two partitions $\mathcal{A} \in \mathbb{P}(U)$, $\mathcal{B} \in \mathbb{P}(V)$, the recombinator and the sampling operator satisfy*

$$\bar{R}^U_{\mathcal{A}} \otimes \bar{H}^V_{\mathcal{B}} = \sum_{\substack{\mathcal{C} \succcurlyeq \mathcal{A}\cup\mathcal{B}\\ \mathcal{C}|_V = \mathcal{B}}} \bar{H}^S_{\mathcal{C}}.$$

Proof Using (18) followed by Fact 2 (C) and Fact 1 we get

$$\bar{R}^U_{\mathcal{A}} \otimes \bar{H}^V_{\mathcal{B}} = \bar{R}^U_{\mathcal{A}} \otimes \Big(\sum_{\mathcal{D} \succcurlyeq \mathcal{B}} \mu(\mathcal{B}, \mathcal{D})\ \bar{R}^V_{\mathcal{D}} \Big) = \sum_{\mathcal{D} \succcurlyeq \mathcal{B}} \mu(\mathcal{B}, \mathcal{D})\ \bar{R}^S_{\mathcal{D} \cup \mathcal{A}}$$
$$= \sum_{\mathcal{D} \succcurlyeq \mathcal{B}} \mu(\mathcal{B}, \mathcal{D}) \sum_{\mathcal{E} \succcurlyeq \mathcal{D} \cup \mathcal{A}} \bar{H}^S_{\mathcal{E}}.$$

Changing the summation order and applying (1) finally leads to

$$\bar{R}^U_{\mathcal{A}} \otimes \bar{H}^V_{\mathcal{B}} = \sum_{\mathcal{C} \succcurlyeq \mathcal{A} \cup \mathcal{B}} \bar{H}^S_{\mathcal{C}} \sum_{\mathcal{B} \preccurlyeq \mathcal{D} \preccurlyeq \mathcal{C}|_V} \mu(\mathcal{B}, \mathcal{D}) = \sum_{\substack{\mathcal{C} \succcurlyeq \mathcal{A} \cup \mathcal{B} \\ \mathcal{C}|_V = \mathcal{B}}} \bar{H}^S_{\mathcal{C}}.$$

□

Remark 6 In a perfectly analogous way, one can show

$$\bar{H}^U_{\mathcal{A}} \otimes \bar{H}^V_{\mathcal{B}} = \sum_{\substack{\mathcal{C} \succcurlyeq \mathcal{A} \cup \mathcal{B} \\ \mathcal{C}|_U = \mathcal{A},\, \mathcal{C}|_V = \mathcal{B}}} \bar{H}^S_{\mathcal{C}}.$$

This illustrates once more that, unlike the $\bar{R}_{\mathcal{A}}$, the $\bar{H}_{\mathcal{A}}$ do *not* have a product structure; this reflects the dependence inherent to drawing without replacement.

5.4 Correlations (or linkage disequilibria)

Linkage disequilibria (LDE) are used in population genetics to quantify the deviation from independence of allele frequencies at the various sites in a sequence. From three sites onwards, many different notions of linkage disequilibria are available in the literature, see Bürger (2000, Chap. V.4.2) for an overview.

We will use as LDEs the general correlation functions, which are widely used in statistical physics, see Dyson (1962) or Mehta (1991, Chap. 5.1.1). This results in an explicit formula for multilocus LDEs for an arbitrary number of sites in terms of sums of products of marginal frequencies, see also Baake and Baake (2003, Appendix) or Gorelick and Laubichler (2004). As we will see, common definitions for two and three sites coincide with ours.

For any given subset $U \subseteq S$ and $\mathcal{A} \in \mathbb{P}(U)$, we first define *correlation operators* as

$$L^U_{\mathcal{A}} = \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} \mu(\mathcal{B}, \mathcal{A})\, R^U_{\mathcal{B}}. \tag{22}$$

Note that the summation is now over all *refinements* of $\mathcal{A}$, in contrast to our sampling functions, which involve all coarsenings of $\mathcal{A}$. The restriction to subsystems stems from the fact that one usually considers deviation from independence on small subsets of S.

The $L^U_{\mathcal{A}}$ have a product structure, $L^U_{\mathcal{A}} = \prod_{j=1}^{|\mathcal{A}|} L_{\mathbf{1}}^{A_j}$, which is obvious from (22) together with the product structure of the recombinators [Fact 2 (C)] and that of the Möbius function (2). Eq. (22) has the inverse

$$R^U_{\mathcal{A}} = \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} L^U_{\mathcal{B}} = \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} \prod_{j=1}^{|\mathcal{B}|} L_{\mathbf{1}}^{B_j}$$

due to *inversion from below* (see Sect. 2). The latter can be reformulated as

$$L^U_{\mathcal{A}} = R^U_{\mathcal{A}} - \sum_{\mathcal{B} \prec \mathcal{A}} \prod_{j=1}^{|\mathcal{B}|} L_{\mathbf{1}}^{B_j}. \tag{23}$$

The case $\mathcal{A} = \mathbf{1}|_U$, $U \subseteq S$, now is of special interest. In line with population-genetics understanding, we define the *multilocus linkage disequilibrium with respect to the sites in* U by letting $L_{\mathbf{1}}^U$ act on the marginal measure $\pi_U.\omega$, $\omega \in \mathcal{M}_+(\mathbb{X})\backslash 0$:

$$L_{\mathbf{1}}^U(\pi_U.\omega) = \sum_{\mathcal{A} \in \mathbb{P}(U)} \mu(\mathcal{A}, \mathbf{1}|_U)\, R^U_{\mathcal{A}}(\pi_U.\omega),$$

cf. (22). Note that $L_{\mathbf{1}}^U(\pi_U.\omega)$ is again a measure on $\pi_U(\mathbb{X})$, but no longer positive in general. With the help of (23), it can be reformulated as

$$L_{\mathbf{1}}^U(\pi_U.\omega) = R_{\mathbf{1}}^U(\pi_U.\omega) - \sum_{\mathcal{B} \prec \mathbf{1}|_U} \prod_{j=1}^{|\mathcal{B}|} L_{\mathbf{1}}^{B_j}(\pi_{B_j}.\omega),$$

which is Eq. (1) in Gorelick and Laubichler (2004). Likewise, this alternative formulation of multilocus LDEs agrees with previous ones from Geiringer (1944), Bennett (1954) and Hastings (1984) up to $|U| \leqslant 3$.

Example 2 For $S = \{1, 2, 3, 4\}$ the LDE with respect to the sites in $U = \{2, 4\}$ reads

$$\begin{aligned} L_{\mathbf{1}}^U(\pi_{\{2,4\}}.\omega)(x) &= R_{\mathbf{1}}^U(\pi_{\{2,4\}}.\omega)(x) - R^U_{\{\{2\},\{4\}\}}(\pi_{\{2,4\}}.\omega)(x) \\ &= \frac{1}{\|\omega\|}\,\omega(*, x_2, *, x_4) - \frac{1}{\|\omega\|^2}\,\omega(*, x_2, *, *)\,\omega(*, *, *, x_4). \end{aligned}$$

Similarly for $U = \{1, 3, 4\}$ we get

$$\begin{aligned} L_{\mathbf{1}}^U(\pi_{\{1,3,4\}}.\omega)(x) = {} & \frac{1}{\|\omega\|}\,\omega(x_1, *, x_3, x_4) - \frac{1}{\|\omega\|^2}\,\omega(x_1, *, *, *)\,\omega(*, *, x_3, x_4) \\ & - \frac{1}{\|\omega\|^2}\,\omega(x_1, *, x_3, *)\,\omega(*, *, *, x_4) \end{aligned}$$

$$- \frac{1}{\|\omega\|^2}\,\omega(x_1,*,*,x_4)\,\omega(*,*,x_3,*)$$
$$+ 2\,\frac{1}{\|\omega\|^3}\,\omega(x_1,*,*,*)\,\omega(*,*,x_3,*)\,\omega(*,*,*,x_4).$$

The correlation operators can also be expressed in terms of our sampling operators. Eqs. (22) and (21), together with a change of the summation order, lead to

$$\begin{aligned} L^U_{\mathcal{A}} &= \sum_{\mathcal{B}\preccurlyeq\mathcal{A}} \mu(\mathcal{B},\mathcal{A}) \sum_{\mathcal{C}\succcurlyeq\mathcal{B}} \frac{N!}{(N-|\mathcal{C}|)!\,N^{|\mathcal{B}|}}\, H^U_{\mathcal{C}} \\ &= \sum_{\mathcal{C}\in\mathbb{P}(U)} H^U_{\mathcal{C}} \sum_{\mathcal{B}\preccurlyeq\mathcal{A}\wedge\mathcal{C}} \frac{N!}{(N-|\mathcal{C}|)!\,N^{|\mathcal{B}|}}\, \mu(\mathcal{B},\mathcal{A}). \end{aligned} \tag{24}$$

For a counting measure $z \in E$ and $U \subseteq S$ with $|U| = k \leqslant 3 \leqslant N$, Eq. (24) yields a particularly nice explicit expression for the LDEs:

$$L^U_{\mathbf{1}}(\pi_U.z) = \frac{N!}{N^k(N-k)!} \sum_{\mathcal{A}\in\mathbb{P}(U)} \mu(\mathcal{A},\mathbf{1}|_U)\, H^U_{\mathcal{A}}(\pi_U.z), \tag{25}$$

as is easily verified. For larger k, the explicit formula gets more involved.

Let us now consider $L^U_{\mathcal{A}}$ for $\mathcal{A} \in \mathbb{P}(U)\backslash\mathbf{1}|_U$. Due to its product structure, the collection of all $L^V_{\mathbf{1}}(\pi_V.\omega)$, $V \subseteq U$, determines all $L^U_{\mathcal{A}}(\pi_U.\omega)$, $\mathcal{A} \in \mathbb{P}(U)$. This is why, for a deterministic ω, the $L^U_{\mathcal{A}}(\pi_U.\omega)$, $\mathcal{A} \neq \mathbf{1}|_U$, are of no particular interest of their own. This changes, however, when ω is random (like Z_t). For we typically do not know the law of Z_t completely; rather, we have access to the expectation of certain functions of Z_t. More precisely, let φ be the law of Z_t and $\mathbf{E}_\varphi$ denote the expectation with respect to φ [that is, for a function f of Z_t, $\mathbf{E}_\varphi[f] = \int f(z)\,\mathrm{d}\varphi(z)$]. It is important to note that the product structure of the recombined measure does not carry over to the expectation. That is, for $\mathcal{A} \in \mathbb{P}(U)$, $\mathbf{E}_\varphi[R^U_{\mathcal{A}}(\pi_U.Z_t)] \neq R^U_{\mathcal{A}}(\mathbf{E}_\varphi[\pi_U.Z_t])$ in general, see the discussion in Baake and Herms (2008); this is indeed a subtle point that sometimes goes wrong, as in Polanska and Kimmel (2005, Eq. (12)), or Bobrowski et al. (2010, pp. 471/472). As a consequence, one also has $\mathbf{E}_\varphi[L^U_{\mathcal{A}}(\pi_U.Z_t)] \neq \prod_{i=1}^{|\mathcal{A}|} L^{A_i}_{\mathbf{1}}(\mathbf{E}_\varphi[\pi_{A_i}.Z_t])$ in general. In the stochastic case, therefore, it is interesting to consider the $L^U_{\mathcal{A}}$ for $\mathcal{A} \neq \mathbf{1}|_U$ as well. The expectations $\mathbf{E}_\varphi[L^U_{\mathcal{A}}(\pi_U.Z_t)]$ contain information on how the mean LDEs in one part of the sequence depend on the mean LDEs in other parts of the sequence. In the next section, we will obtain an ODE system for the $\mathbf{E}_\varphi[H^S_{\mathcal{A}}(Z_t)]$, $\mathcal{A} \in \mathbb{P}(S)$, and these translate into $\mathbf{E}_\varphi[R^S_{\mathcal{A}}(Z_t)]$ and thus into $\mathbf{E}_\varphi[L^S_{\mathcal{A}}(Z_t)]$ via (24). Marginalisation can then be used to calculate the corresponding quantities on $U \subset S$, such as $\mathbf{E}_\varphi[L^U_{\mathcal{A}}(\pi_U.Z_t)]$ for $\mathcal{A} \in \mathbb{P}(U)$.

6 Duality

Duality is a powerful tool to obtain information about one process by studying another, the dual process. The latter may, in an optimal case, have a much smaller state space than the original one. Duality results are essential in interacting particle systems in physics and in population genetics. They are often related to time reversal. The most famous example in population genetics is arguably the moment duality between the Wright–Fisher diffusion forward in time and the block counting process of Kingman's coalescent backward in time (Donnelly 1986; Möhle 2001). Mano (2013) extended this result by incorporating recombination into the two-locus, two-allele case. His results are based on the original version of the ARG and thus on the diffusion limit.

We will briefly explain the general duality concept and then prove that our processes $\{Z_t\}_{t\geqslant 0}$ and $\{\Sigma_t\}_{t\geqslant 0}$ are duals of each other. For the general principle, let $X = \{X_t\}_{t\geqslant 0}$ and $Y = \{Y_t\}_{t\geqslant 0}$ be two Markov processes with state spaces E and F. Define by $M(E \times F)_b$ the set of all bounded measurable functions on $E \times F$. The following definition of duality with respect to a function goes back to Liggett (1985); see also the recent review by Jansen and Kurt (2014).

Definition 3 (*Duality*) The Markov processes X and Y, with laws φ and ψ, respectively, are said to be *dual* with respect to a function $H \in M(E \times F)_b$ if, for all $x \in E,\ y \in F$ and $t \geqslant 0$,

$$\mathbf{E}_\varphi\left[H(X_t, y) \mid X_0 = x\right] = \mathbf{E}_\psi\left[H(x, Y_t) \mid Y_0 = y\right]. \tag{26}$$

If E and F are finite, every function $H \in M(E \times F)_b$ may be represented by a matrix with bounded entries $H(v, w),\ v \in E,\ w \in F$. If, further, X and Y are time-homogeneous with generator matrices Λ and Θ respectively, the expectations in (26) may be written in terms of the corresponding semigroups, i.e.,

$$\begin{aligned}\mathbf{E}_\varphi\left[H(X_t, y) \mid X_0 = x\right] &= \sum_{v\in E} (\mathrm{e}^{t\Lambda})_{xv} H(v, y),\\ \mathbf{E}_\psi\left[H(x, Y_t) \mid Y_0 = y\right] &= \sum_{w\in F} (\mathrm{e}^{t\Theta})_{yw} H(x, w).\end{aligned} \tag{27}$$

Since the duality equation (26) is automatically satisfied at $t = 0$, it is sufficient to check the identity of the derivatives at $t = 0$. That is, Eq. (26) holds for all times if and only if

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}_\varphi\left[H(X_t, y) \mid X_0 = x\right]\big|_{t=0} &= \sum_{v\in E} \Lambda_{xv}\, H(v, y)\\ = \sum_{w\in F} H(x, w)\,\Theta_{yw} &= \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}_\psi\left[H(x, Y_t) \mid Y_0 = y\right]\big|_{t=0}\end{aligned} \tag{28}$$

for all $x \in E,\ y \in F$. As a short-hand of (28), one can write $\Lambda H = H\Theta^T$, where T denotes transpose.

We will now present a duality result that justifies our construction of a marginalised sample at present via the partitioning process and sampling from the initial population (cf. Fig. 4). Indeed, it is not coincidence that we have denoted our sampling functions by $H_{\mathcal{A}}$ and our generators by Λ and Θ.

Theorem 1 *The population process $\{Z_t\}_{t\geqslant 0}$ and the partitioning process $\{\Sigma_t\}_{t\geqslant 0}$ with the generators Λ and Θ and resulting laws φ and ψ, respectively, are dual with respect to the sampling function H defined in* (20)*. Explicitly,*

$$\mathbf{E}_{\varphi}\big[H_{\mathcal{A}}(Z_t) \mid Z_0 = z\big] = \mathbf{E}_{\psi}\big[H_{\Sigma_t}(z) \mid \Sigma_0 = \mathcal{A}\big] \tag{29}$$

for all $\mathcal{A} \in \mathbb{P}(S)$ and $z \in E$.

Before we embark on the proof, let us briefly comment on the meaning of this result.

Remark 7 Eq. (29) is the formal equivalent of the construction in Fig. 4. To see this, recall the random variables $\widetilde{X}_{t,\mathcal{A}}$ from (19). With their help, the left-hand side of (29) may be reformulated as a probability distribution,

$$\mathbf{E}_{\varphi}\big[H_{\mathcal{A}}(Z_t) \mid Z_0 = z\big] = \mathbf{E}_{\varphi}\big[\mathbf{P}\big[\widetilde{X}_{t,\mathcal{A}} = \cdot\big] \mid Z_t, Z_0 = z\big] = \mathbf{P}_{\varphi}\big[\widetilde{X}_{t,\mathcal{A}} = \cdot \mid Z_0 = z\big],$$

since the expectation is over all realisations of Z_t. The right-hand side is the probability distribution considered by Bobrowski et al. (2010). Likewise, the right-hand side of (29) is equal to

$$\mathbf{E}_{\psi}\big[H_{\Sigma_t}(z) \mid \Sigma_0 = \mathcal{A}\big] = \mathbf{E}_{\psi}\big[\mathbf{P}\big[\widetilde{X}_{0,\Sigma_t} = \cdot\big] \mid \Sigma_t, \Sigma_0 = \mathcal{A}\big] = \mathbf{P}_{\psi}\big[\widetilde{X}_{0,\Sigma_t} = \cdot \mid \Sigma_0 = \mathcal{A}\big],$$

since the expectation is over all realisations of Σ_t. The right-hand side is the distribution of types when sampling from the initial population according to the partition Σ_t, where it is understood that the initial population consists of the types $X_0^1, \ldots, X_0^N$ with $\sum_{k=1}^N \delta_{X_0^k} = z$. Recall that time runs forward in Z_t, X_t^k, and $\widetilde{X}_{t,\mathcal{A}}$, but backward in Σ_t.

In order to avoid case distinctions in the calculations in the remainder of this section, let us agree on the following conventions concerning (partitions of) empty sets. Namely, we set $\mathcal{A}_{\varnothing} := \varnothing$, $\overline{H}_{\varnothing}(\pi_{\varnothing}.z) = \overline{R}_{\varnothing}(\pi_{\varnothing}.z) := \pi_{\varnothing}.z = \|z\| = N$, and $\mu(\varnothing, \varnothing) := 1$. We now collect some auxiliary results in the following lemma.

Lemma 3 *Consider a counting measure $z \in E$, a partition $\mathcal{A} \in \mathbb{P}(S)$ with $|\mathcal{A}| = m \leqslant N$ and corresponding index set $M = \{1, \ldots, m\}$, and a partition $\mathcal{B} \in \mathbb{P}(S)$. Then, the following statements hold:*

(A) $\displaystyle\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\,\big[\,\overline{H}_{\mathcal{A}}(z+\delta_x) - \overline{H}_{\mathcal{A}}(z)\big] = \sum_{j\in M}\Big(\overline{H}_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{B}|_{A_j}}\Big)(z).$

(B) $\displaystyle\sum_{x\in\mathbb{X}} z(x)\,\big[\,\overline{H}_{\mathcal{A}}(z-\delta_x) - \overline{H}_{\mathcal{A}}(z)\big] = -m\;\overline{H}_{\mathcal{A}}(z).$

Before we prove the lemma, let us give some explanations.

Remark 8 Note first that, with the above conventions, the right-hand side of identity (A) evaluates to

$$\sum_{j\in M}\Big(\overline{H}_{\mathcal{A}_{M\setminus j}}\otimes R_{\mathcal{B}|_{A_j}}\Big)(z)=NR_{\mathcal{B}}(z)\quad\text{if }\mathcal{A}=\mathbf{1}.$$

Let us now provide an interpretation for the statements of the lemma. Evaluating statement (A) for a given type $y\in\mathbb{X}$ yields the equivalent formulation

$$\Big(\sum_{x\in\mathbb{X}}\big(R_{\mathcal{B}}(z)\big)(x)\ \overline{H}_{\mathcal{A}}(z+\delta_x)\Big)(y)=\big(\overline{H}_{\mathcal{A}}(z)\big)(y)+\sum_{j\in M}\Big(\Big(\overline{H}_{\mathcal{A}_{M\setminus j}}\otimes R_{\mathcal{B}|_{A_j}}\Big)(z)\Big)(y).$$

Let us read the left-hand side as the expected number of y individuals when drawing the parts of $\mathcal{A}$ without replacement from the population z to which one individual with type distribution $R_{\mathcal{B}}(z)$ has beed added. The statement then says that this can be achieved either by drawing *all* parts of $\mathcal{A}$ from z without replacement, *or* by drawing *all but one* of them from z without replacement and the parts of $\mathcal{B}$ induced by the remaining block independently of each other and of all other blocks.

Likewise, evaluating statement (B) for some type $y\in\mathbb{X}$ gives

$$\Big(\sum_{x\in\mathbb{X}}\frac{z(x)}{N}\,\overline{H}_{\mathcal{A}}(z-\delta_x)\Big)(y)=\frac{N-m}{N}\,\big(\overline{H}_{\mathcal{A}}(z)\big)(y).$$

Let us note in passing that the left-hand side is always well-defined, since $z-\delta_x<0$ can only occur with $z(x)=0$, in which case the term vanishes. This left-hand side yields the expected number of y individuals when drawing the parts of $\mathcal{A}$ from the population z *after* removal of one randomly sampled individual. The statement then tells us that this is the same as *first* drawing the parts of $\mathcal{A}$ from *all* of z and then deciding whether none of the m affected individuals has been removed, which is the case with probability $(N-m)/N$.

Proof of Lemma 3 We first observe that

$$\sum_{x\in\mathbb{X}}\big(R_{\mathcal{B}}(z)\big)(x)\,(\pi_U.\delta_x)=\pi_U.\sum_{x\in\mathbb{X}}\big(R_{\mathcal{B}}(z)\big)(x)\,\delta_x=\pi_U.\big(R_{\mathcal{B}}(z)\big)=R_{\mathcal{B}|_U}(\pi_U.z)\tag{30}$$

by Fact 2. We next evaluate $\sum_{x\in\mathbb{X}}\big(R_{\mathcal{B}}(z)\big)(x)\big[\,\overline{R}_{\mathcal{A}}(z+\delta_x)-\,\overline{R}_{\mathcal{A}}(z)\big]$ by expanding $\overline{R}_{\mathcal{A}}$ to separate the action on z from that on δ_x, summing against $R_{\mathcal{B}}(z)$ (using (30)),

applying Fact 1 and changing summation:

$$\begin{aligned}
&\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\left[\,\bar{R}_{\mathcal{A}}(z+\delta_x) - \bar{R}_{\mathcal{A}}(z)\right] \\
&\quad = \sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x) \sum_{\varnothing\neq J\subseteq M} \Big(\bar{R}_{\mathcal{A}_{M\setminus J}}(\pi_{A_{M\setminus J}}.z)\Big)\otimes\big(\pi_{A_J}.\delta_x\big) \\
&\quad = \sum_{\varnothing\neq J\subseteq M} \Big(\bar{R}_{\mathcal{A}_{M\setminus J}}\otimes R_{\mathcal{B}|_{A_J}}\Big)(z) = \sum_{\varnothing\neq J\subseteq M}\ \sum_{\mathcal{C}\succcurlyeq\mathcal{A}_{M\setminus J}} \Big(\bar{H}_{\mathcal{C}}\otimes R_{\mathcal{B}|_{A_J}}\Big)(z) \\
&\quad = \sum_{\mathcal{D}\succcurlyeq\mathcal{A}}\ \sum_{j=1}^{|\mathcal{D}|} \Big(\bar{H}_{\mathcal{D}\setminus D_j}\otimes R_{\mathcal{B}|_{D_j}}\Big)(z),
\end{aligned}$$

where, in the last step, every A_J reappears as one D_j. Using this together with (18) and (1), we obtain

$$\begin{aligned}
&\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\left[\,\bar{H}_{\mathcal{A}}(z+\delta_x) - \bar{H}_{\mathcal{A}}(z)\right] \\
&\quad = \sum_{\mathcal{C}\succcurlyeq\mathcal{A}} \mu(\mathcal{A},\mathcal{C}) \sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\left[\,\bar{R}_{\mathcal{C}}(z+\delta_x) - \bar{R}_{\mathcal{C}}(z)\right] \\
&\quad = \sum_{\mathcal{C}\succcurlyeq\mathcal{A}} \mu(\mathcal{A},\mathcal{C}) \sum_{\mathcal{D}\succcurlyeq\mathcal{C}}\ \sum_{j=1}^{|\mathcal{D}|} \Big(\bar{H}_{\mathcal{D}\setminus D_j}\otimes R_{\mathcal{B}|_{D_j}}\Big)(z) \\
&\quad = \sum_{\mathcal{D}\succcurlyeq\mathcal{A}}\ \sum_{j=1}^{|\mathcal{D}|} \Big(\bar{H}_{\mathcal{D}\setminus D_j}\otimes R_{\mathcal{B}|_{D_j}}\Big)(z) \sum_{\mathcal{A}\preccurlyeq\mathcal{C}\preccurlyeq\mathcal{D}} \mu(\mathcal{A},\mathcal{C}) \\
&\quad = \sum_{j\in M} \Big(\bar{H}_{\mathcal{A}_{M\setminus j}}\otimes R_{\mathcal{B}|_{A_j}}\Big)(z),
\end{aligned}$$

which is statement (A).

In an analoguous way, we can prove statement (B):

$$\begin{aligned}
&\sum_{x\in\mathbb{X}} z(x)\left[\,\bar{R}_{\mathcal{A}}(z-\delta_x) - \bar{R}_{\mathcal{A}}(z)\right] \\
&\quad = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \sum_{x\in\mathbb{X}} z(x) \Big(\bar{R}_{\mathcal{A}_{M\setminus J}}(\pi_{A_{M\setminus J}}.z)\Big)\otimes\Big(\bar{R}_{\mathbf{1}}^{A_J}(\pi_{A_J}.\delta_x)\Big) \\
&\quad = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \Big(\bar{R}_{\mathcal{A}_{M\setminus J}}\otimes\ \bar{R}_{\mathbf{1}}^{A_J}\Big)(z) = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \Big(\bar{R}_{\mathcal{A}_{M\setminus J}\cup A_J}\Big)(z)
\end{aligned}$$

$$= \sum_{\varnothing \neq J \subseteq M} (-1)^{|J|} \sum_{\mathcal{B} \succcurlyeq \mathcal{A}_{M\setminus J} \cup A_J} \overline{H}_{\mathcal{B}}(z) = \sum_{\mathcal{C} \succcurlyeq \mathcal{A}} \overline{H}_{\mathcal{C}}(z) \sum_{j=1}^{|\mathcal{C}|} \sum_{\varnothing \neq K \subseteq C_j} (-1)^{|K|}$$

$$= \sum_{\mathcal{C} \succcurlyeq \mathcal{A}} \overline{H}_{\mathcal{C}}(z) \sum_{j=1}^{|\mathcal{C}|} \Big[(1-1)^{|C_j|} - 1 \Big] = - \sum_{\mathcal{C} \succcurlyeq \mathcal{A}} |\mathcal{C}| \; \overline{H}_{\mathcal{C}}(z),$$

where, in the second-last step, every A_J reappears as a C_j. We therefore get

$$\sum_{x \in \mathbb{X}} z(x) \big[\, \overline{H}_{\mathcal{A}}(z - \delta_x) - \overline{H}_{\mathcal{A}}(z) \big] = \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \mu(\mathcal{A}, \mathcal{B}) \sum_{x \in \mathbb{X}} z(x) \big[\, \overline{R}_{\mathcal{B}}(z - \delta_x) - \overline{R}_{\mathcal{B}}(z) \big]$$

$$= - \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \mu(\mathcal{A}, \mathcal{B}) \sum_{\mathcal{C} \succcurlyeq \mathcal{B}} |\mathcal{C}| \; \overline{H}_{\mathcal{C}}(z) = - \sum_{\mathcal{C} \succcurlyeq \mathcal{A}} |\mathcal{C}| \; \overline{H}_{\mathcal{C}}(z) \sum_{\mathcal{A} \preccurlyeq \mathcal{B} \preccurlyeq \mathcal{C}} \mu(\mathcal{A}, \mathcal{B})$$

$$= -|\mathcal{A}| \; \overline{H}_{\mathcal{A}}(z),$$

as claimed. □

We can now proceed as follows.

Proof of Theorem 1 We start with the partitioning process. We first note that

$$\sum_{\substack{\mathcal{B} \succcurlyeq \mathcal{A}_{M\setminus j} \cup \mathcal{J} \\ \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}}} \frac{\big(N - (m-1)\big)!}{(N - |\mathcal{B}|)!} = N^{|\mathcal{J}|} \tag{31}$$

for $j \in M$ and $|\mathcal{J}| \leqslant 2$. This is easily verified by direct calculation; namely, for $|\mathcal{J}| = 1$, the sum on the left-hand side equals $(N - (m-1)) + (m-1) = N$; for $|\mathcal{J}| = 2$, it evaluates to

$$(N - (m-1))\,(N - m) + (N - (m-1))\,(2m - 1) + (m-1)^2 = N^2.$$

We now use the formulation of the process via (10) and (11) in the first step, normalisation and (31) in the second, Lemma 2 in the third, and finally another normalisation step to calculate

$$\sum_{\mathcal{B} \in \mathbb{P}(S)} \Theta_{\mathcal{A}\mathcal{B}} \, H_{\mathcal{B}}(z)$$

$$= \sum_{j \in M} \sum_{\mathcal{J} \in \mathbb{O}_{\leqslant 2}(A_j)} \frac{r_{\mathcal{J}}}{N^{|\mathcal{J}|}} \sum_{\substack{\mathcal{B} \succcurlyeq \mathcal{A}_{M\setminus j} \cup \mathcal{J} \\ \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}}} \frac{\big(N - (m-1)\big)!}{(N - |\mathcal{B}|)!} \, \big(H_{\mathcal{B}} - H_{\mathcal{A}}\big)(z)$$

$$
\begin{aligned}
&= \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{O}_{\leqslant 2}(A_j)} \frac{r_{\mathcal{J}}}{N^{|\mathcal{J}|}} \Bigg(\Bigg(\sum_{\substack{\mathcal{B}\gtrdot \mathcal{A}_{M\setminus j}\cup\mathcal{J} \\ \mathcal{B}|_{A_{M\setminus j}}=\mathcal{A}_{M\setminus j}}} \frac{\big(N-(m-1)\big)!}{N!}\ \overline{H}_{\mathcal{B}} \Bigg) - N^{|\mathcal{J}|}\, H_{\mathcal{A}} \Bigg)(z) \\
&= \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{O}_{\leqslant 2}(A_j)} \frac{r_{\mathcal{J}}}{N^{|\mathcal{J}|}} \left(\frac{\big(N-(m-1)\big)!}{N!} \Big(\overline{H}_{\mathcal{A}_{M\setminus j}} \otimes\ \overline{R}_{\mathcal{J}} \Big) - N^{|\mathcal{J}|}\, H_{\mathcal{A}} \right)(z) \\
&= \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{O}_{\leqslant 2}(A_j)} r_{\mathcal{J}} \Big(H_{\mathcal{A}_{M\setminus j}} \otimes\ R_{\mathcal{J}} - H_{\mathcal{A}} \Big)(z). \qquad (32)
\end{aligned}
$$

We now turn to the type distribution process. Here we first evaluate, with Lemma 3 (B):

$$
\begin{aligned}
&\sum_{y\in\mathbb{X}} z(y) \big[\, \overline{H}_{\mathcal{A}}\big(z+\delta_x-\delta_y\big) - \ \overline{H}_{\mathcal{A}}(z)\big] \\
&\quad = \sum_{y\in\mathbb{X}} (z+\delta_x)(y)\ \ \overline{H}_{\mathcal{A}}\big((z+\delta_x)-\delta_y\big) - \sum_{y\in\mathbb{X}} (z+\delta_x)(y)\ \ \overline{H}_{\mathcal{A}}(z) \\
&\quad = \sum_{y\in\mathbb{X}} (z+\delta_x)(y) \big[\, \overline{H}_{\mathcal{A}}\big((z+\delta_x)-\delta_y\big) - \ \overline{H}_{\mathcal{A}}(z+\delta_x) + \ \overline{H}_{\mathcal{A}}(z+\delta_x) - \ \overline{H}_{\mathcal{A}}(z)\big] \\
&\quad = (N+1-m) \big[\, \overline{H}_{\mathcal{A}}(z+\delta_x) - \ \overline{H}_{\mathcal{A}}(z)\big] - m\ \overline{H}_{\mathcal{A}}(z).
\end{aligned}
$$

From this, we obtain via summation against $R_{\mathcal{B}}(z)$ and use of Lemma 3 (A) that

$$
\begin{aligned}
&\sum_{x\in\mathbb{X}} \sum_{y\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\, z(y) \big[\, \overline{H}_{\mathcal{A}}(z+\delta_x-\delta_y) - \ \overline{H}_{\mathcal{A}}(z)\big] \\
&\qquad = (N+1-m) \sum_{j\in M} \Big(\overline{H}_{\mathcal{A}_{M\setminus j}} \otimes\ R_{\mathcal{B}|_{A_j}} \Big)(z) - m\ \overline{H}_{\mathcal{A}}(z). \qquad (33)
\end{aligned}
$$

We now have to examine $\sum_{z'\in E} \Lambda_{zz'} H_{\mathcal{A}}(z')$ for an arbitrary partition $\mathcal{A}$ of S. To this end, we use (7) and normalisation, followed by (33) and a change of summation involving (9) to calculate

$$
\begin{aligned}
\sum_{z'\in E} \Lambda_{zz'}\, H_{\mathcal{A}}(z') &= \sum_{x,y\in\mathbb{X}} \lambda(z;\ y,\ x) \big[H_{\mathcal{A}}(z+\delta_x-\delta_y) - H_{\mathcal{A}}(z) \big] \\
&= \frac{(N-m)!}{N!} \sum_{\mathcal{B}\in\mathbb{O}_{\leqslant 2}(S)} r_{\mathcal{B}} \sum_{x,y\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\, z(y) \big[\, \overline{H}_{\mathcal{A}}(z+\delta_x-\delta_y) - \ \overline{H}_{\mathcal{A}}(z)\big] \\
&= \sum_{\mathcal{B}\in\mathbb{O}_{\leqslant 2}(S)} r_{\mathcal{B}} \Big[\Big(\frac{(N-(m-1))!}{N!} \sum_{j\in M} \overline{H}_{\mathcal{A}_{M\setminus j}} \otimes\ R_{\mathcal{B}|_{A_j}} \Big) - \frac{(N-m)!}{N!}\, m\ \overline{H}_{\mathcal{A}} \Big](z) \\
&= \sum_{\mathcal{B}\in\mathbb{O}_{\leqslant 2}(S)} r_{\mathcal{B}} \sum_{j\in M} \Big(H_{\mathcal{A}_{M\setminus j}} \otimes\ R_{\mathcal{B}|_{A_j}} - H_{\mathcal{A}} \Big)(z)
\end{aligned}
$$

$$= \sum_{j \in M} \sum_{\mathcal{J} \in \mathbb{O}_{\leqslant 2}(A_j)} \sum_{\substack{\mathcal{B} \in \mathbb{O}_{\leqslant 2}(S) \\ \mathcal{B}|_{A_j} = \mathcal{J}}} r_{\mathcal{B}} \left(H_{\mathcal{A}_{M \setminus j}} \otimes R_{\mathcal{J}} - H_{\mathcal{A}} \right)(z)$$

$$= \sum_{j \in M} \sum_{\mathcal{J} \in \mathbb{O}_{\leqslant 2}(A_j)} r_{\mathcal{J}} \left(H_{\mathcal{A}_{M \setminus j}} \otimes R_{\mathcal{J}} - H_{\mathcal{A}} \right)(z),$$

which agrees with (32) and proves the claim. □

We can now harvest some interesting consequences. First, Eq. (32) contains a meaningful expression for the derivative:

Corollary 1 *For $\mathcal{A} \in \mathbb{P}(S)$, $z \in E$, and the population process $\{Z_t\}_{t \geqslant 0}$, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{E}_\varphi \left[H_{\mathcal{A}}(Z_t) \mid Z_0 = z \right] \big|_{t=0} = \sum_{j \in M} \sum_{\mathcal{J} \in \mathbb{O}_{\leqslant 2}(A_j)} r_{\mathcal{J}}^{A_j} \left(R_{\mathcal{J}}^{A_j} \otimes H_{\mathcal{A}_{M \setminus j}}^{\mathcal{A}_{M \setminus j}} - H_{\mathcal{A}}^{S} \right)(z).$$

The right-hand side has a plausible explanation. Namely, when block A_j splits into $\mathcal{J}$, the other blocks in $\mathcal{A}$ retain their current type distribution (namely, $H_{\mathcal{A}_{M \setminus j}}(\pi_{A_{M \setminus j}}.z)$). Independently of this, the parts of $\mathcal{J}$ pick their types from *all* individuals (with replacement), including those individuals that already carry other parts of $\mathcal{A}_{M \setminus j}$, which is expressed by the tensor product with $R_{\mathcal{J}}(\pi_{A_j}.z)$.

Next, via (27) together with the fact that $H_{\mathcal{A}}(z) = \mathbf{E}_\varphi \left[H_{\mathcal{A}}(Z_t) \mid Z_t = z \right]$, Eqs. (28) and (29) give rise to a system of differential equations for the expectations, namely:

Corollary 2 *For $\mathcal{A} \in \mathbb{P}(S)$ and the population process $\{Z_t\}_{t \geqslant 0}$, one has*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{E}_\varphi \left[H_{\mathcal{A}}(Z_t) \right] = \sum_{\mathcal{B} \in \mathbb{P}(S)} \Theta_{\mathcal{A}\mathcal{B}} \, \mathbf{E}_\varphi \left[H_{\mathcal{B}}(Z_t) \right].$$

This will be the basis for our concrete calculations in the next section.

7 Applications and examples

Let us now apply our results to the cases of $n = 2$ and $n = 3$ sites. Expectations will always be with respect to φ, so we will abbreviate $\mathbf{E}_\varphi$ by $\mathbf{E}$ throughout. We will assume that $Z_0 = z$, i.e., that the initial population is deterministic.

7.1 Two sites

For $U = S = \{1, 2\}$, with the abbreviation $r := r_{\{\{1\},\{2\}\}}$, the ODE system of Corollary 2 reads

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}\big[H_{\{\{1,2\}\}}(Z_t)\big] &= r\,\frac{N-1}{N}\,\mathbf{E}\big[\big(H_{\{\{1\},\{2\}\}} - H_{\{\{1,2\}\}}\big)(Z_t)\big]\\ \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}\big[H_{\{\{1\},\{2\}\}}(Z_t)\big] &= \frac{2}{N}\,\mathbf{E}\big[\big(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}}\big)(Z_t)\big],\end{aligned} \tag{34}$$

where we have dropped the upper index, which is always U. It follows that

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}\big[\big(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}}\big)(Z_t)\big]\\ &\quad= -\left(\frac{2}{N} + r\,\frac{N-1}{N}\right)\mathbf{E}\big[\big(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}}\big)(Z_t)\big].\end{aligned} \tag{35}$$

Since $L_{\{\{1,2\}\}} = \frac{N-1}{N}\,(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})$, it follows that the expected two-point LDE also decays at rate $2/N + r(N-1)/N$. In the case of two alleles per site, an equivalent formula has appeared in Bobrowski and Kimmel (2003, Ex. 1). The corresponding result in the diffusion limit goes back to Ohta and Kimura (1969), see also Durrett (2008, Chap. 8.2). As noted there, it may seem surprising that the correlations also decay via resampling (even if $r = 0$); but recall that our Moran model with recombination is an absorbing Markov chain where a single type goes to fixation in the long run, that is, Z_t will ultimately end up in a point measure.

The expected type distribution is now easily obtained from (34) and (35) via

$$\begin{aligned}\mathbf{E}\big[H_{\{\{1,2\}\}}(Z_t)\big] &= \mathbf{E}\big[H_{\{\{1,2\}\}}(Z_0)\big] - r\,\frac{N-1}{N}\int_0^t \mathbf{E}\big[\big(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}}\big)(Z_\tau)\big]\,\mathrm{d}\tau\\ &= \frac{Z_0}{N} - \frac{r(N-1)}{r(N-1)+2}\left(1 - \exp\left(-\frac{r(N-1)+2}{N}\,t\right)\right)\\ &\quad\times \mathbf{E}\big[\big(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}}\big)(Z_0)\big],\end{aligned}$$

where we have used that $\mathbf{E}\big[H_{\{\{1,2\}\}}(Z_0)\big] = Z_0/N$. The asymptotic behaviour is given by

$$\lim_{t\to\infty}\mathbf{E}\left[\frac{Z_t}{N}\right] = \frac{2}{2+r(N-1)}\,\frac{Z_0}{N} + \frac{r(N-1)}{2+r(N-1)}\,H_{\{\{1\}\{2\}\}}(Z_0). \tag{36}$$

Since Z_t will ultimately absorb in a point measure, we also know that

$$\lim_{t\to\infty}\mathbf{E}\left[\frac{Z_t}{N}\right] = \sum_{x\in\mathbb{X}}\mathbf{P}[Z_t \text{ absorbs in } x]\,\delta_x,$$

and thus $\mathbf{P}[Z_t \text{ absorbs in } x] = \lim_{t\to\infty}\mathbf{E}[Z_t/N](x)$ for all $x \in \mathbb{X}$. We can therefore read off the fixation probabilities from (36). With probability $\frac{2}{2+r(N-1)}$ (the relative intensity of resampling), the type that wins is drawn from the initial distribution. With probability $\frac{r(N-1)}{2+r(N-1)}$ (the relative intensity of recombination), it is drawn from the distribution that results when the leading and the trailing segments are sampled from the initial population without replacement.

7.2 Three sites

Now, consider $U = S = \{1, 2, 3\}$, together with the abbreviations $r_1 = r_{\{\{1\}\{2,3\}\}}$ and $r_2 = r_{\{\{1,2\}\{3\}\}}$. Let us order the partitions of $\mathbb{P}(U)$ as follows:

$$\{\{1,2,3\}\} \quad \{\{1\},\{2,3\}\} \quad \{\{1,2\},\{3\}\} \quad \{\{1,3\},\{2\}\} \quad \{\{1\},\{2\},\{3\}\}.$$

The generator of the partitioning process then reads

$$\Theta = \begin{pmatrix} -\frac{N-1}{N}(r_1+r_2) & \frac{N-1}{N}r_1 & \frac{N-1}{N}r_2 & 0 & 0 \\ \frac{2}{N}-\frac{N-1}{N^2}r_2 & -\frac{2}{N}-\frac{(N-1)^2}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{(N-1)(N-2)}{N^2}r_2 \\ \frac{2}{N}-\frac{N-1}{N^2}r_1 & \frac{N-1}{N^2}r_1 & -\frac{2}{N}-\frac{(N-1)^2}{N^2}r_1 & \frac{N-1}{N^2}r_1 & \frac{(N-1)(N-2)}{N^2}r_1 \\ \frac{2}{N}-\frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & -\frac{2}{N}-\frac{(N-1)^2}{N^2}(r_1+r_2) & \frac{(N-1)(N-2)}{N^2}(r_1+r_2) \\ 0 & \frac{2}{N} & \frac{2}{N} & \frac{2}{N} & -\frac{6}{N} \end{pmatrix}. \tag{37}$$

Recall that, by Corollary 2, we have $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[H(Z_t)] = \Theta\mathbf{E}[H(Z_t)]$, where $H(Z_t) := (H^U_{\mathcal{A}}(Z_t))_{\mathcal{A}\in\mathbb{P}(U)}$. We now transform this system into a system in terms of correlation functions. Therefore, let $L(Z_t) = (L^U_{\mathcal{A}}(Z_t))_{\mathcal{A}\in\mathbb{P}(U)}$. From (24), we know that $L(Z_t) = TH(Z_t)$, where the transformation matrix is given by

$$T = \frac{(N-1)(N-2)}{N^2}\begin{pmatrix} 1 & -1 & -1 & -1 & 2 \\ \frac{1}{N-2} & 1+\frac{1}{N-2} & \frac{-1}{N-2} & \frac{-1}{N-2} & -1 \\ \frac{1}{N-2} & -\frac{1}{N-2} & 1+\frac{1}{N-2} & -\frac{1}{N-2} & -1 \\ \frac{1}{N-2} & -\frac{1}{N-2} & -\frac{1}{N-2} & 1+\frac{1}{N-2} & -1 \\ \frac{1}{(N-1)(N-2)} & \frac{1}{N-2} & \frac{1}{N-2} & \frac{1}{N-2} & 1 \end{pmatrix}.$$

Consequently, $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[L(Z_t)] = T\Theta T^{-1}\mathbf{E}[L(Z_t)]$, where

$$T\Theta T^{-1} = \begin{pmatrix} -\frac{6}{N}-\frac{(N-1)(N-2)}{N^2}(r_1+r_2) & 0 & 0 & 0 & 0 \\ \frac{2}{N}-\frac{(N-1)}{N^2}(r_1+r_2) & -\frac{2}{N}-\frac{N-1}{N}r_2 & 0 & 0 & 0 \\ \frac{2}{N}-\frac{(N-1)}{N^2}(r_1+r_2) & 0 & -\frac{2}{N}-\frac{N-1}{N}r_1 & 0 & 0 \\ \frac{2}{N}-\frac{(N-1)}{N^2}(r_1+r_2) & 0 & 0 & -\frac{2}{N}-\frac{N-1}{N}(r_1+r_2) & 0 \\ -\frac{1}{N^2}(r_1+r_2) & \frac{2}{N}-\frac{1}{N}r_2 & \frac{2}{N}-\frac{1}{N}r_1 & \frac{2}{N}-\frac{1}{N}(r_1+r_2) & 0 \end{pmatrix}. \tag{38}$$

In contrast to (37), the matrix $T\Theta T^{-1}$ has a nice subtriangular structure, from which we can already read off that the *expected three-point LDE* $\mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)]$ (cf. (25)) decays exponentially according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}\big[L_{\{\{1,2,3\}\}}(Z_t)\big] = -\left(\frac{6N+(N-1)(N-2)(r_1+r_2)}{N^2}\right)\mathbf{E}\big[L_{\{\{1,2,3\}\}}(Z_t)\big].$$

As in the case of two sites, the decay rate contains contributions from resampling as well as from recombination. To extract more information, we recast $T\Theta T^{-1}$

into the diagonal form $V^{-1}T\Theta T^{-1}V = D$, where the entries of the diagonal matrix D are those on the diagonal of $T\Theta T^{-1}$, i.e., its eigenvalues. Consequently, $\frac{\mathrm{d}}{\mathrm{d}t}\, V^{-1}\,\mathbf{E}[L(Z_t)] = DV^{-1}\,\mathbf{E}[L(Z_t)]$.

With the help of the subtriangular structure of $T\Theta T^{-1}$, the matrix V^{-1} can be calculated explicitly. It is again subtriangular, but somewhat unwieldy. To streamline the results, we now turn to the diffusion limit, with generator Θ'' of Definition 2. Then T and T^{-1} converge to matrices T'' and $(T'')^{-1}$, respectively, with elements $T''_{\mathcal{AB}} = \mu(\mathcal{B},\mathcal{A})\,\delta_{\mathcal{B}\preccurlyeq\mathcal{A}}$ and $(T'')^{-1}_{\mathcal{AB}} = \delta_{\mathcal{B}\preccurlyeq\mathcal{A}}$, $\mathcal{A},\mathcal{B}\in\mathbb{P}(U)$ (the latter is due to inverison from below). This yields

$$T''\Theta''(T'')^{-1} = \begin{pmatrix} -(6+\varrho_1+\varrho_2) & 0 & 0 & 0 & 0 \\ 2 & -(2+\varrho_2) & 0 & 0 & 0 \\ 2 & 0 & -(2+\varrho_1) & 0 & 0 \\ 2 & 0 & 0 & -(2+\varrho_1+\varrho_2) & 0 \\ 0 & 2 & 2 & 2 & 0 \end{pmatrix},$$

where $\varrho_i = \lim_{N\to\infty} Nr_i$, $i = 1, 2$. Note that the rescaling of time has already been absorbed in Θ''. In place of V^{-1}, we now get

$$(V'')^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{2}{(2+\varrho_2)(4+\varrho_1)} & \frac{1}{2+\varrho_2} & 0 & 0 & 0 \\ \frac{2}{(2+\varrho_1)(4+\varrho_2)} & 0 & \frac{1}{2+\varrho_1} & 0 & 0 \\ \frac{1}{2(2+\varrho_1+\varrho_2)} & 0 & 0 & \frac{1}{2+\varrho_1+\varrho_2} & 0 \\ \frac{4(\varrho_1\varrho_2+(2+\varrho_1+\varrho_2)(6+\varrho_1+\varrho_2))}{(2+\varrho_1)(2+\varrho_2)(2+\varrho_1+\varrho_2)(6+\varrho_1+\varrho_2)} & \frac{2}{2+\varrho_2} & \frac{2}{2+\varrho_1} & \frac{2}{2+\varrho_1+\varrho_2} & 1 \end{pmatrix},$$

which diagonalises $T''\Theta''(T'')^{-1}$. This shows that, in contrast to $|U| = 2$, the linear combinations of $\mathbf{E}[L_{\mathcal{A}}(Z_t)]$'s that decay exponentially have coefficients depending on the recombination rates (with exception of $\mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)]$). As an example, $(4+\varrho_1)\,\mathbf{E}[L_{\{\{1\}\{2,3\}\}}(Z_t)] + 2\,\mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)]$ is one such combination and decays at rate $2+\varrho_2$. Solution of the complete system is still possible due to the triangular structure; however, it is somewhat tedious since it involves the linear combination given in the last line of $(V'')^{-1}$. Further progress may be possible if alternative scalings are employed, such as the *loose linkage approach* suggested recently by Jenkins et al. (2015).

8 Conclusion

Let us summarise our findings. We have described a marginal ancestral recombination process (ARP) and proved a duality result that relates the ARP with the Moran model forward in time, via so-called sampling functions. This was achieved by extending the recombinator formalism, which had previously proved useful in the context of deterministic recombination equations, to the stochastic setting. The ARP, together with the duality result, reveals the genealogical structure hidden in the work of Bobrowski et al. (2010), who approached the matter by functional-analytic means and forward in time. It also leads to an explicit and closed system of ordinary differential equations for the expected sampling functions, from which the expected linkage disequilibria of all orders can be calculated. It is quite remarkable that such a closed ODE system

exists: after all, the sampling functions are nonlinear, and the attempt to write down the differential equation for the expectation of a nonlinear quantity usually results in a hierarchy of equations that does not close; see Baake and Hustedt (2011) for more on the moment closure problem in the case of recombination. We would like to emphasise that the favourable structure is due to the marginalisation, which gives efficient access to correlation functions, but not to variances, for example.

Unlike Bobrowski et al. (2010), we have not included mutation so far. However, since mutation acts independently of recombination, it should be straightforward to superimpose it on the population process as well as the partitioning process. It will be rewarding to study the interplay of mutation (which increases LDE) with recombination and resampling (which decrease LDE) within the framework established here.

Acknowledgments It is our pleasure to thank Noemi Kurt, Cristian Giardina, and Frank Redig for a primer to duality theory, Fernando Cordero for helpful discussions, and Michael Baake for his help to improve the manuscript. The authors gratefully acknowledge the support from the Priority Programme *Probabilistic Structures in Evolution (SPP 1590)*, which is funded by Deutsche Forschungsgemeinschaft (German Research Foundation, DFG).

References

Aigner M (1979) Combinatorial theory. Springer, Berlin (reprint 1997)
Baake M (2005) Recombination semigroups on measure spaces. Monatsh Math 146:267–278
Baake M, Baake E (2003) An exactly solved model for mutation, recombination and selection. Can J Math 55:3–41
Baake E, Herms I (2008) Single-crossover dynamics: finite versus infinite populations. Bull Math Biol 70:603–624
Baake E, Hustedt T (2011) Moment closure in a Moran model with recombination. Markov Process Relat Fields 17:429–446
Baake E, von Wangenheim U (2014) Single-crossover recombination and ancestral recombination trees. J Math Biol 68:1371–1402
Baake E, Baake M, Salamat M (2016) The general recombination equation in continuous time and its solution. Discrete Contin Dyn Syst 36:63–95
Bennett JH (1954) On the theory of random mating. Ann Eugen 18:311–317
Berge C (1971) Principles of combinatorics. Academic Press, New York
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
Bobrowski A, Kimmel M (2003) A random evolution related to a Fisher–Wright–Moran model with mutation, recombination and drift. Math Methods Appl Sci 26:1587–1599
Bobrowski A, Wojdyła T, Kimmel M (2010) Asymptotic behavior of a Moran model with mutations, drift and recombination among multiple loci. J Math Biol 61:455–473
Bürger R (2000) The mathematical theory of selection, recombination, and mutation. Wiley, New York
Donnelly P (1986) Dual processes in population genetics. In: Tautu P (ed) Stochastic spatial processes (LNM 1212). Springer, Berlin, pp 94–105
Durrett R (2008) Probability models for DNA sequence evolution, 2nd edn. Springer, New York
Dyson FJ (1962) Statistical theory of energy levels of complex systems III. J Math Phys 3:166–175
Ethier SN, Kurtz TG (1986) Markov processes: characterization and convergence. Wiley, New York (reprint 2005)
Geiringer H (1944) On the probability theory of linkage in Mendelian heredity. Ann Math Stat 15:25–57
Golding GB (1984) The sampling distribution of linkage disequilibrium. Genetics 108:257–274
Gorelick R, Laubichler MD (2004) Decomposing multilocus linkage disequilibrium. Genetics 166:1581–1583
Griffiths RC, Marjoram R (1996) Ancestral inference from samples of DNA sequences with recombination. J Comput Biol 3:479–502

Hastings A (1984) Linkage disequilibrium, selection, and recombination at three loci. Genetics 106:153–14
Hein J, Schierup MH, Wiuf C (2005) Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, Oxford
Hudson RR (1983) Properties of a neutral allele model with intragenetic 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
Jenkins PA, Song YS (2010) An asymptotic sampling formula for the coalescent with recombination. Ann Appl Probab 20:1005–1028
Jenkins PA, Griffiths R (2011) Inference from samples of DNA sequences using a two-locus model. J Comput Biol 18:109–127
Jenkins PA, Fearnhead P, Song YS (2015) Tractable stochastic models of evolution for loosely linked loci. Electron J Probab 20:1–26
Liggett TM (1985) Interacting particle systems. Springer, Berlin (reprint 2005)
Mano S (2013) Duality between the two-locus Wright–Fisher diffusion model and the ancestral process with recombination. J Appl Probab 50:256–271
McVean GAT, Cardin NJ (2005) Approximating the coalescent with recombination. Philos Trans R Soc B 360:1387–1393
Mehta ML (1991) Random matrices. Academic Press, San Diego
Möhle M (2001) Forward and backward diffusion approximations for haploid exchangeable population models. Stoch Proc Appl 95:133–149
Ohta T, Kimura M (1969) Linkage disequilibrium due to random genetic drift. Genet Res 13:47–55
Polanska J, Kimmel M (1999) A model of dynamics of mutation, genetic drift and recombination in DNA-repeat genetic loci. Arch Control Sci 9:143–157
Polanska J, Kimmel M (2005) A simple model of linkage disequilibrium and genetic drift in human genomic SNPs: importance of demography and SNP age. Hum Hered 60:181–195
Rota G-C (1964) On the foundations of combinatorial theory I. Theory of Möbius functions. Z Wahrscheinlichkeitstheorie 2:340–368
Song YS, Song JS (2007) Analytic computation of the expectation of the linkage disequilibrium coefficient r^2. Theor Popul Biol 71:49–60
Stanley RP (1986) Enumerative combinatorics, vol I. Wadsworth & Brooks/Cole, Monterey
von Wangenheim U, Baake E, Baake M (2010) Single-crossover recombination in discrete time. J Math Biol 60:727–760
Wakeley J (2009) Coalescent theory: an introduction. Roberts and Co., Greenwood Village
Wang Y, Rannala B (2008) Bayesian inference of fine-scale recombination rates using population genomic data. Philos Trans R Soc B 363:3921–3930
Wiuf C, Hein J (1997) On the number of ancestors to a DNA sequence. Genetics 147:1459–1468