

SOLVING THE SELECTION-RECOMBINATION EQUATION:
ANCESTRAL LINES AND DUAL PROCESSES

FREDERIC ALBERTI AND ELLEN BAAKE

Received: February 3, 2021

Revised: May 13, 2021

Communicated by Roland Speicher

ABSTRACT. The deterministic selection-recombination equation describes the evolution of the genetic type composition of a population under selection and recombination in a law of large numbers regime. So far, an explicit solution has seemed out of reach; only in the special case of three sites with selection acting on one of them has an approximate solution been found, but without an obvious path to generalisation. We use both an analytical and a probabilistic, genealogical approach for the case of an *arbitrary* number of neutral sites linked to one selected site. This leads to a recursive integral representation of the solution. Starting from a variant of the ancestral selection-recombination graph, we develop an efficient genealogical structure, which may, equivalently, be represented as a weighted partitioning process, a family of Yule processes with initiation and resetting, and a family of initiation processes. We prove them to be dual to the solution of the differential equation forward in time and thus obtain a stochastic representation of the deterministic solution, along with the Markov semigroup in closed form.

2020 Mathematics Subject Classification: 60J75, 92D15, 60C05, 05C80

Keywords and Phrases: Moran model with selection and recombination, selection-recombination differential equation, ancestral selection-recombination graph, interactive particle system, duality, population genetics

1 INTRODUCTION

The *recombination equation* is a well-known nonlinear system of ordinary differential equations from mathematical population genetics (see [14] for the general background), which describes the evolution of the genetic composition of a population evolving under recombination. The genetic composition is identified with a probability measure on a space of sequences of finite length; and recombination is the genetic mechanism by which, loosely speaking, two parent individuals create the mixed sequence of their offspring during sexual reproduction, by means of one or several crossovers between the parental sequences. Elucidating the underlying structure and finding solutions was a challenge for a century, namely since the first studies by Jennings [30] in 1917 and Robbins [47] in 1918. The matter finally became simple and transparent when the corresponding stochastic backward (or dual) process was considered, which describes how the genetic material of an individual from the current population is partitioned across an increasing number of ancestors when the lines of descent are traced back into the past [6, 4]. This gives rise to a Markov process on the set of partitions of the set of sequence sites; namely, a variant of the *ancestral recombination graph* [28, 24, 25, 29, 12, 36], see also [18, Ch. 3.4]. With its help, one obtains a stochastic representation of the solution of the (deterministic) recombination equation, and a recursive solution of the Markov semigroup, see [4, 6], and [5] for a review. Furthermore, it provides the deeper reason for the underlying linear structure, which had been observed previously in the context of genetic algebras [40, 39]. The recombination equation may therefore be considered solved.

We now take the next step and attack the *selection-recombination equation*, which describes evolution under the joint action of recombination *and* selection, where selection means that fit individuals flourish at the expense of less fit ones. The selection-recombination equation first appeared in a paper by Kimura [32] in 1956. This differential equation, as well as the analogous discrete dynamical system, has since been studied intensely for a large variety of selection and recombination mechanisms, see, for example, [31, 33], [39, Chs. 9.5,9.6], as well as [14, Ch. II] and [10, Chs. 7,8] for comprehensive reviews. Most research has focussed on the long-term behaviour, which can be complex and may display subtle and counterintuitive dependence on the parameters; in particular, Hopf bifurcations and stable limit cycles may occur [26, 2]. Much research has been devoted to the case where recombination is much faster than selection, so that time-scale separation applies and the dynamics is confined to a submanifold, see [43, 46].

While a large body of knowledge has accumulated on the long-term behaviour, explicit solutions have seemed out of reach even in the simplest nontrivial cases. Indeed, the monograph [1] by Akin on the differential geometry of population genetics starts with the sentence ‘The differential equations which model the action of selection and recombination are nonlinear equations which are impossible to solve explicitly.’ The only situation where an *approximate* solution has

been found is a sequence of length three in a two-letter alphabet, where only one of the sites is under selection, and recombination involves one breakpoint (or *crossover*) at a time between the parental sequences (Stephan, Song, and Langley 2006 [49]). The approximation (in terms of special functions) seems sufficiently precise, but the derivation is cumbersome and does not reveal the underlying mathematical structure; in particular, it does not convey any hope for a generalisation beyond three sites.

The goal of this article is to reconsider the selection-recombination equation with one selected site and single crossovers, to provide a systematic and transparent approach that also generalises to an *arbitrary* number of sites, and to establish an *exact* solution via a recursion. We do this in two ways that complement each other: firstly, we solve the differential equation in the usual (forward) direction of time by analytic methods with a slight algebraic flavour. Secondly, we extend the probabilistic approach used in [6, 4] for the pure recombination equation by tracing back the (potentially) ancestral lines of individuals in the current population, this time by a variant of the *ancestral selection-recombination graph* [17, 37, 13]. This gives rise to a Markov process on the set of *weighted* partitions of the set of sequence sites, dual to the selection-recombination equation. The corresponding Markov semigroup is available in closed form, and the resulting stochastic representation yields deep insight into the genealogical content of the solution of the differential equation. Moreover, it gives access to the long-term behaviour.

The paper is organised as follows. Sections 2 and 3 introduce the selection-recombination equation, both in its own right and in terms of a dynamical law of large numbers of the corresponding Moran model, an interactive particle system that describes a *finite* population under selection and recombination. A recursive integral representation of the solution is given in Section 4. In Sections 5 and 6, we construct the stochastic process backward in time and provide the genealogical argument behind our recursion. The corresponding dual process is formulated, and the formal duality result is proved, in Section 7. Finally, the solution is presented in Section 8 in closed form, and its long-term behaviour investigated. In the Appendix, we discuss *marginalisation consistency*, which describes the forward dynamics when only a subset of the sites is considered. This is a fairly obvious, but nevertheless powerful, property in the case without selection [6]. In the presence of selection, however, it is more subtle and only true for certain subsets, but all the more interesting.

2 THE SELECTION-RECOMBINATION EQUATION

We model the distribution of the genetic types in a sufficiently large (hence effectively infinite) population under selection and recombination. The *genetic type* of an individual is represented by a sequence x on the set $S := \{1, \dots, n\}$ of sites and in the *type space*

$$X := \prod_{i \in S} X_i = X_1 \times \dots \times X_n \quad \text{with } X_i = \{0, 1\}. \quad (1)$$

The *type distribution* is identified with a probability measure $\omega \in \mathcal{P}(X)$ on the type space, where $\mathcal{P}(X)$ denotes the set of all probability measures on X . More generally, we define $\mathcal{M}(X)$ to be the set of all finite signed measures on X . It is convenient to think of ω as an element of the vector space

$$V = V_S := \bigotimes_{i \in S} V_i = V_1 \otimes \dots \otimes V_n, \quad (2)$$

where each V_i is a copy of \mathbb{R}^2 and \bigotimes denotes the tensor product of vector spaces. Here, a vector $v_i \in \mathbb{R}^2$ of the form $v_i = (p_i, 1 - p_i)^T$ for some $p_i \in [0, 1]$ is identified with the probability distribution $p_i \delta_0 + (1 - p_i) \delta_1$ on X_i with δ_0 (δ_1) denoting the point measure on 0 (on 1). The elementary tensors $v_1 \otimes \dots \otimes v_n$ correspond to products of one-dimensional marginals. Hence, it is easy to see that the tensor product of vector spaces in (2) provides an equivalent description of $\mathcal{M}(X)$.

For a subset $A \subseteq S$, we define the canonical projection

$$\pi_A : X \longrightarrow \prod_{i \in A} X_i =: X_A, \quad x \mapsto (x_i)_{i \in A} =: x_A. \quad (3)$$

The push-forward of any $\nu \in \mathcal{M}(X)$ by π_A is denoted by $\pi_A \cdot \nu$, which we abbreviate by ν^A . Thus, ν^A is the marginal measure (or marginal distribution, if ν is a probability measure) with respect to the sites in A . More explicitly,

$$\nu^A(E) = \nu(\pi_A^{-1}(E)) \quad \text{for all } E \subseteq X_A. \quad (4)$$

Note that $X_S = X$ and $\nu^S = \nu$. Moreover, X_\emptyset is the set with the single element e , which we think of as the empty sequence. Thus, $\mathcal{M}(X_\emptyset)$ is isomorphic to \mathbb{R} , and

$$\nu^\emptyset = \nu(X). \quad (5)$$

Furthermore, we write $\alpha \otimes \nu$ or $\nu \otimes \alpha$ instead of $\alpha\nu$, for all $\nu \in \mathcal{M}(X_A)$ and $\alpha \in \mathbb{R}$, in line with the usual identification of the empty tensor product with the base field. In particular, if $\nu \in \mathcal{P}(X_\emptyset)$, one has $\nu = 1$, and the above convention just means to omit such factors from products. Later, we need to project not only from X , but also from factors X_A . In order to keep the notation simple, all these projections will be denoted by the same letter π . In particular, we will write, for any two subsets $A \subseteq S$ and $B \subseteq S$ and any $\nu \in \mathcal{M}(X)$,

$$\pi_A \cdot (\pi_B \cdot \nu) = \pi_B \cdot (\pi_A \cdot \nu) = \pi_{A \cap B} \cdot \nu. \quad (6)$$

In line with Eq. (5), this implies that $\pi_A \cdot \nu = \nu(X_B) \in \mathbb{R}$ for any finite signed measure $\nu \in \mathcal{M}(X_B)$ and $A \subseteq S$ with $A \cap B = \emptyset$.

To describe the action of selection, we first fix a site $1 \leq i_* \leq n$, which we will refer to as the *selected site*. An individual of type $x \in X$ is deemed to be *fit* or *of beneficial type* if $x_{i_*} = 0$ and *unfit* or *of deleterious type* otherwise, regardless of the letters at all other sites. We also introduce

$$f(\nu) := \nu(\pi_{i_*}^{-1}(0)) = \nu^{\{i_*\}}(0) \quad (7)$$

for the proportion of fit individuals in a population with type distribution ν , and the *selection operator* $F : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ via

$$F(\nu)(x) = (1 - x_{i_*})\nu(x). \tag{8}$$

Interpreting the type distribution as an element of V as given in (2), the selection operator can also be written in tensor notation as

$$F = P_{i_*} \otimes \text{id}_{S^*}. \tag{9}$$

Here

$$P := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

the subscripts indicate the site(s) at which the matrices act, and we set $S^* := S \setminus i_*$, where we use the shorthand $S \setminus i_*$ for $S \setminus \{i_*\}$ (note that $\text{card}(S^*) = n - 1$). In words, F is the canonical projection to the subspace spanned by all elements of the form

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes v \text{ with } v \in \bigotimes_{i \in S^*} V_i,$$

and we recall that $(1, 0)^T$ and $(0, 1)^T$ correspond to the point measures δ_0 and δ_1 on X_{i_*} . Furthermore, we define

$$b(\nu) := \begin{cases} \frac{F\nu}{f(\nu)} & \text{if } f(\nu) \neq 0, \\ \nu & \text{if } f(\nu) = 0, \end{cases} \tag{10}$$

and

$$d(\nu) := \begin{cases} \frac{(1-F)\nu}{1-f(\nu)} & \text{if } f(\nu) \neq 1, \\ \nu & \text{if } f(\nu) = 1. \end{cases} \tag{11}$$

We will also write $F\nu$ instead of $F(\nu)$ where there is no risk of confusion. The measure $b(\nu)$ (the measure $d(\nu)$) is the type distribution in the beneficial (deleterious) subpopulation.

Selection now works as follows. Unfit individuals reproduce at rate 1, while fit individuals reproduce at rate $1 + s$, $s > 0$. Put differently, every individual, regardless of its type, has the neutral reproduction rate 1, while the fit individuals have an additional (selective) rate s . The net effect of this is that, in each infinitesimal time interval of length dt , an infinitesimal portion $sf(\omega_t) dt$ of ω_t is replaced by $b(\omega_t)$. That is, the dynamics of the type distribution of our population under selection alone can be described by the ordinary differential equation

$$\dot{\omega}_t = sf(\omega_t)(b(\omega_t) - \omega_t). \tag{12}$$

With the notation (10), Eq. (12) turns into the *deterministic selection equation*

$$\dot{\omega}_t = s(F - f(\omega_t))\omega_t =: \Psi_{\text{sel}}(\omega_t). \tag{13}$$

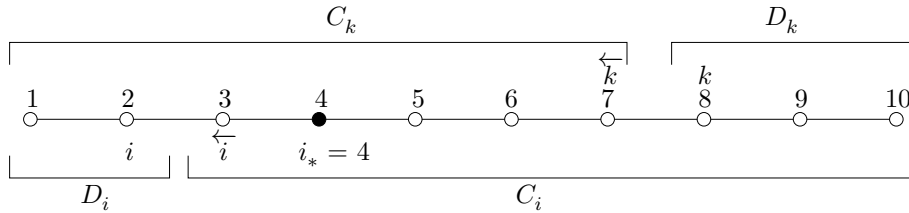


Figure 1: A sequence of length 10 with selected site, and two instances of predecessor, head, and tail; see the text for more.

We will sometimes speak of s as the *selection intensity*.

Next, we describe the action of single-crossover recombination. To this end, it is vital to introduce the following partial order on S .

DEFINITION 2.1. For two sites $i, j \in S$, we say that i precedes j , or $i \preceq j$, if either $i_* \leq i \leq j$ or $i_* \geq i \geq j$. We write $i \prec j$ if $i \preceq j$ and $i \neq j$. We furthermore define the *i -tail* as the set

$$D_i := \{j \in S : i \preceq j\}$$

of all sites that succeed i , including i itself. We define the *i -head* C_i as the complement of the i -tail, $C_i := S \setminus D_i = \overline{D_i}$ (throughout, the overbar will denote the complement with respect to S); see Figure 1. Note that $D_{i_*} = S$ and $C_{i_*} = \emptyset$. Finally, if $i \neq i_*$, we denote by \overleftarrow{i} the *predecessor* of i ; that is, the maximal $j \in S$ with $j \prec i$ (note that $\overleftarrow{i} = i_*$ is possible).

REMARK 2.2. 1. Definition 2.1 may appear awkward in that $i_* \in D_{i_*}$ but $i_* \in C_i$ for $i \in S^*$. However, it will become clear in Section 7 why this is exactly how it must be.

2. In the limiting case $s = 0$, we may single out any site as the selected one; say $i_* = n$, so that $D_i = [1 : i]$ and $C_i = [i + 1 : n]$, where $[a : b]$ for $a, b \in S$ denotes the interval $[a, a + 1, \dots, b - 1, b]$, which is empty if $b < a$.
 \diamond

For $i \in S^*$, we now define the *recombinator* $R_i : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ by

$$R_i(\nu) := \nu^{C_i} \otimes \nu^{D_i}, \quad (14)$$

with the notation of (3) and (4); we will also write $R_i\nu$ instead of $R_i(\nu)$. Then, the dynamics under the influence of single-crossover recombination is captured by the *deterministic recombination equation*

$$\dot{\omega}_t = \sum_{i \in S^*} \rho_i (R_i - \text{id}) \omega_t =: \Psi_{\text{rec}}(\omega_t) \quad (15)$$

with recombination rates $\rho_i \geq 0$ for $i \in S^*$; for consistency, we set $\rho_{i_*} := 0$. On an intuitive level, Eq. (15) means that during each infinitesimal time interval of length dt and for every $i \in S^*$, an infinitesimal portion of size $\rho_i dt$ of the population is killed off and replaced by the offspring of two randomly chosen parent individuals of types $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ (which occur in the current population with frequencies $\omega_t(x)$ and $\omega_t(y)$, respectively); the offspring then has type (x_{C_i}, y_{D_i}) . This means that, for $i < i_*$ ($i > i_*$), a single-crossover event takes place between sites i and $i + 1$ (sites $i - 1$ and i); in any case, we say that recombination happens at site i . This way, we address the *links* between neighbouring sites, as in [3], but in a way that depends on the location of the selected site.

Occasionally (cf. Section 6), it will be handy to employ a more general notion of recombinators in terms of partitions; a *partition* of S is a set \mathcal{A} of nonempty, disjoint subsets of S that exhaust S . We will refer to the elements of \mathcal{A} as *blocks*. We denote by $\mathcal{P}(S)$ the set of partitions of S (not to be confused with $\mathcal{P}(X)$, the set of probability measures on X). A partition is called an *interval* (or *ordered*) *partition* if all its blocks are intervals, that is, consist of contiguous numbers. Given an (arbitrary) partition \mathcal{A} of S and a nonempty subset $U \subseteq S$, we define by $\mathcal{A}|_U := \{U \cap A : A \in \mathcal{A}\} \setminus \{\emptyset\}$ the partition induced by \mathcal{A} on U . Generalising Eq. (14), we define, for an arbitrary partition \mathcal{A} of S ,

$$\tilde{R}_{\mathcal{A}}(\nu) := \bigotimes_{A \in \mathcal{A}} \nu^A. \tag{16}$$

Clearly, $R_i = \tilde{R}_{\{C_i, D_i\}}$ for $i \in S^*$. The formulation in terms of partitions is natural because it describes how the offspring sequence is pieced together from the parental sequences. We refer the interested reader to [6, 4] and the recent review [5] for a comprehensive discussion of the properties of $\tilde{R}_{\mathcal{A}}$ and for the general recombination equation, which involves arbitrary partitions rather than single crossovers only.

We now return to the single-crossover case and assume that selection and recombination act independently of each other. Combining (13) and (15), we obtain the *deterministic selection-recombination equation* (SRE)

$$\dot{\omega}_t = \Psi(\omega_t), \quad \text{where } \Psi := \Psi_{\text{sel}} + \Psi_{\text{rec}}. \tag{17}$$

The independence, as implied by the additivity, reflects the assumption that both selection and recombination are rare, so that one can neglect the possibility that recombination happens during selective reproduction; see Remark 3.1 below, and [27] for the worked argument in the analogous case of the selection-mutation equation.

We will throughout denote by $\omega := (\omega_t)_{t \geq 0}$ the solution of Eq. (17). Let us mention at this point that, for $i_* \in A \subseteq S$, the SRE is *marginalisation consistent* in the sense that the marginal $\omega^A := (\omega_t^A)_{t \geq 0}$ satisfies

$$\dot{\omega}_t^A = \pi_A \cdot \Psi_{\text{sel}}(\omega_t) + \pi_A \cdot \Psi_{\text{rec}}(\omega_t) = \Psi_{\text{sel}}^A(\omega_t^A) + \Psi_{\text{rec}}^A(\omega_t^A)$$

with initial condition ω_0^A , for suitably defined Ψ_{sel}^A and Ψ_{rec}^A ; this will be laid out in Appendix A. Although this is not essential for the core of the paper, it helps to understand the graphical constructions in Section 5, and is also of independent interest. There is no such consistency for $i_* \notin A$, which is a source of difficulties and pitfalls in the selective case.

3 THE MORAN MODEL WITH SELECTION AND RECOMBINATION

To gain a better understanding of Eq. (17) and to prepare for the genealogical arguments in Section 5, we briefly recall the Moran model with selection and recombination. This is a *stochastic* model that describes selection and recombination in a *finite* population, from which (17) is recovered via a dynamical law of large numbers (LLN). We will use the representation as an *interacting particle system* (IPS). The Moran IPS is a Markov chain $(\Xi_t^{(N)})_{t \geq 0}$ with state space X^N , the set of type configurations of a population of N individuals, labelled by $1 \leq \alpha \leq N$. Starting from some initial configuration $(\Xi_0^{(N)}(\alpha))_{\alpha \in [1:N]}$, it evolves as follows.

- Every individual β reproduces asexually at a fixed rate according to its fitness. That is, unfit individuals reproduce at rate 1 whereas fit individuals reproduce at rate $1 + s$, where $s > 0$ is again the selection intensity. Upon reproduction, the single offspring inherits the parent's type and replaces a uniformly chosen individual α in the population (possibly its own parent). We will realise the different reproduction rates of the two types by distinguishing between *neutral reproduction events*, which happen at rate 1 to all individuals regardless of their type, and *selective reproduction events*, which are additionally performed by fit individuals at rate s . This distinction is a crucial ingredient in the ancestral selection graph [34].
- At rate ρ_i , $i \in S^*$, every individual β reproduces sexually, choosing a partner γ uniformly at random, possibly β itself. (Biologically, this means that we include the possibility of *selfing*.) The offspring is of type $(\Xi_{C_i}^{(N)}(\beta), \Xi_{D_i}^{(N)}(\gamma))$ and replaces another uniformly chosen individual α , possibly one of its own parents.

Formally, we can summarise the transitions in the Moran IPS, starting at $\xi \in X^N$, as follows.

$$\begin{aligned} \xi &\rightarrow \xi_{\text{neut}}^{\{\alpha, \beta\}} \text{ at rate } \frac{1}{N} \text{ for all } 1 \leq \alpha, \beta \leq N, \\ \xi &\rightarrow \xi_{\text{sel}}^{\{\alpha, \beta\}} \text{ at rate } \frac{s}{N} \text{ for all } 1 \leq \alpha, \beta \leq N, \quad \text{and} \\ \xi &\rightarrow \xi_{\text{rec}}^{\{\alpha, \beta, \gamma, i\}} \text{ at rate } \frac{\rho_i}{N^2} \text{ for all } 1 \leq \alpha, \beta, \gamma \leq N \text{ and } i \in S^*, \end{aligned}$$

where, for $1 \leq \varepsilon \leq N$, the new state vectors explicitly read

$$\begin{aligned} \xi_{\text{neut}}^{\{\alpha,\beta\}}(\varepsilon) &= \begin{cases} \xi(\beta), & \varepsilon = \alpha, \\ \xi(\varepsilon), & \text{otherwise,} \end{cases} \\ \xi_{\text{sel}}^{\{\alpha,\beta\}}(\varepsilon) &= \begin{cases} \xi(\beta), & \varepsilon = \alpha \text{ and } \xi_{i_*}(\beta) = 0, \\ \xi(\varepsilon), & \text{otherwise,} \end{cases} \end{aligned} \tag{18}$$

and

$$\xi_{\text{rec}}^{\{\alpha,\beta,\gamma,i\}}(\varepsilon) := \begin{cases} (\xi_{C_i}(\beta), \xi_{D_i}(\gamma)), & \varepsilon = \alpha, \\ \xi(\varepsilon), & \text{otherwise.} \end{cases}$$

REMARK 3.1. The reader may wonder why we include both sexual and asexual reproduction. However, the ‘asexual’ reproduction events are actually sexual ones in which no recombination has occurred; that is, $C = \emptyset$ and $D = S$, so the offspring is a full copy of the first parent, and the second parent is irrelevant. Selective reproduction never occurs together with recombination due to the independence built into the SRE. \diamond

Consider now the process $Z^{(N)} := (Z_t^{(N)})_{t \geq 0}$, where $Z_t^{(N)}$ is the empirical measure

$$Z_t^{(N)} := \frac{1}{N} \sum_{\alpha=1}^N \delta_{\Xi_t^{(N)}(\alpha)}.$$

Proposition 3.1 in [15] in combination with Theorem 2.1 from [19] (see also [8]) shows that, as $N \rightarrow \infty$ without rescaling of parameters or time, the processes $Z^{(N)}$ converge almost surely locally uniformly to ω , the solution of the deterministic SRE (17). This is because the Moran models, indexed with population size, form a density-dependent family, for which a dynamical LLN applies; see [20, Ch. 11].

For our purpose, it is particularly profitable to use the *graphical representation* of the Moran IPS, see Figure 2. Here, individuals are represented by horizontal lines, labelled $1 \leq \alpha \leq N$ from bottom to top, and reproduction events are depicted as arrows between the lines with the parent at the tail, the offspring at the tip, and the offspring replacing the individual at the target line (arrows pointing to their own tails have no effect and are omitted). In line with (18) and for reasons to become clear when taking the ancestral perspective in Section 5, we distinguish two types of arrows: *neutral* arrows (with normal arrowheads), which appear between every ordered pair of lines at rate $1/N$ regardless of the types of the lines; and *selective* arrows (with star-shaped arrowheads), which are laid down at rate s/N between every ordered pair of lines, again regardless of types. Similarly, a recombination event in which the individual at line α is replaced by the joint offspring of lines β and γ is encoded as a square (on the α -th line) with the recombination site i inscribed. The square has two arms connecting to β and γ and labelled C and D , indicating that β and γ contribute the i -head and i -tail, respectively. These graphical elements appear

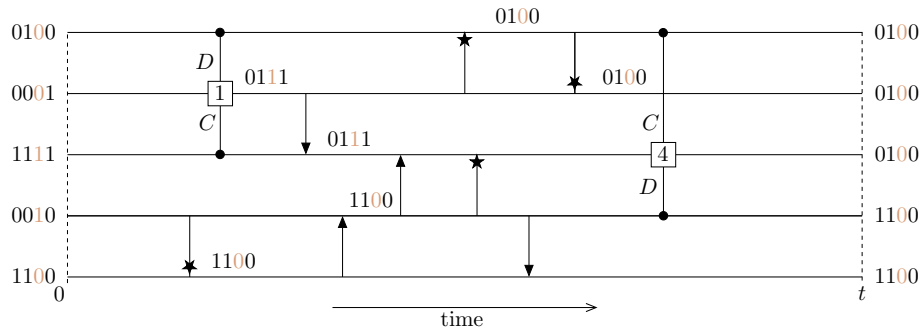


Figure 2: Graphical representation of the Moran IPS. Time runs from left to right. Arrows corresponding to neutral reproduction events are depicted with normal arrowheads, selective arrows with star-shaped arrowheads; recombination events are symbolised by squares containing the recombination point, and arms connecting to the parents that contribute the head (C) and tail (D) segments. The selected site is marked in light brown.

at rate ϱ_i/N^2 for every ordered triple of lines and every $i \in S^*$. If $\beta = \gamma$, the recombination event turns into a neutral reproduction event.

REMARK 3.2. In view of this graphical construction, another perspective on the transition rates in the Moran IPS is natural. We can say that, with rates ϱ_i , each individual is replaced by the joint offspring of two uniformly chosen parents with the crossover point at site i . Likewise, at rate 1, each individual is replaced by the offspring of a *single* uniformly chosen parent; and with rate $s f(Z_t^{(N)})$, it is replaced by the offspring of a parent chosen uniformly from the subset of fit individuals. This point of view will be particularly useful when looking back in time in Section 5. \diamond

Using different kinds of arrows for the two types of reproduction events (rather than simply letting fit individuals shoot reproduction arrows at a faster rate) allows for an *untyped* construction of the Moran IPS. That is, we first lay down the graphical elements between the lines regardless of the types and only then assign an initial type configuration. This type configuration is finally propagated forward in time under the rule that only individuals of beneficial type use the selective arrows to place their offspring, while neutral arrows and recombination arms are used by all individuals, regardless of type.

4 RECURSIVE SOLUTION OF THE SELECTION-RECOMBINATION EQUATION

Our first main result will be a recursive solution of the SRE. The recursion starts at i_* and works along the site indices in agreement with the partial order of Definition 2.1. If the original indices are used, the recursion must be

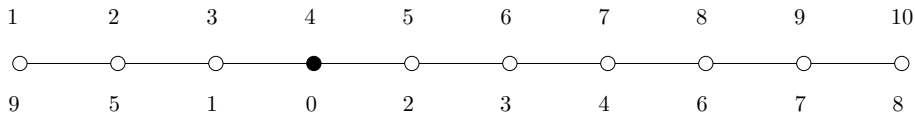


Figure 3: A nondecreasing permutation of sites. The original labels of the sites, $1 \leq i \leq n$, are at the top; below each site with label i , we have noted the corresponding k for which $i_k = i$.

formulated individually for every choice of i_* ; in particular, it looks quite different depending on whether i_* is at one of the ends or in the interior of the sequence. To establish the recursion in a unified framework, we introduce a relabelling; let us fix a nondecreasing (in the sense of the partial order from Definition 2.1) permutation $(i_k)_{0 \leq k \leq n-1}$ of S (cf. Fig. 3) and denote the corresponding heads and tails by upper indices, that is, $C^{(k)} := C_{i_k}$ and $D^{(k)} := D_{i_k}$ (cf. Figure 1). Note that $i_0 = i_*$, $D^{(0)} = S$ and $C^{(0)} = \emptyset$. Note also that this definition implies that for all $\ell \geq k$, one has either $D^{(\ell)} \subseteq D^{(k)}$ (if $\ell \succ k$) or $D^{(\ell)} \subseteq C^{(k)}$ (if ℓ and k are incomparable). Furthermore, we define $\varrho^{(k)} := \varrho_{i_k}$ and $R^{(k)} = R_{i_k}$ for $k > 0$.

First, we recapitulate the solution of the pure selection equation, that is, we solve (17) in the special case that all recombination rates vanish. Then, in accordance with the labelling given by $(i_k)_{1 \leq k \leq n-1}$, we will successively add sites at which we allow recombination. We set the scene as follows.

DEFINITION 4.1. For $\varrho^{(1)}, \dots, \varrho^{(n-1)}$ as above and every $k \in [0 : n - 1]$, we set

$$\Psi_{\text{rec}}^{(k)} := \sum_{\ell=1}^k \varrho^{(\ell)} (R^{(\ell)} - \text{id}), \quad \Psi^{(k)} := \Psi_{\text{sel}} + \Psi_{\text{rec}}^{(k)}$$

(with the usual convention that the empty sum is 0, whence $\Psi_{\text{rec}}^{(0)} = 0$ and $\Psi^{(0)} = \Psi_{\text{sel}}$). We then define the SRE *truncated at k* as the differential equation

$$\dot{\omega}_t^{(k)} = \Psi^{(k)}(\omega_t^{(k)}).$$

We understand $(\omega^{(k)})_{0 \leq k \leq n-1}$ as the family of the corresponding solutions, all with the same initial condition ω_0 . In particular, $\omega^{(0)}$ is the solution of the pure selection equation (13). We also define $\psi^{(k)} = (\psi_t^{(k)})_{t \geq 0}$ as the flow semigroup associated to the differential equation defined via $\Psi^{(k)}$. In line with (17), we have $\omega = \omega^{(n-1)}$ (which is to say $\omega_t = \omega_t^{(n-1)}$ for all $t \geq 0$) and $\Psi = \Psi^{(n-1)}$, and we likewise set $\psi = \psi^{(n-1)}$. We also write φ instead of $\psi^{(0)}$.

PROPOSITION 4.2. *The solution of the pure selection equation (13) with initial condition $\omega_0 \in \mathcal{P}(X)$ is given by*

$$\omega_t^{(0)} = \varphi_t(\omega_0) = \frac{e^{st}F(\omega_0) + (1 - F)(\omega_0)}{e^{st}f(\omega_0) + 1 - f(\omega_0)}, \quad t \geq 0, \tag{19}$$

with f and F as given in (7) and (8). In particular,

$$f(\omega_t) = \frac{e^{st} f(\omega_0)}{e^{st} f(\omega_0) + 1 - f(\omega_0)} \quad (20)$$

is increasing over time and $\omega_t^{(0)} = \varphi_t(\omega_0)$ is a convex combination of the initial type distributions of the fit (that is, beneficial) and unfit (that is, deleterious) subpopulations introduced in Eqs. (10) and (11), namely,

$$\omega_t^{(0)} = f(\omega_t^{(0)})b(\omega_0) + (1 - f(\omega_t^{(0)}))d(\omega_0).$$

This implies in particular

$$b(\varphi_t(\omega_0)) = b(\omega_0) \quad \text{and} \quad d(\varphi_t(\omega_0)) = d(\omega_0). \quad (21)$$

Proof. A straightforward verification. To see Eq. (21), recall that F is a projection and $b(\nu)$ is in the image of F , while $d(\nu)$ is in the image of $1 - F$ for any $\nu \in \mathcal{P}(X)$. \square

REMARK 4.3. Eq. (20) generalises the well-known solution of the selection equation for a single site, which is simply a logistic equation, cf. [18, p. 198]. Eq. (21) reflects the plausible fact that, while the proportion of fit individuals increases at the cost of the unfit ones (as quantified in Eq. (19)), the type composition *within* the set of fit types remains unchanged, and likewise for the set of unfit types. \diamond

The main result in this section is the following recursion for the solutions of the (truncated) SREs.

THEOREM 4.4. *The family of solutions $(\omega^{(k)})_{1 \leq k \leq n-1}$ of Definition 4.1 satisfies the recursion*

$$\omega_t^{(k)} = e^{-\varrho^{(k)}t} \omega_t^{(k-1)} + \pi_{C^{(k)}} \cdot \omega_t^{(k-1)} \otimes \pi_{D^{(k)}} \cdot \int_0^t \varrho^{(k)} e^{-\varrho^{(k)}\tau} \omega_\tau^{(k-1)} d\tau$$

for $1 \leq k \leq n-1$ and $t \geq 0$, where $\omega^{(0)}$ is the solution of the pure selection equation given in Proposition 4.2.

We will first give an analytic proof, followed, in the next section, by a genealogical proof based on the ancestral selection-recombination graph (ASRG); this will provide additional insight.

To deal with the nonlinearity of recombination and to exploit the underlying *linear* structure (see [4]) more efficiently, we now introduce a variant of the product of two measures that are defined on X_A and X_B , where A and B need not be disjoint. Namely, given sets $I, J \subseteq S$ and finite signed measures ν_I, ν_J on X_I and X_J , respectively, we define

$$\nu_I \boxtimes \nu_J := (\pi_{I \setminus J} \cdot \nu_I) \otimes \nu_J,$$

which is a finite signed measure on $X_{I \cup J}$ (recall that $\pi_\emptyset \cdot \nu = \nu(X_I)$ for all finite signed measures ν on X_I , $I \subseteq S$). Note that ν_I here means any finite signed measure on X_I , whereas ν^I stands for the specific measure on X_I that is obtained from ν on X via $\nu^I = \pi_I \cdot \nu$.

PROPOSITION 4.5. *For $I, J, K \subseteq S$ and finite signed measures ν_I, ν_J, ν_K on X_I, X_J , and X_K , respectively, the operation \boxtimes has the following properties.*

- (i) $(\nu_I \boxtimes \nu_J) \boxtimes \nu_K = \nu_I \boxtimes (\nu_J \boxtimes \nu_K)$ (associativity).
- (ii) If $I \cap J = \emptyset$, we have $\nu_I \boxtimes \nu_J = \nu_I \otimes \nu_J = \nu_J \boxtimes \nu_I$ (reduction to product measure and commutativity).
- (iii) If $I \subseteq J$, then $\nu_I \boxtimes \nu_J = \nu_I(X_I)\nu_J$ (cancellation property).

Proof. For associativity, note that

$$\begin{aligned} (\nu_I \boxtimes \nu_J) \boxtimes \nu_K &= ((\pi_{I \setminus J} \cdot \nu_I) \otimes \nu_J) \boxtimes \nu_K \\ &= (\pi_{(I \cup J) \setminus K} \cdot (\pi_{I \setminus J} \cdot \nu_I) \otimes \nu_J) \otimes \nu_K \\ &= \pi_{I \setminus (J \cup K)} \cdot \nu_I \otimes \pi_{J \setminus K} \cdot \nu_J \otimes \nu_K \\ &= \pi_{I \setminus (J \cup K)} \otimes (\nu_J \boxtimes \nu_K) = \nu_I \boxtimes (\nu_J \boxtimes \nu_K), \end{aligned}$$

where we have used in the third step that

$$((I \cup J) \setminus K) \cap (I \setminus J) = I \setminus (J \cup K).$$

When $I \cap J = \emptyset$, one has

$$\nu_I \boxtimes \nu_J = \pi_{I \setminus J} \cdot \nu_I \otimes \nu_J = \pi_I \cdot \nu_I \otimes \nu_J = \nu_I \otimes \nu_J = \nu_J \otimes \nu_I,$$

which implies the claimed reduction to \otimes and thus commutativity. Finally, for $I \subseteq J$,

$$\nu_I \boxtimes \nu_J = (\pi_{I \setminus J} \cdot \nu_I) \otimes \nu_J = (\pi_\emptyset \cdot \nu_I) \otimes \nu_J = \nu_I(X_I)\nu_J$$

establishes the cancellation property. □

Under the conditions of Proposition 4.5, we now denote by $\nu_J \boxplus \nu_K$ the formal sum of ν_J and ν_K (and use \boxminus for the corresponding formal difference). Note that the formal sum turns into a proper sum (and hence \boxplus reduces to $+$) when $J = K$. Furthermore, we define

$$\nu_I \boxtimes (\nu_J \boxplus \nu_K) := (\nu_I \boxtimes \nu_J) \boxplus (\nu_I \boxtimes \nu_K). \tag{22}$$

Clearly, the right-hand side reduces to a proper sum when $I \cup J = I \cup K$. Generalising the formal sum above, let $\mathfrak{A}(X_U)$ be the real vector space of formal sums

$$\nu := \lambda_1 \nu_{U_1} \boxplus \dots \boxplus \lambda_q \nu_{U_q},$$

where $q \in \mathbb{N}$, $\lambda_1, \dots, \lambda_q \in \mathbb{R}$, $U_1, \dots, U_q \subseteq U \subseteq S$, and $\nu_{U_1}, \dots, \nu_{U_q}$ are finite signed measures on X_{U_1}, \dots, X_{U_q} , respectively. We also write $\nu(X_U) := \sum_{i=1}^q \lambda_i \nu_{U_i}(X_{U_i})$.

REMARK 4.6. If one extends the definition of \boxtimes canonically to all of $\mathfrak{A}(X_U)$ (recalling that the projections are linear), $(\mathfrak{A}(X_U), \boxtimes, \boxplus)$ becomes an associative, unital algebra with neutral element $\mathbf{1}$, the measure with weight 1 on X_\emptyset . Recall that, when multiplying $\nu \in \mathfrak{A}(X_I)$ and $\mu \in \mathfrak{A}(X_J)$ for disjoint I and J , the multiplication \boxtimes agrees with the measure product \otimes . \diamond

Now, we can rewrite $\Psi_{\text{rec}}^{(k)}$ of Definition 4.1 as

$$\Psi_{\text{rec}}^{(k)}(\nu) = \nu \boxtimes \left(\bigoplus_{\ell=1}^k \varrho^{(\ell)}(\pi_{D_\ell} \cdot \nu - \mathbf{1}) \right); \tag{23}$$

note that the right-hand side indeed reduces to a proper (rather than a formal) sum of measures via (22), because every summand is a measure on X_S .

We shall see later that, when combined with selection, this representation is superior to the use of recombinators because it nicely brings out the recursive structure; this will streamline calculations and naturally connect to the graphical construction. The fact that the head alone determines the fitness of an individual manifests itself in the right multiplicativity of Ψ_{sel} and its associated flow φ (compare Definition 4.1), as we shall see next.

LEMMA 4.7. *For all $\mu \in \mathcal{P}(X)$ and all $\nu \in \mathfrak{A}(X_{S^*})$,*

$$F(\mu \boxtimes \nu) = F(\mu) \boxtimes \nu.$$

If, in addition, $\nu(X_{S^}) = 1$, one has*

$$\Psi_{\text{sel}}(\mu \boxtimes \nu) = \Psi_{\text{sel}}(\mu) \boxtimes \nu$$

and therefore $\varphi_t(\mu \boxtimes \nu) = \varphi_t(\mu) \boxtimes \nu$ for every $t \geq 0$.

Proof. To keep the notation simple, we assume $U_1, U_2 \subseteq S^*$ and $\nu = \nu_{U_1} \boxplus \nu_{U_2}$ with finite signed measures ν_{U_1} and ν_{U_2} on X_{U_1} and X_{U_2} , respectively. By the tensor product representation of F from (9), we have

$$\begin{aligned} & F(\mu \boxtimes \nu_{U_1} + \mu \boxtimes \nu_{U_2}) \\ &= F(\mu \boxtimes \nu_{U_1}) + F(\mu \boxtimes \nu_{U_2}) = F(\pi_{\overline{U_1}} \cdot \mu \otimes \nu_{U_1}) + F(\pi_{\overline{U_2}} \cdot \mu \otimes \nu_{U_2}) \\ &= (P_{i_*} \otimes \text{id}_{\overline{U_1} \setminus i_*})(\pi_{\overline{U_1}} \cdot \mu) \otimes \text{id}_{U_1}(\nu_{U_1}) \\ &\quad + (P_{i_*} \otimes \text{id}_{\overline{U_2} \setminus i_*})(\pi_{\overline{U_2}} \cdot \mu) \otimes \text{id}_{U_2}(\nu_{U_2}) \\ &= \pi_{\overline{U_1}} \cdot (P_{i_*} \otimes \text{id}_{S^*})(\mu) \otimes \text{id}_{U_1}(\nu_{U_1}) + \pi_{\overline{U_2}} \cdot (P_{i_*} \otimes \text{id}_{S^*})(\mu) \otimes \text{id}_{U_2}(\nu_{U_2}) \\ &= F(\mu) \boxtimes \nu_{U_1} + F(\mu) \boxtimes \nu_{U_2}, \end{aligned}$$

which gives the first claim. Taking the first claim together with the fact that $f(\mu \boxtimes \nu) = f(\mu)$ if $\nu(X_{S^*}) = 1$, we get the second and the third claim. \square

Now, the postponed proof becomes straightforward.

Proof of Theorem 4.4. Let $\Psi^{(k)}$ be as in Definition 4.1. With the shorthand

$$\nu_t^{(k-1)} := \pi_{D^{(k)}} \cdot \int_0^t \varrho^{(k)} e^{-\varrho^{(k)} \tau} \omega_\tau^{(k-1)} d\tau,$$

one has $\nu_t^{(k-1)}(X_{D^{(k)}}) = 1 - e^{-\varrho^{(k)} t}$, and the right-hand side of the recursion formula from Theorem 4.4 can be expressed as

$$\mu_t^{(k)} := \omega_t^{(k-1)} \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)}). \tag{24}$$

First, we show that

$$\mu_t^{(k)} \boxtimes \pi_{D^{(\ell)}} \cdot \mu_t^{(k)} = (\omega_t^{(k-1)} \boxtimes \pi_{D^{(\ell)}} \cdot \omega_t^{(k-1)}) \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)}) \tag{25}$$

for all $1 \leq \ell \leq k$. To see this, write the left-hand side as $\omega_t^{(k-1)} \boxtimes \kappa \boxtimes \chi$, where

$$\begin{aligned} \kappa &:= e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)} \quad \text{and} \\ \chi &:= \pi_{D^{(\ell)}} \cdot (\omega_t^{(k-1)} \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)})) = \pi_{D^{(\ell)}} \cdot \mu_t^{(k)}. \end{aligned}$$

Recall that, by our monotonicity assumption on the permutation of sites, we have either $D^{(k)} \subseteq D^{(\ell)}$ or $D^{(k)} \cap D^{(\ell)} = \emptyset$. In the first case, (25) follows by cancelling κ using Proposition 4.5 (note that $\kappa(X_{D^{(k)}}) = 1$). In the second case, χ is just $\pi_{D^{(\ell)}} \cdot \omega_t^{(k-1)}$, so $\kappa \boxtimes \chi = \chi \boxtimes \kappa$, again by Proposition 4.5. Now we compute, using (23) and (24) in the first step, (25) and Lemma 4.7 in the second, Definition 4.1 in the third, and Proposition 4.5 in the last:

$$\begin{aligned} &\Psi^{(k)}(\mu_t^{(k)}) \\ &= \Psi_{\text{sel}}(\omega_t^{(k-1)} \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)})) + \sum_{\ell=1}^k \varrho^{(\ell)} \mu_t^{(k)} \boxtimes (\pi_{D^{(\ell)}} \cdot \mu_t^{(k)} \boxplus \mathbf{1}) \\ &= (\Psi_{\text{sel}}(\omega_t^{(k-1)}) + \sum_{\ell=1}^k \varrho^{(\ell)} \omega_t^{(k-1)} \boxtimes (\pi_{D^{(\ell)}} \cdot \omega_t^{(k-1)} \boxplus \mathbf{1})) \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)}) \\ &= (\Psi^{(k-1)}(\omega_t^{(k-1)}) \boxplus \varrho^{(k)} \omega_t^{(k-1)} \boxtimes (\pi_{D^{(k)}} \cdot \omega_t^{(k-1)} \boxplus \mathbf{1})) \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)}) \\ &= \dot{\omega}_t^{(k-1)} \boxtimes (e^{-\varrho^{(k)} t} \mathbf{1} \boxplus \nu_t^{(k-1)}) \\ &\quad + \omega_t^{(k-1)} \boxtimes (\varrho^{(k)} e^{-\varrho^{(k)} t} \pi_{D^{(k)}} \cdot \omega_t^{(k-1)} \boxplus \varrho^{(k)} e^{-\varrho^{(k)} t} \mathbf{1}). \end{aligned}$$

Identifying $\varrho^{(k)} e^{-\varrho^{(k)} t} \pi_{D^{(k)}} \cdot \omega_t^{(k-1)}$ with $\dot{\nu}_t^{(k-1)}$, we see that the last line is just the time derivative of $\mu_t^{(k)}$ of (24). □

REMARK 4.8. We could have proved Theorem 4.4 also without the help of formal sums and the new operations $\boxplus, \boxminus, \boxtimes$. However, we decided on the current presentation in order to familiarise the reader with this — admittedly somewhat abstract — formalism, as it is the key to stating the duality result in Section 7 in closed form. It will also allow us later to state the solution itself in closed form; see Corollary 8.2. ◇

REMARK 4.9. Note that the only property of Ψ_{sel} that entered the proof of Theorem 4.4 is the second property in Lemma 4.7. Therefore, the result remains true if Ψ_{sel} is replaced by a more general operator with this property. In particular, Theorem 4.4 remains true when frequency-dependent selection and/or mutation at the selected site is included. \diamond

An important application of Theorem 4.4 is the following recursion for the first-order correlation functions $\omega_t^{(k)} - R^{(k)}\omega_t^{(k)}$ between the type frequencies at the sites contained in $C^{(k)}$ and those contained in $D^{(k)}$, for solutions of the truncated equations. These objects are referred to as *linkage disequilibria* in biology and also of independent interest [18, Ch. 3.3].

LEMMA 4.10 (correlation functions). *The family of solutions $(\omega^{(k)})_{0 \leq k \leq n-1}$ of Definition 4.1 satisfies, for $1 \leq k \leq n-1$,*

$$(\text{id} - R^{(k)})\omega_t^{(k)} = e^{-\rho^{(k)}t}(\text{id} - R^{(k)})\omega_t^{(k-1)}.$$

Proof. The claim can be verified directly via Theorem 4.4 by using $R^{(k)}\omega_t^{(k)} = \omega_t^{(k)} \boxtimes \pi_{D^{(k)}}.\omega_t^{(k)}$. \square

5 LOOKING BACK IN TIME:

THE ANCESTRAL SELECTION-RECOMBINATION GRAPH

Our next goal is to reveal the genealogical content of the recursive solution of Theorem 4.4. We will accomplish this by a change of perspective: Instead of focusing on the evolution of the type distribution (of the entire population) forward in time as described by the SRE (17), we will trace a single individual's genealogy *back in time*.

The crucial tool for this purpose is the *ancestral selection-recombination graph* (ASRG) of [17, 37, 13]. As the name suggests, it combines the *ancestral selection graph* (ASG) of [34] and the *ancestral recombination graph* (ARG) of [28, 24, 25]. We will introduce the ASRG here as tailored for the SRE; it allows to trace back, in a Markovian way, all lines that may carry information about the type (and the ancestry) of an individual at present. This is similar to [16, 7] for the selection part and to [6, 4] for the recombination part, where the ancestral graphs consist of all *potentially ancestral lines* of an individual at present. At this point, we understand the notion of *potentially ancestral* in a broad sense, indicating lines that are potentially ancestral to some line in the graph, but not necessarily to the individual at present. Indeed, some of these lines are not potentially ancestral to the present individual itself (that is, the notion of potential ancestry is not transitive); they will be pruned away later on. Consider first a finite population of size N . Recalling the definition of the Moran IPS in Section 3, we can sample from the type distribution at present time t via the following procedure (see Figure 4).

- (1) Select an arbitrary label α from $\{1, \dots, N\}$ for the individual to be considered.

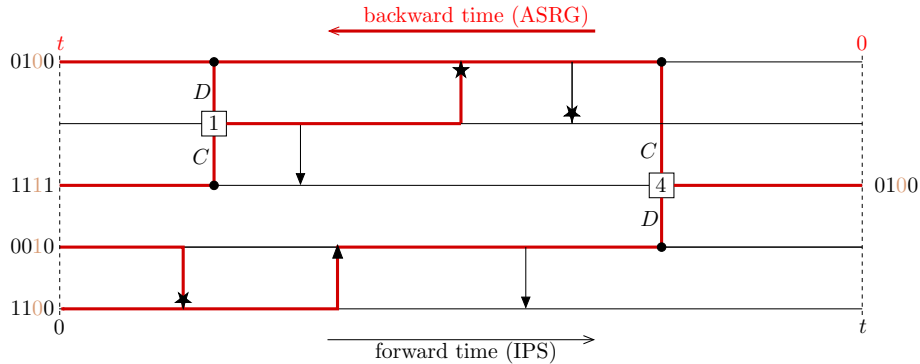


Figure 4: Sampling from the type distribution at present using the graphical representation of the Moran IPS. The ASRG is marked in red and the selected site in light brown. Notice the two different time axes for the IPS and the ASRG, respectively; while the types are propagated through the IPS from left to right, the genealogy is constructed in the opposite direction, starting with a present-day individual on the right.

- (2) Construct the untyped version of the Moran IPS.
- (3) Start the graph by tracing back the single line emerging from the individual at time t . Proceed as follows in an iterative way in the backward direction of time until the initial time is reached; note that forward time 0 (forward time t) corresponds to backward time t (backward time 0).
 - (a) If a line currently in the graph is hit by the tip of a neutral arrow, it is relocated to the line at the tail.
 - (b) If a line in the graph is hit by a selective arrow, we trace back both its potential ancestors, namely the incoming branch (at the tail of the arrow) and the continuing branch (at the tip). That is, we add the incoming line to the graph, which results in a *branching event*.
 - (c) If a line is hit by a recombination square at site i , we have a *splitting event* and trace back the lines that contribute the head (C_i) and the tail (D_i), respectively, while the line hit by the square is discontinued.
- (4) Assign types to all lines in the graph at time 0 by sampling without replacement from the initial counting measure $NZ_0^{(N)}$. Then, propagate the types forward along the lines obtained in step (3), according to the same rules as in the Moran IPS. That is, selective branchings are resolved by applying the *pecking order* derived from the Moran IPS and illustrated in Fig. 5, namely: the incoming branch is parental to the descendant line if it has a 0 at the selected site; otherwise, the continuing branch

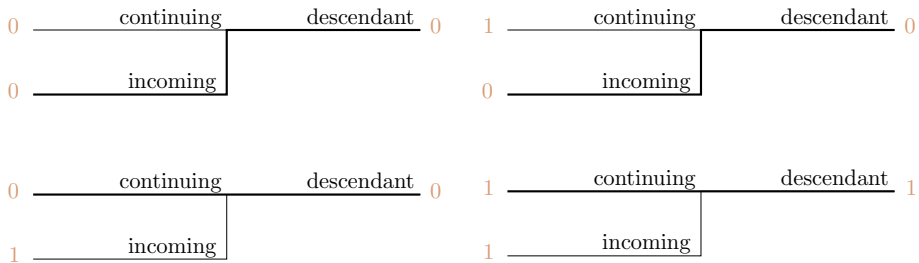


Figure 5: The pecking order between incoming line and continuing line, and the resulting type of the descendant. In each case, the ancestral line is bold. To keep the picture simple, we have only indicated the letter at the selected site. Likewise, the picture applies to the case $n = i_* = 1$.

is parental. Splitting events are resolved by piecing together heads and tails. This way, a type is associated with every line element of the graph.

The graph resulting from steps (1)–(3), along with the graphical elements indicating reproduction and recombination, is called the *untyped* ASRG, whereas the outcome of step (4) is the *typed* ASRG. While steps (3a) and (3c) are obvious, let us comment on the crucial branching step (3b). It reflects the fact that whether the incoming or the continuing branch is the true parent depends on the type of the incoming branch, which is not known in the untyped situation; in this sense, every branching event encodes a case distinction. Let us also mention that, in all events (3a)–(3c), it may happen that a line coalesces with a line that is already in the graph. Likewise, it is possible that, in a splitting event, the same parent contributes both the head and the tail; the event then turns into a relocation.

Steps (1)–(4) yield the type of the present individual considered, but also serve to elucidate the true ancestry of each site in this individual. In step (4), the paths along which the individuals contributing to the type of the present-day individual are propagated are called (*true*) *ancestral lines*, as opposed to the *potentially ancestral lines* in the untyped ASRG. More precisely, for $i \in S$, the path along which the type of the ancestor of site i is propagated is called the *ancestral line of site i* . It is obtained explicitly by adding step

- (5) Trace back the ancestry of site i by starting from the individual at present, following back the true ancestral line (determined in step (4)) in every branching event. This is the bold line in Fig. 5, and the one following either the C or D branch at every splitting event, depending on whether $i \in C$ or $i \in D$. That is, we remove from the ASRG those lines that do not contribute genetic material to site i in the present individual.

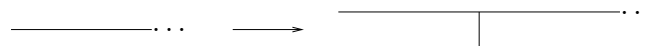
Clearly, in step (2), we need not construct the full graphical representation of the interacting particle system. Instead, it suffices to consider those events

that occur on the lines in the ASRG of the sampled individual, that is, the lines (to be) traced back in step (3). We therefore obtain the same ASRG (in distribution) if steps (2) and (3) are replaced by the following single one.

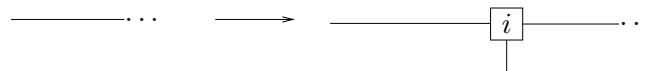
- (2'&3') Starting from the single line at forward time t , move backward and independently at rates $1, s$, and ϱ_i , let each line in the graph be hit by neutral arrows, selective arrows, and recombination events at site i , $i \in S^*$, with the (potential) parent individual(s) chosen uniformly without replacement from $[1 : N]$; update the graph accordingly.

Note that we use the homogeneity of the Poisson process here, which entails that the graphical elements are laid down according to the same law in either direction of time. Note also that the probability of choosing, for any kind of event, parent(s) already contained in the genealogy is of order $1/N$; the same is true for the probability to choose the same parent twice in a recombination event. In the limit $N \rightarrow \infty$, therefore, the coalescence rate vanishes. Likewise, selective reproduction (recombination) events always result in branching (splitting), with the incoming branch (both arms) outside the current set of lines. Furthermore, we disregard the position of the lines within the IPS; this is allowed because the types associated with each line form a *permutation-invariant* or *exchangeable* family of random variables. In particular, therefore, relocations may safely be ignored. The resulting random graph is called the ASRG *in the LLN regime*. Since we will only be concerned with this limit in the remainder of the paper, we will often omit this specification.

DEFINITION 5.1. For any given $t > 0$, the *ancestral selection-recombination graph (ASRG) in the LLN regime* is a random graph-valued function in backward time starting from a single node at time 0 and growing from right to left until time t , where branching events



occur at rate s on every line, and splitting events



occur at rate ϱ_i , $i \in S^*$, per line; all events are mutually independent. The rightmost node is called the *root* of the ASRG and the leftmost nodes are called the *leaves*.

The ASRG is almost surely finite, that is, an ASRG of finite length contains only a finite number of branchings / splittings. Note that we dispense with the star-shaped arrowheads used in the IPS for the selective events; rather, we

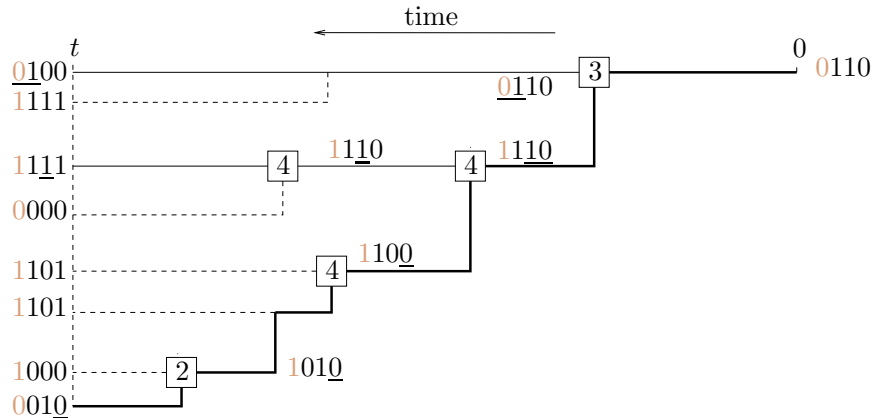
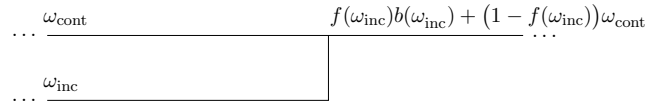


Figure 6: Tracing back the ancestry of an individual with 4 sites $i_0 = 1$, $i_1 = 2$, $i_2 = 3$ and $i_3 = 4$ under selection and recombination; the selected site $i_* = 1$ is light brown. The bold line is ancestral to site 4, the thin solid lines are ancestral to sites 1, 2, or 3, and the dashed lines are not ancestral to any site. Each branch is decorated with its type, and the sites to which it is ancestral are underlined.

use the convention that the incoming branch be placed below the continuing branch. This is again allowed due to exchangeability. For the same reason, we dispense with the labelling of the recombination arms and instead adopt the convention that the sites in the head always come from the individual on the upper line, which we place on the same level as the descendant line. The sites in the tail are provided by the line attached from below. For an example realisation of the ASRG and the construction of the type of an individual at present along with the ancestral line of one specific site, see Fig. 6.

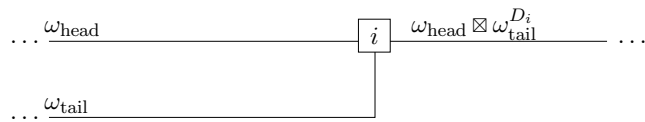
The ASRG implies the following sampling procedure for ω_t . First, construct a realisation of the ASRG, run for time t . Then, assign types to its leaves, sampled independently from ω_0 , and propagate them through the graph as described above.

REMARK 5.2. In order to connect the graphical constructions in this section to the viewpoint from the previous section, let us describe the type propagation in slightly more formal terms. Given a realisation of the ASRG of length t , we assign to each node a type distribution as follows. First, each leaf is assigned the initial type distribution ω_0 . If an internal node v arises due to a branching, we associate to v the distribution $\omega_v := f(\omega_{\text{inc}})b(\omega_{\text{inc}}) + (1 - f(\omega_{\text{inc}}))\omega_{\text{cont}}$, that is,



where ω_{inc} and ω_{cont} are the type distributions associated to the nodes that connect to v via the incoming and continuing branch.

Likewise, if v is due to splitting (at site i , say), we associate with it $\omega_{\text{head}} \boxtimes \omega_{\text{tail}}^{D_i}$, where ω_{head} and ω_{tail} are the distributions associated to the nodes that connect to v via the ancestral lines of the head and tail, respectively,



Finally, the distribution for the root equals that of the unique internal node connected to it. \diamond

EXAMPLE 5.3. In the case of pure selection ($k = 0$), our ASRG reduces to an ordered version of the ASG in the deterministic limit; this is equivalent to a special case of the *pruned lookdown ASG* in the LLN regime, as introduced in [16, 7] in the context of a probabilistic representation of the solution of the deterministic selection-mutation equation. Since there are no coalescence events in this regime, the number of lines in the graph, that is, the number of potential ancestors of an individual sampled at time t , is a simple Yule process $K = (K_t)_{t \geq 0}$ with branching rate s . This is a continuous-time branching process where, at any time t , every individual branches into two at rate s , independently of all others. In the case considered here, the process starts with $K_0 = 1$. Clearly, the pecking order implies that the individual at present will be drawn from the unfit subpopulation $d(\omega_0)$ if all K_t potential ancestors are of deleterious type, which happens with probability $(1 - f(\omega_0))^{K_t}$. Otherwise (with probability $1 - (1 - f(\omega_0))^{K_t}$), the individual will be sampled from the fit subpopulation $b(\omega_0)$. Thus, we obtain a *stochastic representation* of the solution of the selection equation by averaging over all realisations of the Yule process at time t :

$$\begin{aligned} \omega_t^{(0)} &= \varphi_t(\omega_0) \\ &= \mathbb{E}[(1 - f(\omega_0))^{K_t} \mid K_0 = 1]d(\omega_0) \\ &\quad + (1 - \mathbb{E}[(1 - f(\omega_0))^{K_t} \mid K_0 = 1])b(\omega_0). \end{aligned} \tag{26}$$

It is well known that K_t , given $K_0 = 1$, follows $\text{Geom}(e^{st})$ (cf. [22, Ch. II.4] or [50, Ex. 2.19]), where $\text{Geom}(\sigma)$ denotes the distribution of the number of

independent Bernoulli trials with success probability σ up to and including the first success. The probability generating function is given by

$$g(z) = \mathbb{E}[z^{K_t} \mid K_0 = 1] = \frac{e^{-st}z}{1 - (1 - e^{-st})z}. \quad (27)$$

Consequently,

$$\mathbb{E}[(1 - f(\omega_0))^{K_t} \mid K_0 = 1] = \frac{e^{-st}(1 - f(\omega_0))}{e^{-st}(1 - f(\omega_0)) + f(\omega_0)} = 1 - f(\omega_t^{(0)}) \quad (28)$$

with $f(\omega_t^{(0)})$ of Proposition 4.2. Inserting this into (26), we obtain $\omega_t^{(0)}$ of Proposition 4.2.

Anticipating the results in Section 6, this can be viewed as a special case of the general duality relation with respect to the duality function

$$h(m, \mu) = (1 - f(\mu))^m d(\mu) + (1 - (1 - f(\mu))^m) b(\mu) \quad (29)$$

(cf. Definition 7.1 and Proposition 7.9), which is the distribution of an individual's type at present, given it has m potential ancestors sampled from $\mu \in \mathcal{P}(X)$. \diamond

EXAMPLE 5.4. Likewise, in the case of pure recombination, the ASRG reduces to a stochastic partitioning process $\Sigma = (\Sigma_t)_{t \geq 0}$; this is a special case of [6, Sec. 6] or [4], where recombination is tackled as a more general partitioning process, rather than the single-crossover case treated here. In our case, Σ is a continuous-time Markov chain on the lattice of interval partitions of S whose law is simply stated as follows. Start with $\Sigma_0 = \{S\}$ and, if the current state is Σ_t , a transition to state $\Sigma'_t := \Sigma_t \wedge \{C_i, D_i\}$ occurs at rate ρ_i for $i \in S^*$. Here, $\mathcal{A} \wedge \mathcal{B}$ denotes the coarsest common refinement of partitions \mathcal{A} and \mathcal{B} , that is, $\mathcal{A} \wedge \mathcal{B} := \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\} \setminus \{\emptyset\}$. Note that this includes silent events, where $\Sigma'_t = \Sigma_t$. Given Σ_t , one can sample an individual of type $x = (x_1, \dots, x_n)$ from the distribution ω_t as follows. First, construct a realisation $\sigma = \{A_1, \dots, A_k\}$ of Σ_t . Then, sample individuals $\mathcal{X}(1), \dots, \mathcal{X}(k)$ i.i.d. from the initial type distribution ω_0 and set

$$x := (\pi_{A_1}(\mathcal{X}(1)), \dots, \pi_{A_k}(\mathcal{X}(k))),$$

which has distribution $\tilde{R}_\sigma(\omega_0)$; see Eq. (16). Averaging over all realisations of Σ_t gives

$$\omega_t = \mathbb{E}[\tilde{R}_{\Sigma_t}(\omega_0) \mid \Sigma_0 = \{S\}]. \quad (30)$$

As in the previous example, this can again be interpreted as a special case of a duality relation, this time with respect to the duality function $\tilde{H}(\mathcal{A}, \mu) = \tilde{R}_{\mathcal{A}}(\mu)$, see Definition 7.1. \diamond

We now turn to the genealogical proof of Theorem 4.4; recall that the start of the recursion, the solution $\omega^{(0)}$ of the pure selection equation, was already

considered in Example 5.3. We reuse the permutation $(i_k)_{0 \leq k \leq n-1}$ of sites defined in Section 4 and, in perfect analogy with the family $(\omega^{(k)})_{0 \leq k \leq n-1}$, define for $0 \leq k \leq n-1$ the ASRG truncated at k to be an ASRG with $\varrho^{(\ell)} = 0$ for all $\ell > k$. We denote the ASRG truncated at k by $\text{ASRG}^{(k)}$, or by $\text{ASRG}_t^{(k)}$ if we want to indicate its duration. Clearly, the $\text{ASRG}^{(k)}$ is the ASRG corresponding to $\omega^{(k)}$. In particular, the $\text{ASRG}^{(0)}$ is just the ASG (without recombination), and the type at the root of an $\text{ASRG}_t^{(k)}$ follows $\omega_t^{(k)}$. The key ingredient to the genealogical proof of the recursion is the following proposition, which links the type of the root of an $\text{ASRG}^{(k)}$ to the type at the root of an $\text{ASRG}^{(k-1)}$, or two independent copies thereof.

PROPOSITION 5.5. *For $1 \leq k \leq n-1$ and any given $t > 0$, let B be a Bernoulli variable with success probability $1 - e^{-\varrho^{(k)}t}$. Conditional on $\{B = 1\}$, let T be an $\text{Exp}(\varrho^{(k)})$ random variable conditioned on being $\leq t$, where $\text{Exp}(\sigma)$ denotes the exponential distribution with parameter σ . Furthermore, denote by $\mathfrak{X} \in X$ the type at the root of an $\text{ASRG}_t^{(k-1)}$, and by $\tilde{\mathfrak{X}}$ the type at the root of an $\text{ASRG}_T^{(k-1)}$, independent of the $\text{ASRG}_t^{(k-1)}$ that delivers \mathfrak{X} . The type \mathfrak{Z} at the root of an $\text{ASRG}_t^{(k)}$ is then, in distribution, given by*

$$\mathfrak{Z} = (1 - B)\mathfrak{X} + B(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\tilde{\mathfrak{X}})).$$

Before we prove this, let us give some intuition. We work with the untyped $\text{ASRG}_t^{(k)}$ and consider the line ancestral to $D^{(k)}$. It is clear that this is a single line because, due to the partial order, none of the splitting events in the $\text{ASRG}^{(k)}$ partition $D^{(k)}$. Note that the location of the true ancestral line is not yet known, since this is only decided in step (4), when propagating the types forward, as in Figure 6.

We distinguish two cases. With probability $e^{-\varrho^{(k)}t}$, no splitting at site i_k has occurred along this line, so the tail is ‘glued’ to the head. Thus, \mathfrak{Z} may be constructed as in the absence of recombination at site i_k , that is, via an $\text{ASRG}_t^{(k-1)}$; this gives the first term on the right-hand side. With probability $1 - e^{-\varrho^{(k)}t}$, a splitting at site i_k has occurred along the ancestral line of $D^{(k)}$. We then consider the time of the last, that is, of the *leftmost* splitting event at site i_k on the line in question and identify this time with $t - T$ (since such splitting events occur at rate $\varrho^{(k)}$ and due to the homogeneity of the Poisson process, T is indeed distributed as stated). The ancestry of the sites in $C^{(k)}$ is then unaffected by the split and thus follows an $\text{ASRG}_t^{(k-1)}$ (in line with marginalisation consistency, see Theorem A.2). But the sites contained in $D^{(k)}$ now come from a different individual. Since $t - T$ is the time of the *leftmost* splitting event, we know that no further splits at site i_k have occurred at any point further back in the past. This means that, at this point, the tail of the individual at the root of an independent $\text{ASRG}_T^{(k-1)}$ enters the ancestral line. The combination of head and tail as described gives the second term on the right-hand side.

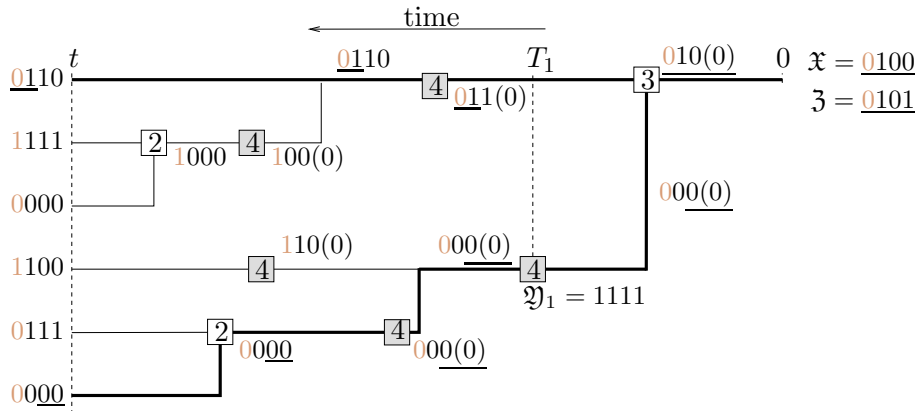


Figure 7: Determining the type at the root of a $\text{cASRG}^{(4)}$. The graph is a $\text{cASRG}^{(4)}$, the selected site is light brown, ancestral lines in the $\text{ASRG}^{(3)}$ are printed in bold, and ancestral letters are underlined. The shaded recombination squares indicate splitting events at site 4, where a new copy of an $\text{ASRG}^{(4)}$ is attached for the remaining time. Parentheses mark the 4th site in the $\text{ASRG}^{(3)}$ that is replaced by the tail of the new copy. Thus, \mathfrak{x} is obtained by ignoring the shaded squares as well as the parentheses, and \mathfrak{z} is then obtained by replacing the 0 in brackets in the type of the lower branch of the rightmost recombination event by the 1 from \mathfrak{Q}_1 .

In order to turn these heuristics into a proof, we have to make the construction of the ancestral line of $D^{(k)}$ explicit. To this end, we mimick the recursion forward in time by coupling the $\text{ASRG}_t^{(k)}$ to an $\text{ASRG}_t^{(k-1)}$. To keep things transparent, we introduce the following simplified construction; see Fig. 7.

DEFINITION 5.6 (collapsed ASRG). Let $1 \leq k \leq n - 1$ be given. A *collapsed ASRG truncated at k* , or $\text{cASRG}^{(k)}$, is an $\text{ASRG}^{(k-1)}$ decorated with i_k -recombination squares laid down according to independent Poisson point processes at rate $\rho^{(k)}$ on every horizontal line segment.

We can then construct a realisation of the $\text{ASRG}_t^{(k)}$ by attaching to every i_k -recombination square of a $\text{cASRG}^{(k)}$ an independent copy of an $\text{ASRG}^{(k)}$ for the remaining time; that is, for any i_k -recombination square at time $\tau \in [0, t]$, we attach an $\text{ASRG}_{t-\tau}^{(k)}$. So splitting events take the form of attachment events. In the subsequent sampling step, this attachment provides the k -tail, while the k -head comes from the original $\text{ASRG}_t^{(k-1)}$. We now describe how to use the collapsed ASRG to sample a root individual of an $\text{ASRG}_t^{(k)}$, that is, to sample from $\omega_t^{(k)}$. First, one constructs a realisation of the $\text{cASRG}_t^{(k)}$. Then, types are assigned to the leaves according to ω_0 in an i.i.d. fashion and propagated forward, where selective branchings and splitting (attachment) events are re-

solved as in the ASRG. Assume an i_k -square is encountered on a given line at some (forward) time $\tau \in [0, t]$, and the type just before the i_k -square (that is, at time $\tau - 0$) is \mathfrak{x} . We then draw a type \mathfrak{y} from $\omega_\tau^{(k)}$, independently of \mathfrak{x} , for the individual contributing the tail. The type on the line then jumps from \mathfrak{x} at time $\tau - 0$ to type $\mathfrak{z} = (\pi_{C^{(k)}}(\mathfrak{x}), \pi_{D^{(k)}}(\mathfrak{y}))$ at time τ , see Fig. 8. Keeping in mind the original motivation behind Definition 5.6 and thinking of the i_k -squares as splitting events (at site i_k) at which a new realisation of an $\text{ASRG}^{(k)}$ is attached, it is clear that this gives the correct result.

Proof of Proposition 5.5. Let $1 \leq k < n$ and $t > 0$ be fixed and let a realisation of the $\text{cASRG}_t^{(k)}$ be given, together with an assignment of types to its leaves. Elements of the proof are illustrated in Fig. 7. Note first that

- \mathfrak{X} is, in distribution, equal to the type of the root when ignoring the i_k -squares.

We consider the line ancestral to $D^{(k)}$ in the underlying $\text{ASRG}_t^{(k-1)}$. The location of this line is now well defined, since we sample the types and can perform steps (4) and (5). Note that the line ancestral to $D^{(k)}$ is, at the same time, the line ancestral to $\overset{\leftarrow}{i_k}$, the predecessor of i_k ; this is because no splits happen at i_k in the $\text{ASRG}_t^{(k-1)}$. We consider the following quantities.

- Let B_1 be the Bernoulli variable that takes the value 0 (the value 1) if there is no (at least one) recombination square on the ancestral line of $D^{(k)}$. Clearly, B_1 has success probability $1 - e^{-\varrho^{(k)}t}$.
- Conditional on $\{B_1 = 1\}$, let T_1 be the waiting time for the first i_k -square, in the backward direction of time, on the line ancestral to $D^{(k)}$ (that is, the *rightmost* i_k -square on this line in our graphical representation). Clearly, T_1 is an $\text{Exp}(\varrho^{(k)})$ -random variable conditioned to be $\leq t$, and independent of \mathfrak{X} .
- Let $\mathfrak{Y}_1 \in X$ be the type of the root of the independent $\text{ASRG}_{t-T_1}^{(k)}$ attached upon encountering the i_k -square at time T_1 , that is, an independent sample from $\omega_{t-T_1}^{(k)}$.

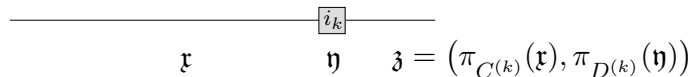


Figure 8: Upon encountering an i_k -square, the head of type \mathfrak{x} is combined with the tail of a newly sampled individual (from $\omega_\tau^{(k)}$) to form the type of the descendant.

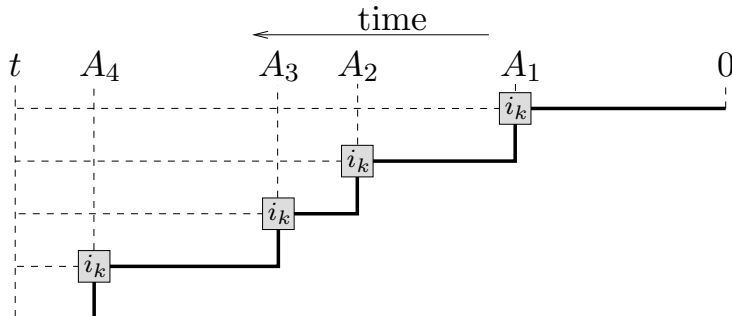


Figure 9: The ancestral line of $D^{(k)}$ after expanding all recombination events arriving at elements of $W \cap [0, t]$ used in the proof of Proposition 5.5. The ancestral lines of the corresponding heads (dashed) need not be considered any further. Note that the maximal element A_4 is the leftmost one.

We then have (cf. Fig. 7)

$$\mathfrak{Z} = (1 - B_1)\mathfrak{X} + B_1(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{Y}_1)). \tag{31}$$

We now iterate Eq. (31), see Figure 9. In the first step, we draw \mathfrak{X} and B_1 as above. If $B_1 = 1$, we also draw T_1 according to $\text{Exp}(\varrho^{(k)})$, conditioned on being $\leq t$. If $B_1 = 0$, we set $\mathfrak{Z} = \mathfrak{X}$. If $B_1 = 1$, by Eq. (31) we must construct \mathfrak{Y}_1 , which contributes the tail. Since \mathfrak{Y}_1 is the type at the root of an $\text{ASRG}_{t-T_1}^{(k)}$, we do this by applying Eq. (31) to \mathfrak{Y}_1 instead of \mathfrak{Z} , that is, we repeat the first step but replace t by $t - T_1$. So we determine whether or not there is a recombination square on the ancestral line between 0 and $t - T_1$; if there is one, we determine the waiting time for it, and so forth. More explicitly, let B_2 be the new indicator variable, which is Bernoulli with success probability $1 - e^{-\varrho^{(k)}(t-T_1)}$. If $B_2 = 0$, let \mathfrak{X}_1 be the type at the root of an independent copy of the $\text{ASRG}_{t-T_1}^{(k-1)}$. If $B_2 = 1$, let T_2 be the waiting time for the new event; T_2 follows $\text{Exp}(\varrho^{(k)})$ conditioned to be $\leq t - T_1$; and let \mathfrak{Y}_2 be the type at the root of an independent $\text{ASRG}_{t-T_1-T_2}^{(k-1)}$. Inserting this back into Eq. (31), we obtain

$$\begin{aligned} \mathfrak{Z} &= (1 - B_1)\mathfrak{X} + B_1(1 - B_2)(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{X}_1)) \\ &\quad + B_1B_2(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{Y}_2)); \end{aligned}$$

note that, if $B_1 = 0$, B_2 has not been declared, but the terms involving it remain well-defined since B_1 vanishes. Iterating this further gives

$$\mathfrak{Z} = (1 - B_1)\mathfrak{X} + \sum_{i \geq 1} B_1 \cdots B_i (1 - B_{i+1})(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{X}_i)), \tag{32}$$

where \mathfrak{X}_i is the type at the root of an independent $\text{ASRG}_{t-\sum_{j=1}^i T_j}^{(k-1)}$, and we adhere to the above convention concerning undeclared B_i . Note that, with

probability 1, exactly one of the terms on the right-hand side is nonzero; in particular, $B_1 \cdot \dots \cdot B_i = 0$ whenever $\sum_{j=1}^i T_j > t$, so everything is well defined. Let us now interpret the arrival times T_j of the i_k -squares as arrival times in a Poisson point set W with intensity measure $\rho^{(k)} \mathbb{1}_{t \geq 0} dt$ and elements $A_1 < A_2 < \dots$. When $A_i \leq t$, we have $A_i = \sum_{j=1}^i T_j$. Furthermore, $B_1 = \mathbb{1}_{\{A_1 \leq t\}}$ and, for $i \geq 1$,

$$B_1 \cdot \dots \cdot B_i = \mathbb{1}_{\{A_i \leq t\}}, \quad \text{as well as} \quad B_1 \cdot \dots \cdot B_i (1 - B_{i+1}) = \mathbb{1}_{\{A_i \leq t < A_{i+1}\}}. \quad (33)$$

We now rewrite B_1 as $B_1 = \mathbb{1}_{\{W \cap [0, t] \neq \emptyset\}}$. Together with (33), this entails that the nonzero term in (32) is the first one if $W \cap [0, t]$ is empty; and if the set is nonempty, then the nonzero term is the one with the index i that satisfies $A_i = \max(W \cap [0, t])$. Conditionally on $B_1 = 1$, we therefore set $T := t - \max(W \cap [0, t])$. The claim then follows by identifying B with B_1 , and by noting that T has the same distribution as T_1 , namely $\text{Exp}(\rho^{(k)})$ conditioned to be $\leq t$. \square

REMARK 5.7. Remembering the motivation for the collapsed ASRG^(k), we think of every i_k -square as the anchor point for a new independent copy of the ASRG^(k), which is collapsed to keep things tidy. In the above proof, we iteratively expand the i_k -squares *on the ancestral line of $D^{(k)}$* until there are no more recombination events left on that line. Therefore, the Poisson point set W has an interpretation as the collection of all recombination events on the ancestral line of $D^{(k)}$. The proof has made precise the previously heuristic notion of the last splitting event at site i_k encountered on the ancestral line of $D^{(k)}$ in the backward direction of time; that is, the leftmost event in the graphical representation, see Fig. 9. \diamond

REMARK 5.8. When sampling \mathfrak{Y}_1 via the newly attached ASRG^(k) in (31), one might wonder whether it would suffice to construct the potential ancestry of the tail alone — after all, the head of \mathfrak{Y}_1 does not enter \mathfrak{Z} . However, it cannot be overemphasised that this is not the case! Although \mathfrak{Y}_1 only contributes the tail, the branching events in its ancestry can only be resolved if the letter at the selected site is known, whence we need to also trace back the ancestry of the head attached to the new tail. We are haunted here by the fact that marginalisation consistency does not hold for the tail, see the Appendix, in particular Remark A.3. \diamond

We are now all set to re-prove Theorem 4.4. Indeed, Proposition 5.5 connects the random variable \mathfrak{Z} , delivered by an ASRG^(k), with random variables \mathfrak{X} and $\tilde{\mathfrak{X}}$, delivered by an ASRG^(k-1). This is the crucial observation that we will now exploit.

Genealogical proof of Theorem 4.4. From Proposition 5.5, we can extract the conditional distribution of \mathfrak{Z} given B and T :

$$\mathbb{P}(\mathfrak{Z} = \cdot \mid B, T) = (1 - B)\omega_t^{(k-1)} + B\pi_{C^{(k)}} \cdot \omega_t^{(k-1)} \otimes \pi_{D^{(k)}} \cdot \omega_T^{(k-1)}.$$

Theorem 4.4 now follows by integrating out B and T , keeping in mind their distributions (and denoting the distribution of the latter by λ):

$$\begin{aligned}\omega_t^{(k)} &= \mathbb{P}(B = 0)\omega_t^{(k-1)} \\ &\quad + \mathbb{P}(B = 1)\pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \int_0^\infty \pi_{D^{(k)}}.\omega_\tau^{(k-1)} d\lambda(T) \\ &= e^{-\varrho^{(k)}t}\omega_t^{(k-1)} + \pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \int_0^t \varrho^{(k)}e^{-\varrho^{(k)}t}\pi_{D^{(k)}}.\omega_\tau^{(k-1)} d\tau,\end{aligned}$$

and we are done. \square

6 INTERLUDE

Using our insight from the proof of Proposition 5.5, we now informally describe a more efficient version of the ASRG in order to motivate the more elegant dual process and the formal duality result that are detailed and proved in the next section. We start with an untyped ASG = ASRG⁽⁰⁾, since this marks the beginning of the recursion. Recall that, in the iteration leading from $\omega^{(0)}$ to $\omega^{(1)}$ via the cASRG⁽¹⁾, i_1 -recombination squares are laid down at rate $\varrho^{(1)}$ independently on every line of the ASG. But at most one of these squares turns out as relevant; namely the rightmost square on the ancestral line of $D^{(1)}$, if there is such a square. Recall also that the head of the root of the ASRG⁽¹⁾, that is its sites in $C^{(1)}$, are delivered by the initial ASG, independently of any recombination squares; while the sites in the tail are delivered by an independent copy of the ASRG⁽¹⁾, attached below the square for the remaining time and processed in the same way as the initial one. This procedure stops when no further recombination square is found on the ancestral line of the tail.

In order to reduce the ASRG⁽¹⁾ to its essential parts, we now start over and decorate the ASG with *at most one recombination event*, which will play the role of the relevant one, see Figure 10. Namely, with probability $e^{-\varrho^{(1)}t}$, we include no event, and both head and tail are delivered by the ASG. With probability $1 - e^{-\varrho^{(1)}t}$, we include one event, which happens at time T_1 distributed according to $\text{Exp}(\varrho^{(1)})$ conditioned to be $\leq t$. Since we are in an untyped setting and do not know which of the lines in the ASG will be ancestral to the head, we symbolise the event by an i_1 -box (that is, a box labelled i_1) that covers all lines. At the bottom of the box, we attach an independent copy of the ASG starting with a single line and running for the remaining time. The new ASG is processed in the analogous way, with t replaced by $t - T_1$. This procedure stops when no further i_1 -box is encountered; this is (almost surely) the case after a finite number of steps, see Figure 10 (top left). The initial ASG delivers the head, while the last ASG attached delivers the tail. In particular, at every i_1 -box, the tail delivered by the ASG attached below is combined with the head of whichever of the lines running through the box will turn out to be ancestral to the root of the ASG it belongs to.

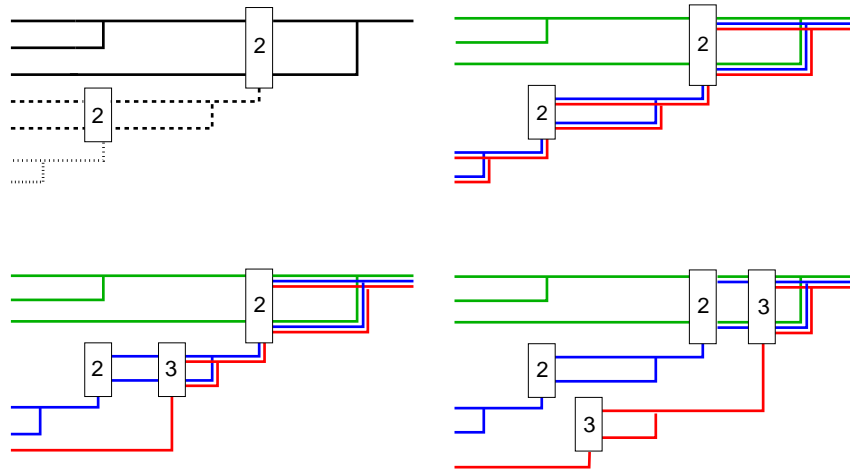


Figure 10: Constructing the essential ASRG for three sites with $i_* = 1$. Top left: An ASG decorated with a 2-box to which decorated ASGs are attached repeatedly; solid, dashed, and dotted lines correspond to steps 1, 2, and 3, respectively. Top right: Labelling and pruning the resulting graph; green, blue and red encode sites 1,2 and 3, respectively. Bottom: Adding 3-boxes to the top right graph (two different realisations bottom left and bottom right).

We now label each line in the graph with the set of sites in the root to which the line is (potentially) ancestral. This will finally allow us to prune away those lines that are not informative for the type of the root, see Figure 10 (top right). We start with the label $A = S$ for the single line at the root. When a branching event occurs to a line labelled A , both branches inherit the label. Upon encountering an i_1 -box, the continuing line is ancestral to $A \cap C^{(1)}$, while the line attached below is ancestral to $A \cap D^{(1)}$. If $A \cap C^{(1)} = \emptyset$ (this applies to a second and any further i_1 -box), we prune the continuing line away, because it is neither ancestral to any sites in A at the root, nor does it affect their ancestry. The latter is true because now the same new tail is provided for *all* potential ancestors of the head, at the same moment; in contrast to the original ASRG, where a new tail may compete with others, see Figure 7.

We finally work up the recursion by decorating the set of lines potentially ancestral to $D^{(2)}$ with i_2 -boxes, adding new ASGs, labelling, and pruning in the analogous way, see Figure 10 (bottom). That is, with probability $e^{-\rho^{(2)}t}$, no i_2 -box appears. With probability $1 - e^{-\rho^{(2)}t}$, we add an i_2 -box, at a time distributed according to $\text{Exp}(\rho^{(2)})$ conditioned to be $\leq t$. A new ASG labelled $D^{(2)}$ is then attached below, starting with a single line, while the continuing lines now carry the label $D^{(1)} \cap C^{(2)}$. If a second $\text{Exp}(\rho^{(2)})$ waiting time still falls within the remaining time, a second i_2 -box occurs, with no lines running through it and a single line labelled $D^{(2)}$ starting a new ASG below; and so on

until no further i_2 -box is encountered.

We continue like this until S^* is exhausted. The resulting graph is the *essential* ASRG. Rather than constructing it via recursion over S^* with successive addition of boxes, labelling, and pruning, it can also be produced in one go in a Markovian manner as follows.

- Start with a single line labelled S .
- Every line independently branches at rate s ; both offspring lines inherit the label of the parent.
- Every set of lines that carry the same label, say A , independently receives an i -box at rate ϱ_i for every $i \in S^*$ with $A \cap D_i \neq \emptyset$, upon which either of the following happens.
 - If $A \cap D_i \neq A$, the lines continue through the box and change their labels to $A \cap C_i$; a single new line labelled $A \cap D_i$ starts below the box.
 - If $A \cap D_i = A$, no lines continue through the box and a single new line labelled A starts below the box.
- Stop when the time horizon t is reached.

Note that the resulting graph may be conceived as a collection of (conditionally) independent ASGs, each with its own label, and joined together by recombination boxes. It is now easy to see that all the relevant information can be condensed into a *weighted partitioning process*, namely a Markov process in continuous time that holds, at any time, an interval partition \mathcal{A} of S into the blocks $A \in \mathcal{A}$ of potentially ancestral sites, together with weights v_A giving the number of lines in the respective ASGs. This will be formalised in the next section.

7 DUALITY

For the genealogical proof of the recursive solution in Theorem 4.4, we relied on the graphical construction, which implicitly assumes a duality between the ASRG and the solution of the SRE. Since the ASRG is somewhat unwieldy from a technical standpoint, our next goal is to construct a simpler dual process. Let us begin with our definition of duality for Markov processes, which is a straightforward extension of the standard concept (see [38, Ch. 3.4.4] or [35] for thorough expositions, and [41] for an early application to population genetics).

DEFINITION 7.1. Let $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$ and $\mathcal{Y} = (\mathcal{Y}_t)_{t \geq 0}$ be continuous-time Markov processes with state spaces E and F , respectively. \mathcal{X} and \mathcal{Y} are said to be *dual* with respect to some bounded measurable function $H : E \times F \rightarrow \mathbb{R}^d$ if

$$\mathbb{E}[H(\mathcal{X}_t, y) \mid \mathcal{X}_0 = x] = \mathbb{E}[H(x, \mathcal{Y}_t) \mid \mathcal{Y}_0 = y]$$

holds for all $t \geq 0, x \in E$, and $y \in F$. Furthermore, H is referred to as a *duality function* for \mathcal{X} and \mathcal{Y} and we abbreviate the duality by $(\mathcal{X}, \mathcal{Y}, H)$.

REMARK 7.2. The slight extension of the standard concept consists in allowing for an \mathbb{R}^d -valued duality function instead of the usual real-valued H . This is, of course, equivalent to introducing a family of d real-valued duality functions. It touches on the interesting problem of finding *all* duality functions for a given pair of Markov processes. The corresponding *duality space* has been introduced in [41] and investigated in [42]. \diamond

Motivated by our observation at the end of Section 6, we now define a suitable dual process for ω , and a corresponding duality function. More precisely, we will find three different processes dual to ω , each providing different insight; namely, the weighted partitioning process, a family of Yule processes with initiation and resetting, and a family of *initiation processes*.

7.1 THE WEIGHTED PARTITIONING PROCESS

For the first dual process, we refer back to the essential ASRG, which can be formalised as a weighted partitioning process. Just as in the case without selection, the partitioning describes how the genotype of a given individual is pieced together from the types of its ancestors. To include selection, a positive integer (weight) is assigned to each block, denoting the number of lines in the ASG labelled with this block. As in the single-site case (cf. Figure 5), the true ancestor will be of deleterious type if and only if *all* potential ancestors are of deleterious type.

DEFINITION 7.3. The *weighted partitioning process* (WPP) is a continuous-time Markov chain $(\Sigma, V) = (\Sigma_t, V_t)_{t \geq 0}$ with (countable) state space

$$F := \bigcup_{k \geq 0} (\mathcal{I}_k(S) \times \mathbb{N}_+^k),$$

where $\mathcal{I}_k(S)$ denotes the set of all interval partitions of S into exactly k blocks, and transitions

- (1) $(\mathcal{A}, v) \rightarrow (\mathcal{A}, w)$ at rate sv_A if $w_A = v_A + 1$ for some $A \in \mathcal{A}$ and $w_B = v_B$ for all $A \neq B \in \mathcal{A}$.
- (2) $(\mathcal{A}, v) \rightarrow (\mathcal{A} \wedge \{C_i, D_i\}, w)$ at rate ϱ_i if, for $i \in S^*$ and the unique $A \in \mathcal{A}$ with $A \not\subseteq C_i, D_i$, $w_{A \cap C_i} = v_A$, $w_{A \cap D_i} = 1$, and $w_B = v_B$ for all $A \neq B \in \mathcal{A}$.
- (3) $(\mathcal{A}, v) \rightarrow (\mathcal{A}, w)$ at rate $\varrho_{\min(A)}^{A \cup \{i_*\}}$ if, for some $A \in \mathcal{A}$, $w_A = 1$ and $w_B = v_B$ for all $B \neq A$ (the minimum is in the sense of \preceq).

Note that transition (3) is silent if $w_A = w_B = 1$.

These transitions are a straightforward translation of the dynamics of the essential ASRG at the end of Section 6; clearly, $(\Sigma_t, V_t) = (\mathcal{A}, v)$ represents the set of ASGs present at time t , where each block A of \mathcal{A} corresponds to one ASG with v_A lines. For every $i \in S^*$, every A splits into $A \cap C_i$ and $A \cap D_i$ at rate ϱ_i independently of all other blocks. If this split is nontrivial, then $A \cap C_i$ inherits the weight of A (reflecting the continuing lines), while the weight of $A \cap D_i$ is set to 1 (reflecting the new ASG attached below and starting with a single line); this gives transition (2). If $A \subseteq C_i$, nothing happens. If $A \subseteq D_i$, the weight is reset to 1 (again reflecting the new ASG attached below); note that this happens whenever the split leaves A intact but separates it from the selected site, which gives rise to the total rate of $\varrho_{\min(A)}^{A \cup \{i_*\}}$ in transition (3). Note also that the marginal Σ is the partitioning process of Example 5.4. Independently of everything else, every block experiences branching at rate s (transition (1)). Based on the WPP, we now define the corresponding candidate for our duality function.

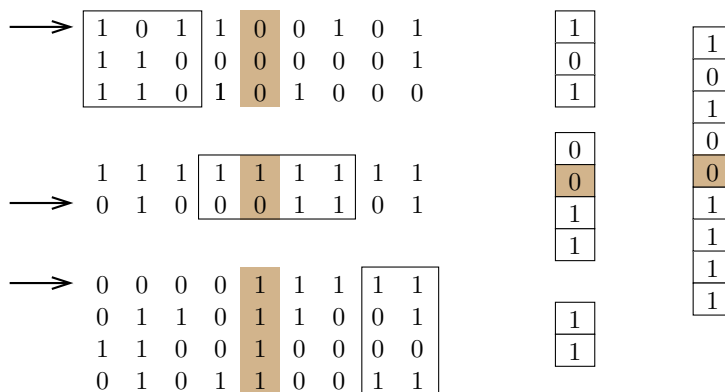


Figure 11: Illustration of the duality function H for a WPP in state (\mathcal{A}, v) , where $\mathcal{A} = \{\{1, 2, 3\}, \{4, 5, 6, 7\}, \{8, 9\}\}$ and $v_{\{1,2,3\}} = 3$, $v_{\{4,5,6,7\}} = 2$, and $v_{\{8,9\}} = 4$. The selected site is $i_* = 5$ and highlighted in light brown. As prescribed by (\mathcal{A}, v) , we sample 3 potential ancestors (displayed horizontally on the left) for the first, 2 for the second, and 4 for the third block of sites, all i.i.d. according to ν . The true ancestor (marked by an arrow) is then sampled uniformly at random from all individuals of beneficial type within the respective samples, except in the case of the third block, since there are no individuals of beneficial type. The resulting marginal types for the individual blocks (middle) are then merged into the sequence on the right. The distribution of this sequence is $H(\mathcal{A}, v; \nu)$.

DEFINITION 7.4. For an interval partition \mathcal{A} of S , associated weights $v := (v_A)_{A \in \mathcal{A}}$, and $\nu \in \mathcal{P}(X)$, we define

$$H(\mathcal{A}, v; \nu) := \bigotimes_{A \in \mathcal{A}} \pi_A \cdot \left((1 - f(\nu))^{v_A} d(\nu) + (1 - (1 - f(\nu))^{v_A}) b(\nu) \right).$$

The function H has the following meaning, which is illustrated in Figure 11. For a given (\mathcal{A}, v) and every $A \in \mathcal{A}$, we sample one sequence according to ν for each of the v_A leaves of an ASG. The type at the root of this ASG is then distributed according to $b(\nu)$ (according to $d(\nu)$) if at least one of the leaves (none of the leaves) carries a beneficial type, just as in the case of pure selection in Example 5.3. Finally, the sequence at the root of the ASRG is pieced together by taking, for every $A \in \mathcal{A}$, the sites in A from the type delivered by the ASG corresponding to A . The resulting sequence has distribution $H(\mathcal{A}, v; \nu)$; note that $H(\mathcal{A}, v; \nu)$ may be understood as a probability vector on X , that is, a vector in \mathbb{R}^{2^n} . Before proving the resulting duality, let us proceed to a more convenient representation of the WPP.

7.2 THE YULE PROCESS WITH INITIATION AND RESETTING.

Since we are only dealing with single-crossover recombination (and, therefore, only interval partitions), we will take advantage of the following one-to-one correspondence between (weighted) partitions and assignments of nonnegative integers to the sites (see Figure 12). Let a vector $m = (m_k)_{1 \leq k \leq n}$ of nonnegative integers with $m_{i_*} > 0$ be given. We then obtain an (interval) partition by the rule that two sites $i < j$ belong to the same block if and only if $m_k = 0$ for all $i < k < j$; intuitively, the nonzero integers tell us where to chop up the sequence. We obtain in this way a partition \mathcal{A} in which, for each block $A \in \mathcal{A}$, $m_{\min(A)} > 0$, while $m_i = 0$ for $\min(A) \neq i \in A$ (where the minimum is with respect to $<$, and is unique since \mathcal{A} is an interval partition). We then assign a weight to block A by setting $v_A := m_{\min(A)}$. Likewise, we may encode a weighted partition as an integer vector m by assigning the weight of each block to its minimal site and 0 to all others. Since i_* is the unique minimal element of S , one always has $m_{i_*} > 0$. Explicitly, $m_{i_*} = v_A$ for the unique A that contains i_* and, for $i \neq i_*$,

$$m_i = \begin{cases} 0, & \text{if } \overleftarrow{i} \text{ and } i \text{ share a block,} \\ v_A \text{ for the unique } A \text{ containing } i, & \text{otherwise,} \end{cases}$$

with \overleftarrow{i} as in Definition 2.1. The new encoding allows us to rewrite H of Definition 7.4 in a convenient way, where we also take advantage of the formalism introduced in Section 4.

LEMMA 7.5. *Let H be as in Definition 7.4. For $m \in \mathbb{N}_0^S$ with $m_{i_*} > 0$, let $(\mathcal{A}(m), v(m))$ be the weighted partition associated with m , and define*

v	3	4	1	2
\mathcal{A}	{1, 2, 3},	{4, 5},	{6, 7, 8},	{9, 10}
m	0 0 3	0 4	1 0 0	2 0

Figure 12: Encoding a weighted partition (top) by an integer vector (bottom). The selected site is light brown.

$$\mathcal{H}(m, \mu) := H((\mathcal{A}(m), v(m)), \mu).$$

Then, one has

$$\mathcal{H}(m, \mu) = \prod_{i \in S} h(m_i, \mu)^{D_i} \quad (34)$$

where

$$h(k, \mu) := (1 - f(\mu))^k d(\mu) + (1 - (1 - f(\mu))^k) b(\mu) \quad (35)$$

for $k \neq 0$ and $h(0, \mu) := 1$. The factors are ordered nondecreasingly with respect to \preceq . \square

REMARK 7.6. When using the product sign \prod for products of elements of $\mathfrak{A}(X)$ indexed by S , we always understand the factors to be ordered nondecreasingly.

REMARK 7.7. At this point, it becomes clear that the special role of $D_{i_*} = S$ in the definition of the D_i (see Remark 2.2) makes perfect sense. Indeed, (34) shows that the contributions to the sequence at the root of the ASRG come from the ASGs associated to the ‘new’ tails D_i that are attached to the original one corresponding to $D_{i_*} = S$. This will become even more evident in the context of the initiation process, see Eq. (39) and Fig. 13.

Proof. Recall that, by the minimality of the selected site, we have $C_{i_*} = \emptyset, D_{i_*} = S$ and therefore $\mathcal{H}(m, \mu) = h(m_{i_*}, \mu)$ if $m_i = 0$ for all $i \neq i_*$. In all other cases, let i be a maximal site with $m_i \neq 0$. The definitions of \mathcal{H} and H then entail

$$\mathcal{H}(m, \mu) = \mathcal{H}(m', \mu)^{C_i} \otimes h(m_i, \mu)^{D_i} = \mathcal{H}(m', \mu) \boxtimes h(m_i, \mu)^{D_i},$$

where m' is obtained from m by setting m_i to zero. The claim then follows via induction. \square

The new encoding also allows us to represent the WPP as a collection of n independent Yule processes with initiation and resetting. In the case $s = 0$, this is similar to the representation of interval partitions in [5] in terms of the sets of breakpoints.

DEFINITION 7.8. A Yule process with initiation and resetting (YPIR) with branching rate $s > 0$, initiation rate $\varrho \geq 0$, and resetting rate $r \geq 0$ is a continuous-time Markov chain on $\mathbb{N}_{\geq 0}$ with transitions

$$\begin{aligned} \text{(Y)} \quad & k \rightarrow k + 1 \quad \text{at rate } sk \text{ for } k > 0, \\ \text{(I)} \quad & 0 \rightarrow 1 \quad \text{at rate } \varrho, \\ \text{(R)} \quad & k \rightarrow 1 \quad \text{at rate } r \text{ for } k > 0. \end{aligned}$$

Note that transition (R) is silent if $k = 1$.

Given the one-to-one correspondence between (\mathcal{A}, v) and m , it is then easy to see that (Σ, V) is equivalent to a collection $M = (M_i)_{i \in S}$ of independent YPIRs. Here, $M_{i_*} = (M_{i_*, t})_{t \geq 0}$ is a basic Yule process with branching rate $s > 0$, that is, the degenerate case of a YPIR with initiation and resetting rates $\varrho_{i_*} := r_{i_*} := 0$; for $i \neq i_*$, $M_i = (M_{i, t})_{t \geq 0}$ is a YPIR with branching rate s , initiation rate ϱ_i and resetting rate

$$r_i := \sum_{\ell \preceq i} \varrho_\ell; \tag{36}$$

note, in particular, that $r_i \geq \varrho_i$. Indeed, the equivalence is clear since the transitions of (Σ, V) and M can be matched in a unique way; compare Definitions 7.3 and 7.8. Note that r_i is the total rate at which i is separated from the selected site; it may be understood as the marginal recombination rate $r_i = \varrho_i^{\{i, i_*\}}$, cf. (52).

Note that the Yule process K (cf. Example 5.3) has the law of M_{i_*} . Let us recapitulate from [7] the duality for the pure selection equation, which is a slight extension of Example 5.3.

PROPOSITION 7.9. Let K be a Yule process with branching rate s . For $k \geq 1$ and $\omega \in \mathcal{P}(X)$, define $h(k, \omega)$ as in Eq. (35). Then,

$$h(k, \varphi_t(\mu)) = \mathbb{E}(h(K_t, \mu) \mid K_0 = k),$$

where φ is the selection semigroup.

Proof. Combining Eqs. (35), (21) and (28), one gets

$$\begin{aligned} h(k, \varphi_t(\mu)) &= (\mathbb{E}[(1 - f(\mu))^{K_t} \mid K_0 = 1])^k d(\varphi_t(\mu)) \\ &\quad + (1 - (\mathbb{E}[(1 - f(\mu))^{K_t} \mid K_0 = 1])^k) b(\varphi_t(\mu)) \\ &= \mathbb{E}(h(K_t, \mu) \mid K_0 = k), \end{aligned}$$

where the last step follows from the fact that a collection of m independent Yule processes, each started with a single line, is equivalent to a Yule process started with k lines. \square

Let us still postpone the duality result in the case with recombination to the next section, since the proof is most convenient on the basis of the initiation process.

7.3 THE INITIATION PROCESS.

Let us first gain some intuition by representing the duality function from Lemma 7.5 in terms of products of elements of the selection semigroup at various times. To this end, recall from Proposition 4.2 that $\varphi_t(\nu)$ is, for all ν and t , a convex combination of the conditional type distributions $d(\nu)$ and $b(\nu)$, and so is $h(k, \nu)$ for all $k \geq 1$, see Eq. (35). Since $f(\varphi_t(\nu))$ is strictly increasing in t (cf. Proposition 4.2), there exists, for all $k \geq 1$ and $s > 0$, a unique $\theta(k) \in \mathbb{R}$ such that $(1 - f(\nu))^k = 1 - f(\varphi_{\theta(k)}(\nu))$ and thus,

$$h(k, \nu) = \varphi_{\theta(k)}(\nu). \quad (37)$$

Note that $\theta(1) = 0$ since $h(1, \nu) = \nu = \varphi_0(\nu)$. Then, setting $\theta(0) := \Delta$ and $\varphi_\Delta(\nu) := 1$ for all ν (in line with $h(0, \cdot) = 1$ in Lemma 7.5), we can write, using Lemma 7.5,

$$\mathcal{H}(m, \nu) = \bigotimes_{i \in S} h(m_i, \mu)^{D_i} = \bigotimes_{i \in S} \varphi_{\theta(m_i)}(\mu)^{D_i} =: \mathcal{G}(\theta(m), \mu), \quad (38)$$

where $\theta(m) := (\theta(m_i))_{i \in S}$. More generally, this leads to the ansatz

$$\mathcal{G}(\theta, \nu) := \bigotimes_{i \in S} \varphi_{\theta_i}(\nu)^{D_i} \quad (39)$$

for a third (putative) duality function. Here, $\theta = (\theta_i)_{i \in S} \in \mathbb{R}_{\geq 0}^{i_*} \times (\mathbb{R}_{\geq 0} \cup \{\Delta\})^{S^*}$ and the symbol Δ is used to indicate that the factor is absent from the product. Recall that the factors in the product are ordered nondecreasingly w.r.t. \preceq and note that its value is the same for all such orderings since incomparable factors commute by virtue of being measures defined on projections of the type space X with respect to disjoint subsets of S .

Recall that m in (34) corresponds to a partition of S in which each block is weighted by a positive integer, counting the number of lines in the associated ASG (as part of an essential ASRG, see Section 6). Similarly, θ in Eq. (39) also encodes a partition of S (the role of 0 now being played by Δ), only this time, the blocks are not weighted by the number of lines in the associated ASGs, but by their runtimes (again, seen as part of an essential ASRG). In the sampling step, we average over all realisations of the ASG with the indicated runtime, and thus obtain \mathcal{G} from \mathcal{H} by replacing the factors $h(m_i, \nu)$ in $\mathcal{H}(m, \nu)$ by

$$\varphi_{\theta_i}(\nu) = \mathbb{E}[h(K_{\theta_i}, \nu) \mid K_0 = 1];$$

this will later make the connection to the transformation (37).

We now give an informal description of the *initiation process* Θ , which will take the role of the YPIR. It is a continuous-time Markov process, and its transition rates relate to that of the YPIR as follows. As Δ takes the role of 0, the transition (I) (initiation) in Definition 7.8 corresponds to a transition from Δ to 0. Similarly, as 0 takes the role of 1, a reset (R) (to 1) of the YPIR

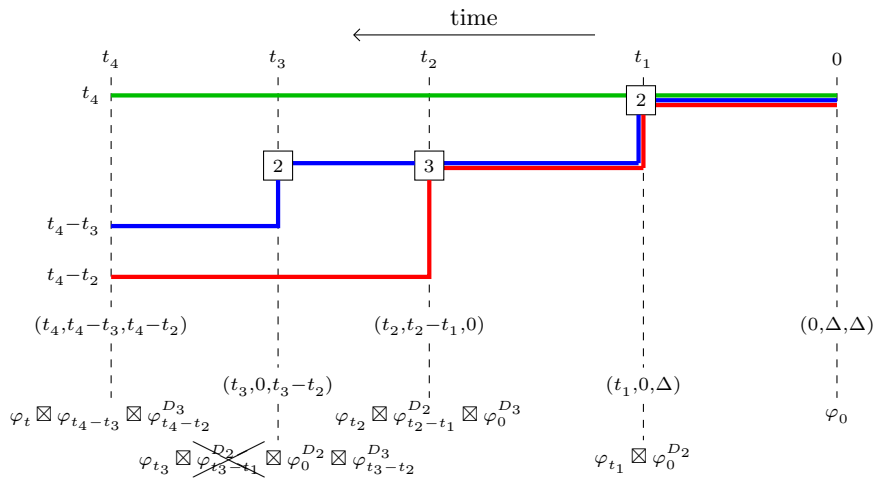


Figure 13: A realisation of the essential ASRG, where every ASG is collapsed into a single line. It describes the evolution of a partitioning process whose blocks are weighted by the time since the corresponding ASG was attached. The colours are as in Figure 10: green, blue and red for site 1,2 and 3; as before, the first site is selected. Below the graph, we indicate the evolution of the associated collection of initiation processes Θ . At the bottom, we see how the function $\mathcal{G}(\Theta_t, \cdot)$, defined in Eq. (39), evolves in time. Every factor corresponds to a different line, and attachment of a new line due to an i -recombination event corresponds to right multiplication by $\varphi_0^{D^i}$; subsequently, the time index in each factor evolves on its own. Notice the cancellation at time t_3 ; it corresponds to the discontinuation of the line at the box representing recombination and the reset of the second component of Θ , due to $\{2\} \cap D_2 = \{2\}$.

corresponds to a reset (to 0) of the initiation process. Keeping in mind that (Y) describes the branching of the ASG (and that we now only want to record its runtime), we replace these random jumps by a deterministic and continuous increase. Thus, Θ_t is either Δ , signifying that it has not yet been initiated, or its value is just the time that has passed since the last reset. Finally, when no resetting occurs, we have $\Theta_t = \Theta_0 + t$.

This can be condensed into the following definition; for an illustration, see Fig. 13.

DEFINITION 7.10. The *initiation process* with *initiation rate* $\varrho \geq 0$ and *resetting rate* $r \geq 0$ is the continuous-time Markov process with values in $\mathbb{R}_{\geq 0} \cup \{\Delta\}$ and generator mapping $u \in C^1(\mathbb{R})$ to \bar{u} , defined via

$$\begin{aligned}\bar{u}(t) &:= \dot{u}(t) + r(u(0) - u(t)) \quad \text{for } t \in \mathbb{R}_{\geq 0}, \\ \bar{u}(\Delta) &:= \varrho(u(0) - u(\Delta)).\end{aligned}\tag{40}$$

We define a collection $\Theta = (\Theta_i)_{i \in S}$ of independent initiation processes where $\Theta_i = (\Theta_{i,t})_{t \geq 0}$ has initiation rate ϱ_i and resetting rate r_i (cf. (36)). In particular, since $\varrho_{i_*} = r_{i_*} = 0$, all stochastic contributions in Eq. (40) vanish for this choice, and what remains is a purely deterministic drift, that is, $\Theta_{i_*,t} = t + \Theta_{i_*,0}$. We denote by \mathcal{L}_i the generator of Θ_i . Furthermore, $\mathcal{L} := \sum_{i \in S} \mathcal{L}_i$, where \mathcal{L}_i acts on the i -th component of the argument.

Note that Θ shares the parameters ϱ_i and r_i with M , but does not depend on s . Rather, for any given s , we will see that Θ and M are related at the level of an expectation. First, we prove the duality $(\omega, \Theta, \mathcal{G})$. From there, we recover (ω, M, \mathcal{H}) and, equivalently, $(\omega, (\Sigma, V), H)$.

PROPOSITION 7.11. For all $i \in S$, the YPIR M_i and the initiation process Θ_i of Definition 7.10 satisfy

$$\mathbb{E}(h(M_t, \nu) \mid M_0 = m) = \mathbb{E}(\varphi_{\Theta_t}(\nu) \mid \Theta_0 = \theta(m))$$

for all $m_i \in \mathbb{N}_0$ and $t \geq 0$.

Proof. It suffices to show that the left- and right-hand side of the statement solve the same initial value problem (with globally Lipschitz continuous right-hand side). By (37), the expressions agree at $t = 0$. It remains to be shown that

$$\mathcal{Q}_i h(\cdot, \nu)(m_i) = \mathcal{L}_i \varphi(\nu)(\theta(m_i)),$$

where \mathcal{Q}_i is the generator of M_i , and \mathcal{L}_i that of Θ_i . Comparing Definitions 7.8 and 7.10, it is obvious that the transitions from m_i to 1 in the YPIR (at rate ϱ_i if $m_i = 0$ and at rate r_i if $m_i > 0$) correspond to transitions to 0 in the initiation process (at rate ϱ_i if $\Theta_i = \Delta$ and at rate r_i if $\Theta_i \in \mathbb{R}_{\geq 0}$). The identity (37) then implies the equality of the corresponding contributions to the left and right-hand side, i.e.

$$h(1, \nu) - h(m_i, \nu) = \begin{cases} \varphi_0(\nu) - \varphi_\Delta(\nu) & \text{for } m_i = 0, \\ \varphi_0(\nu) - \varphi_{\theta(m_i)}(\nu) & \text{for } m_i > 0. \end{cases}$$

Furthermore, it is a direct consequence of Proposition 7.9 together with (37) that the time derivative corresponds to branching of the YPIR, that is,

$$\begin{aligned} \dot{\varphi}_{\theta(m_i)}(\nu) &= \frac{d}{dt} \mathbb{E}(h(K_t, \nu) \mid M_0 = m_i) \Big|_{t=0} \\ &= sm_i(h(m_i + 1, \nu) - h(m_i, \nu)) \end{aligned}$$

by the Kolmogorov backward equation for the Yule process. □

Returning now to \mathcal{H} and \mathcal{G} , we obtain immediately, by independence:

COROLLARY 7.12. *The families M and Θ of independent YPIRs and initiation processes satisfy*

$$\mathbb{E}(\mathcal{H}(M_t, \nu) \mid M_0 = m) = \mathbb{E}(\mathcal{G}(\Theta_t, \nu) \mid \Theta_0 = \theta(m))$$

for all $m \in \mathbb{N}_0^n$ and $t \geq 0$.

We are now set to state the main result of this section, the duality $(\omega, \Theta, \mathcal{G})$.

THEOREM 7.13. *Let Θ be the family of independent initiation processes introduced in Definition 7.10. Then, with \mathcal{G} as in (39), we have, for all $\nu \in \mathcal{P}(X)$ and all $\theta \in \mathbb{R}_{\geq 0}^{i_*} \times (\mathbb{R}_{\geq 0} \cup \{\Delta\})^{S^*}$,*

$$\mathcal{G}(\theta, \psi_t(\nu)) = \mathbb{E}(\mathcal{G}(\theta, \omega_t) \mid \omega_0 = \nu) = \mathbb{E}(\mathcal{G}(\Theta_t, \nu) \mid \Theta_0 = \theta),$$

where $\psi = (\psi_t)_{t \geq 0}$ is the flow of the SRE introduced in Definition 4.1.

Proof. The first equality is clear because ψ is deterministic. For the proof of the second equality (the actual duality relation), it will be useful to think of the solution of the SRE (17) as a deterministic Markov process with generator $\tilde{\Psi} = \tilde{\Psi}_{\text{sel}} + \tilde{\Psi}_{\text{rec}}$ given by

$$\begin{aligned} \tilde{\Psi}f(\nu) &:= \frac{d}{dt} f(\psi_t(\nu)) \Big|_{t=0} = \frac{d}{dt} f(\nu + t\Psi_{\text{sel}}(\nu) + t\Psi_{\text{rec}}(\nu)) \Big|_{t=0} \\ &= \frac{d}{dt} f(\nu + t\Psi_{\text{sel}}(\nu)) \Big|_{t=0} + \frac{d}{dt} f(\nu + t\Psi_{\text{rec}}(\nu)) \Big|_{t=0} \\ &=: \tilde{\Psi}_{\text{sel}}f(\nu) + \tilde{\Psi}_{\text{rec}}f(\nu) \end{aligned}$$

for all $f \in C^1(\mathcal{P}(X))$.

As in the proof of Proposition 7.11, we will show that the left and right-hand side satisfy the same initial value problem. As the values at $t = 0$ agree, see Eq. (37), it suffices to show that

$$\tilde{\Psi}\mathcal{G}(\theta, \cdot)(\nu) = \mathcal{L}\mathcal{G}(\cdot, \nu)(\theta) \tag{41}$$

for all $\nu \in \mathcal{P}(X)$ and all $\theta \in \mathbb{R}_{\geq 0}^{i_*} \times (\mathbb{R}_{\geq 0} \cup \{\Delta\})^{S^*}$. (Indeed, if (41) is satisfied, it trivially applies to all components of the \mathbb{R}^{2^n} -valued function \mathcal{G} and thus establishes duality also in our slightly extended sense; cf. Remark 7.2.) First of all, let us note that, since $\tilde{\Psi}$ is a differential operator, we have

$$\begin{aligned} & \tilde{\Psi}(\mathcal{G}(\theta, \cdot))(\nu) \\ &= \sum_{\substack{j \in S \\ \theta_j \neq \Delta}} \left(\prod_{j \not\prec i \in S \setminus j} \varphi_{\theta_i}(\nu)^{D_i} \right) (\tilde{\Psi}(\varphi_{\theta_j})(\nu))^{D_j} \prod_{j \prec i \in S} \varphi_{\theta_i}(\nu)^{D_i} \end{aligned} \tag{42}$$

by the product rule, where the underdot indicates the variable (in this case i) with respect to which the product is performed; note that since $\varphi_{\Delta}(\nu) = 1$, factors with $\theta_i = \Delta$ play no role. Hence, in order to evaluate the left-hand side of Eq. (41), we only need to compute $(\tilde{\Psi}(\varphi_{\theta_j})(\nu))^{D_j}$ for all $j \in S$ such that $\theta_j \neq \Delta$. Clearly,

$$(\tilde{\Psi}_{\text{sel}}(\varphi_{\theta_j})(\nu))^{D_j} = (\dot{\varphi}_{\theta_j}(\nu))^{D_j} \tag{43}$$

because φ is the flow of the pure selection equation. For the recombination part, we calculate

$$\begin{aligned} & (\tilde{\Psi}_{\text{rec}}(\varphi_{\theta_j})(\nu))^{D_j} \\ &= \left(\frac{d}{dh} \varphi_{\theta_j}(\nu + h\Psi_{\text{rec}}(\nu)) \Big|_{h=0} \right)^{D_j} \\ &= \left(\frac{d}{dh} \varphi_{\theta_j}(\nu \boxtimes (\mathbf{1} \boxplus h \sum_{\ell \in S^*} \varrho_{\ell}(\nu^{D_{\ell}} \boxminus \mathbf{1}))) \Big|_{h=0} \right)^{D_j} \\ &= \left(\varphi_{\theta_j}(\nu) \boxtimes \frac{d}{dh} (\mathbf{1} \boxplus h \sum_{\ell \in S^*} \varrho_{\ell}(\nu^{D_{\ell}} \boxminus \mathbf{1})) \Big|_{h=0} \right)^{D_j} \\ &= \sum_{\ell \in S^*} \varrho_{\ell}(\varphi_{\theta_j}(\nu)^{D_j} \boxtimes \nu^{D_{\ell} \cap D_j} - \varphi_{\theta_j}(\nu)^{D_j}) \\ &= \sum_{\substack{\ell \in S^* \\ \ell \preccurlyeq j}} \varrho_{\ell}(\varphi_0(\nu)^{D_j} - \varphi_{\theta_j}(\nu)^{D_j}) \\ &\quad + \sum_{\substack{\ell \in S^* \\ \ell \succ j}} \varrho_{\ell}(\varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_0(\nu)^{D_{\ell}} - \varphi_{\theta_j}(\nu)^{D_j}). \end{aligned} \tag{44}$$

Here, we have used Lemma 4.7 in the third step, and in the last that $\varphi_0(\nu) = \nu$ together with the fact that the sum over sites incomparable to j vanishes because $D_j \cap D_{\ell} = \emptyset$ if ℓ is incomparable to j . To simplify the first sum, we took advantage of the fact that $\ell \preccurlyeq j$ implies $D_j \subseteq D_{\ell}$ together with the cancellation rule from Proposition 4.5. Similarly, $\ell \succ j$ implies $D_{\ell} \subseteq D_j$, which simplifies the second sum. Inserting (44) and (43) into (42) and recalling Eq. (36), we

have shown so far that

$$\begin{aligned} \tilde{\Psi}\mathcal{G}(\theta, \cdot)(\nu) &= \sum_{\substack{j \in S \\ \theta_j \neq \Delta}} \left(r_j(\mathcal{G}((\theta_{<j}, 0, \theta_{>j}), \nu) - \mathcal{G}(\theta, \nu)) \right. \\ &\quad \left. + \frac{\partial}{\partial \theta_j} \mathcal{G}(\theta, \nu) + \sum_{\ell > j} \varrho_\ell(\mathcal{G}_{j,\ell}(\theta, \nu) - \mathcal{G}(\theta, \nu)) \right), \end{aligned}$$

where we use the obvious convention that $(\theta_{<j}, 0, \theta_{>j})$ is obtained from θ by setting θ_j to 0. Furthermore, $\mathcal{G}_{j,\ell}(\theta, \nu)$ (for $t_j \neq \Delta$ and $j \prec \ell$) arises from $\mathcal{G}(\theta, \nu)$ by inserting the factor $\varphi_0(\nu)^{D_\ell}$ at the immediate right of $\varphi_{\theta_j}(\nu)^{D_j}$. That is, if $\mathcal{G}(\theta, \nu)$ is of the form $\mathcal{G}(\theta, \nu) = \kappa \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \chi$, then

$$\mathcal{G}_{j,\ell}(\theta, \nu) = \kappa \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_0(\nu)^{D_\ell} \boxtimes \chi. \tag{45}$$

Hence, if we can show that

$$\sum_{\substack{j \in S \\ \theta_j \neq \Delta}} \sum_{\ell > j} \varrho_\ell(\mathcal{G}_{j,\ell}(\theta, \nu) - \mathcal{G}(\theta, \nu)) = \sum_{\substack{\ell \in S^* \\ \theta_\ell = \Delta}} \varrho_\ell(\mathcal{G}((\theta_{<\ell}, 0, \theta_{>\ell}), \nu) - \mathcal{G}(\theta, \nu)), \tag{46}$$

it follows that

$$\tilde{\Psi}\mathcal{G}(\theta, \cdot)(\nu) = \sum_{j \in S} \mathcal{L}_j \mathcal{G}((\theta_{<j}, \cdot, \theta_{>j}), \nu)(\theta_j) = \mathcal{L}\mathcal{G}(\cdot, \nu)(\theta).$$

To see Eq. (46), notice that, if $j \neq \max\{j' \preceq \ell : \theta_{j'} \neq \Delta\}$ (in particular, this is the case if $\theta_\ell \neq \Delta$), then $\mathcal{G}_{j,\ell}(\theta, \nu)$ is of the form

$$\kappa \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_0(\nu)^{D_\ell} \boxtimes \varphi_{\theta_{j'}}(\nu)^{D_{j'}} \boxtimes \chi' \tag{47}$$

for some $j' \preceq \ell$ due to the site ordering (cf. Remark 7.6), where $\chi = \varphi_{\theta_j} \boxtimes \chi'$. Since $j' \preceq \ell$ means $D_\ell \subseteq D_{j'}$, (47) is equal to

$$\kappa \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_{\theta_{j'}}(\nu)^{D_{j'}} \boxtimes \chi' = \mathcal{G}(\theta, \nu)$$

by the cancellation rule from Proposition 4.5.

If $j = \max\{j' \preceq \ell : \theta_{j'} \neq \Delta\}$, the factors in (45) are ordered strictly nondecreasingly w.r.t. \preceq , and no cancellations occur; hence we have $\mathcal{G}_{j,\ell}(\theta, \nu) = \mathcal{G}((\theta_{<\ell}, 0, \theta_{>\ell}), \nu)$. Thus, we have verified (46). \square

REMARK 7.14. (i) Another approach to recover Theorem 7.13 would be to prove the right multiplicativity for $h(k, \cdot)$ for $k \geq 1$ by the same argument as in Lemma 4.7, and to replace φ_t by $h(k, \cdot)$ in the proof of Theorem 7.13.

(ii) Note that the particular form of the selection term was not used in the proof of Theorem 7.13; the only property required was the second statement in Lemma 4.7. Therefore, the same procedure can be applied to

any single-locus model with linked neutral sites. Examples include the deterministic mutation-selection equation, for which the dual process can then be expressed as a collection of independent *pruned lockdown* ASGs [7, 11] that are initiated and reset at random.

- (iii) It is also instructive to relate the proof of Theorem 7.13 to the genealogical construction detailed above; see Figure 13. Recall that the factors $\varphi_{\theta_j}(\nu)^{D_j}$ in $\mathcal{G}(\theta, \nu)$ correspond to the different independent ASGs that make up the essential ASRG of Section 5, and which are ancestral to different sets of sites. At rate ρ_ℓ , $\ell \in S^*$, each such ASG is hit independently by an ℓ -box, at which a new ASG is started for the tail. This corresponds to right multiplication of $\varphi_{t_j}(\nu)^{D_j}$ by $\varphi_{t_\ell}(\nu)^{D_\ell}$. Recall that, for such a multiplication, we had to distinguish the three cases of j being either incomparable to ℓ , $\ell \preceq j$ and $\ell \succ j$. In the genealogical picture, these cases correspond to the recombination event being either ignored (if ℓ and j are incomparable, which entails that the ASG in question is only ancestral to sites in C_ℓ); a resetting event if $\ell \preceq j$, which means that the ASG is only ancestral to sites contained in D_ℓ ; or an initiation event if $\ell \succ j$, where a new ASG is initiated for the tail. \diamond

By Corollary 7.12 and (38), Theorem 7.13 also yields the duality of ω and M .

COROLLARY 7.15. *The family M of YPIRs and the solution ω of the SRE (17) are dual with respect to \mathcal{H} of (29), namely*

$$\mathbb{E}[\mathcal{H}(M_t, \nu) \mid M_0 = m] = \mathbb{E}[\mathcal{H}(m, \omega_t) \mid \omega_0 = \nu] = \mathcal{H}(m, \psi_t(\nu)) \quad (48)$$

for all $\nu \in \mathcal{P}(X)$ and all initial values $m \in \mathbb{N}_0^S$ with $m_{i_*} > 0$. Here, ψ is the deterministic flow introduced in Definition 4.1. \square

The following representations analogous to (30) for the solution of the selection-recombination differential equation are now immediate.

COROLLARY 7.16. *Let $\omega = \psi(\omega_0)$ be the solution of the SRE (17). Then, for all $t \geq 0$, we have the stochastic representations*

$$\begin{aligned} \omega_t &= \mathbb{E}[\mathcal{H}(M_t, \omega_0) \mid M_{i,0} = \delta_{i,i_*} \text{ for } i \in S] \\ &= \mathbb{E}[\mathcal{G}(\Theta_t, \omega_0) \mid \Theta_{i_*,0} = 0, \Theta_{i,0} = \Delta \text{ for } i \in S^*] \end{aligned}$$

with \mathcal{H} of (29) and \mathcal{G} of (39). That is, we average over all realisations of the WPP, starting from the trivial partition with weight one, as represented by the family of YPIRs, or the family of initiation processes, started in 0 for $i = i_*$ and started in Δ for $i \in S^*$.

8 THE EXPLICIT SOLUTION AND ITS LONG-TERM BEHAVIOUR

We have just seen that the solution of the SRE has a stochastic representation in terms of a collection of independent YPIRs. Their semigroups are easily expressed in terms of geometric distributions with random success probability.

PROPOSITION 8.1. *Let $i \in S$ and let M_i be a YPIR with branching rate $s > 0$, initiation rate $\varrho_i \geq 0$ and resetting rate $r_i \geq 0$. If $r_i > 0$, let T_i be a random variable with distribution $\text{Exp}(r_i)$; if $r_i = 0$, set $T_i := \infty$ for consistency. The Markov semigroup $p_i = (p_{i,t})_{t \geq 0}$ corresponding to M_i is then given by*

$$\begin{aligned} p_{i,t}(1, \cdot) &= \mathbb{E}[\text{Geom}(e^{-s(T_i \wedge t)})], \\ p_{i,t}(0, n_i) &= \int_0^\infty \varrho e^{-\varrho_i \tau} p_{i,t-\tau}(1, n_i) \, d\tau + \delta_{0, n_i} e^{-\varrho_i t}, \\ p_{i,t}(m_i, n_i) &= \int_0^\infty r_i e^{-r_i \tau} p_{i,t-\tau}(1, n_i) \, d\tau + e^{-r_i t} \text{NegBin}(m_i, e^{-st})(n_i), \end{aligned}$$

where $n_i \geq 0$, $m_i \geq 1$, and $\text{NegBin}(m_i, \sigma)$ is the negative binomial distribution with parameters m_i and σ , and we set $p_{i,t}(1, \cdot) \equiv 0$ for $t < 0$.

Proof. For the first formula, we argue as in the genealogical proof of Theorem 4.4. After the time of the last resetting event, which follows $\text{Exp}(r_i)$, the YPIR experiences no further resetting and hence has the law of a Yule process with branching rate s for the remaining time. The second and third formulae follow from the first by waiting the $\text{Exp}(\varrho_i)$ ($\text{Exp}(r_i)$)-distributed time until the process initiates (resets); recall that $\text{NegBin}(m_i, \sigma)$ is the distribution of the number of independent Bernoulli trials (with success probability σ) up to and including the m_i th success. In the degenerate case $r_i = 0$ and $\varrho_i = 0$, the statement reduces to

$$p_{i,t}(m_i, n_i) = \text{NegBin}(m_i, e^{-st})(n_i), \quad n_i \geq 0, \quad m_i \geq 1, \quad p_{i,t}(0, n_i) = \delta_{0, n_i},$$

which is just the semigroup of the ordinary Yule process. The consistency in the cases where only one of the parameters ϱ_i or r_i vanishes is seen just as easily. \square

Combining Proposition 8.1 with Corollary 7.16 yields a closed expression for the solution.

COROLLARY 8.2. *The solution of the SRE is given by*

$$\omega_t = p_{i_*, t} h(\cdot, \omega_0)^{D_{i_*}}(1) \boxtimes \prod_{i \in S^*} p_{i, t} h(\cdot, \omega_0)^{D_i}(0),$$

where $p_i = (p_{i,t})_{t \geq 0}$ is the semigroup of M_i as in Proposition 8.1.

This explicit representation allows us to investigate the long-term behaviour of the solution. We start with the asymptotics of the semigroup from Proposition 8.1.

COROLLARY 8.3. *As in Proposition 8.1, let p_i be the Markov semigroup of the YPIR M_i . Then, for all $m_i \geq 0$ and $n_i \geq 1$,*

$$p_{i, \infty}(n_i) := \lim_{t \rightarrow \infty} p_{i,t}(m_i, n_i) = \mathbb{E}[\text{Geom}(e^{-sT_i})(n_i)] = \alpha B(n_i, \alpha + 1), \quad (49)$$

where T_i follows $\text{Exp}(r_i)$, $\alpha := r_i/s_i$, and B denotes the beta function. If $\varrho_i = 0$, then $p_{i,t}(0, n_i) = \delta_{0, n_i}$, and Eq. (49) applies for $m_i > 0$.

Proof. Since M_i is irreducible, positive recurrent, and non-explosive (since it is stochastically dominated by a Yule process with branching rate s , which is non-explosive), it has a unique asymptotic distribution $p_{i,\infty}$ such that $p_{i,t}(m_i, \cdot)$ converges to $p_{i,\infty}$ for all initial conditions $m_i > 0$. To see that in fact $p_{i,\infty} = \mathbb{E}(\text{Geom}(e^{-sT_i}))$, it suffices in the case $\varrho_i > 0$ to simply let $t \rightarrow \infty$ in $p_{i,t}(1, \cdot)$ in Proposition 8.1; note that even when starting in 0, the process will jump to one almost surely. This is not the case if $\varrho_i = 0$; in this case, the process started in 0 will stay there forever whence the convergence to $\text{Geom}(e^{-sT_i})$ then only holds for strictly positive m_i . To see the last equality in Eq. (49), we write out the exponential mixture of geometric distributions explicitly and substitute $x = e^{-st}$ to obtain

$$\begin{aligned} p_{i,\infty}(n_i) &= \int_0^\infty r_i e^{-st} (1 - e^{-st})^{n_i-1} r_i e^{-r_i t} dt \\ &= \alpha \int_0^1 x^\alpha (1-x)^{n_i-1} dx = \alpha B(n_i, \alpha + 1) \end{aligned}$$

as claimed. □

REMARK 8.4. In the degenerate case $\varrho_i = r_i = 0$ (where the YPIR is an ordinary Yule process), there is no stationary distribution because the number of lines diverges almost surely. Nonetheless, one may still define (somewhat informally) $p_{i,\infty}(n_i) := 0$ for all $n_i \in \mathbb{N}$ together with $p_{i,\infty}(\infty) = 1$. ◇

Note that $p_{i,\infty}$ of (49) is the *Yule distribution* [51, 48] and may, for $n_i \geq 1$, be rewritten as

$$p_{i,\infty}(n_i) = \alpha B(n_i, \alpha + 1) = \frac{\alpha \Gamma(n_i) \Gamma(\alpha + 1)}{\Gamma(n_i + \alpha + 1)} = \frac{\alpha (n_i - 1)!}{\prod_{k=1}^{n_i} (k + \alpha)},$$

where Γ denotes the gamma function. In particular, $p_{i,\infty}(1) = \alpha / (\alpha + 1)$. For $\alpha \gg 1$, therefore, $p_{i,\infty}$ is close to a point measure on 1; whereas for $\alpha \ll 1$, $p_{i,\infty}(n)$ is close to α / n_i and puts substantial mass on large values, in line with intuition.

From the representation of the solution in Corollaries 8.1–8.3 together with Eq. (35) and $(1-x)^\infty = \delta_{x,0}$ for $x \in [0, 1]$, the long-term behaviour of the solution is now immediate.

COROLLARY 8.5. *Assuming that $\varrho_i > 0$ for all $i \in S^*$, we have*

$$\omega_\infty := \lim_{t \rightarrow \infty} \omega_t = \bigotimes_{i \in S} \pi_i \cdot \left((1 - \gamma_i(1 - f(\omega_0))) b(\omega_0) + \gamma_i(1 - f(\omega_0)) d(\omega_0) \right)$$

for all initial conditions $\omega_0 \in \mathcal{P}(X)$. As always, $f(\omega_0)$ is the initial frequency of the beneficial type. Furthermore, $\gamma_{i_*}(x) = \delta_{x,0}$ (in line with Remark 8.4 and $(1-x)^\infty = \delta_{x,0}$), and for $i \in S^*$, γ_i is the probability generating function of $p_{i,\infty}$, that is,

$$\gamma_i(x) := \sum_{n_i=1}^\infty p_{i,\infty}(n_i) x^{n_i}.$$

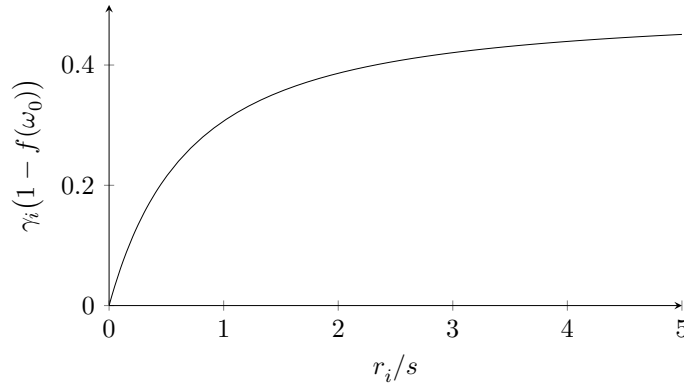


Figure 14: Asymptotic probability of site i being drawn from $\pi_i.d(\omega_0)$ as a function of r_i/s . As recombination becomes stronger, the asymptotic probability approaches the initial probability $1 - f(\omega_0) = 1/2$ assumed here.

REMARK 8.6. From Corollary 8.5, it is clear that $\gamma_i(1 - f(\omega_0))$ is the probability that site i is drawn from $\pi_i.d(\omega_0)$, or equivalently, that it is associated with $i_* = 1$ at equilibrium; see Figure 14 for an illustration of its parameter dependence. For a site i that is far away from i_* in the sense that its total rate of separation from i_* is large in comparison to the selection strength ($s \ll r_i$), the dynamics is close to that of the pure recombination equation; in particular, the marginals $\pi_i.\omega_t$ are approximately time invariant in line with the marginalisation consistency (53) of the pure recombination equation. Accordingly, the long-term behaviour is governed by $\gamma_i(x) \approx x$. In contrast, in the regime $s \gg r_i$, the behaviour is closer to that of the pure selection equation in that p_∞ places much weight on large values, which implies that $\gamma_i(x)$ is very small for small values of x , and the beneficial type prevails. \diamond

A MARGINALISATION CONSISTENCY

Let us consider the dynamics of the *marginal type distributions* under selection and recombination. For $A \subseteq S$, we define the *marginal recombinators* $R_i^A : \mathcal{P}(X_A) \rightarrow \mathcal{P}(X_A)$ by

$$R_i^A \nu := \nu^{A \cap C_i} \otimes \nu^{A \cap D_i} \tag{50}$$

for $i \in A \setminus i_*$, where C_i and D_i denote the head and tail for i as before.

REMARK A.1. Note that $\pi_A.R_i \omega = R_i^A \omega^A$ for all $A \subseteq S$ and $\omega \in \mathcal{P}(X)$, and $R_i^A = \text{id}$ if A is contained in either C_i or D_i , that is, if $\{C_i, D_i\}|_A = \{A\}$ (compare [6, Lemma 1]). \diamond

Consider now the marginal ω^A of ω . It is well known [6, Proposition 6] that

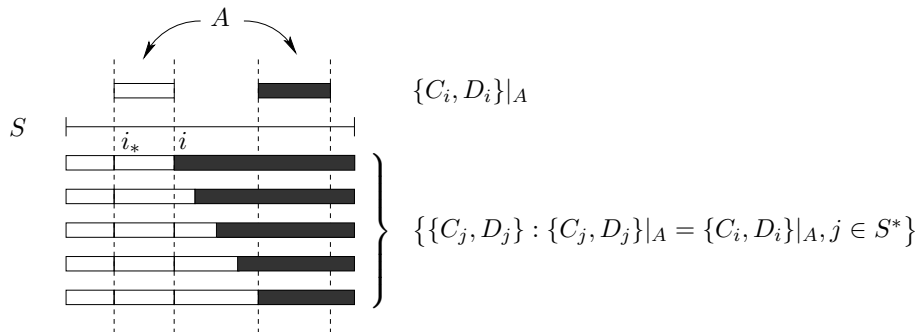


Figure 15: The partitions in Eq. (52) that define the marginal recombination rates.

ω^A satisfies the *marginalised recombination equation*

$$\dot{\omega}_t^A = \pi_A \cdot \Psi_{\text{rec}}(\omega_t) = \sum_{i \in A \setminus i_*} \varrho_i^A (R_i^A \omega_t^A - \omega_t^A) =: \Psi_{\text{rec}}^A(\omega_t^A) \tag{51}$$

with initial condition ω_0^A and *marginal recombination rates*

$$\varrho_i^A := \sum_{\substack{j \in S^* \\ \{C_j, D_j\}|_A = \{C_i, D_i\}|_A}} \varrho_j \quad \text{for all } i \in A \setminus i_*; \tag{52}$$

see Figure 15 for an illustration. In particular,

$$\dot{\omega}_t^{\{i\}} = 0 \quad \text{for } i \in S^* \tag{53}$$

since $R_i^{\{i\}} = \text{id}$. Eq. (51) follows from Remark A.1, the linearity of π_A , and Eq. (52).

Unfortunately, this property does not generalise to the selective case. The reason is that Ψ_{sel} also depends on the proportion $f(\omega_t) = \omega_t^{\{i_*\}}(0)$ of fit individuals and that we lose this information by projecting onto a factor with respect to a subset of S not containing i_* . When A *does* contain i_* , however, we clearly have

$$f(\nu) = f^A(\nu^A) \quad \text{for any } \nu \in \mathcal{P}(X), \tag{54}$$

where f^A is defined analogously to (7), but with S replaced by A . Moreover, the selection operator (8) acts consistently on subsystems that contain the selected site, that is, $\pi_A \cdot F\nu = F^A \nu^A$ for $A \ni i_*$, where the marginalised selection operator is given by $F^A(\nu^A)(x_A) := (1 - x_{i_*})\nu^A(x_A)$. We can thus define $\Psi_{\text{sel}}^A : \mathcal{P}(X_A) \rightarrow \mathcal{P}(X_A)$ via

$$\Psi_{\text{sel}}^A(\nu_A) := s(F^A - f^A(\nu_A))\nu_A$$

such that

$$\pi_A \cdot \Psi_{\text{sel}}(\nu) = \Psi_{\text{sel}}^A(\nu^A) \quad \text{for } A \ni i_* \quad \text{and all } \nu \in \mathcal{P}(X).$$

Combining this with (51), we obtain the following result.

THEOREM A.2 (marginalisation consistency of the SRE). *Let ω be the solution of the initial value problem for the SRE (17) with initial condition ω_0 . Let $A \subseteq S$ contain i_* . Then, the marginal $\omega^A := (\omega_t^A)_{t \geq 0}$ solves the marginal SRE,*

$$\dot{\omega}_t^A = s(F^A(\omega_t^A) - f^A(\omega_t^A)\omega_t^A) + \sum_{i \in A \setminus i_*} \varrho_i^A (R_i^A \omega_t^A - \omega_t^A),$$

with initial condition ω_0^A and marginal recombination rates (52). In particular, ω^A is independent of all ϱ_i with i such that $\{C_i, D_i\}|_A = \{A\}$; or equivalently, with i such that $i \succ j$ for all $j \in A$ comparable to i . \square

REMARK A.3. The problem of marginalisation (in)consistency was already observed by Ewens and Thomson [21] in 1977 for the discrete-time SRE; see also the review in [14, pp. 69–72]. For Theorem A.2 to hold, the assumption that A contains the selected site is crucial: It is otherwise impossible to find a closed expression for the projection of the selective part in (17) in terms of the marginal measure, because we lose the information about the proportion of fit individuals in the case that $i_* \notin A$. It is indeed a common pitfall to assume that Theorem A.2 holds for arbitrary A . This is also implicit in [3]; see the corresponding erratum. \diamond

There is an interesting connection between marginalisation consistency and the recursive solution of the SRE of Theorem 4.4. Applying Theorem A.2 to $A = \{i_*\}$ shows that the marginal type frequency at the selected site is unaffected by recombination. More generally, consider the set $L^{(k)} := \{i_0 = i_*, i_1, \dots, i_k\}$ and note that $L^{(k)} \setminus i_*$ is exactly the set of recombination sites that are considered up to and including the k -th iteration. Obviously, marginalisation consistency holds for $L^{(k)}$ for all $0 \leq k \leq n-1$. Since $\varrho_i^{L^{(k)}} = \varrho_i$ for $i \in L^{(k)} \setminus i_*$, Remark A.1 and Eq. (52) together with Definition 4.1 give

$$\pi_{L^{(k)}} \cdot \dot{\omega}_t = \pi_{L^{(k)}} \cdot \sum_{i \in L^{(k)} \setminus i_*} \varrho_i (R_i \omega_t - \omega_t) = \pi_{L^{(k)}} \cdot \Psi_{\text{rec}}^{(k)}(\omega_t) = \pi_{L^{(k)}} \cdot \dot{\omega}_t^{(k)},$$

and so $\pi_{L^{(k)}} \cdot \omega_t^{(k)} = \pi_{L^{(k)}} \cdot \omega_t$. This implies that if one is only interested in the marginal with respect to $L^{(k)}$, then one may stop the iteration after the k th step.

ACKNOWLEDGEMENTS

It is a pleasure to thank M. Baake for enlightening discussions and substantial suggestions to improve the manuscript, and M. Möhle for an interesting

exchange about duality. We also thank L. Esercito and S. Hummel for stimulating discussions, as well as an anonymous referee for helpful comments. This work was supported by the German Research Foundation (DFG), within the SPP 1590 and the CRC 1283 (project C1).

REFERENCES

- [1] E. Akin *The Geometry of Population Genetics*, Springer, New York (1979).
- [2] E. Akin, Cycling in simple genetic systems, *J. Math. Biol.* 13 (1982), 305–324.
- [3] E. Baake and M. Baake, An exactly solved model for recombination, mutation and selection, *Can. J. Math.* 55 (2003), 3–41; and erratum, *Can. J. Math.* 60 (2008), 264–265.
- [4] E. Baake and M. Baake, Haldane linearisation done right: Solving the nonlinear recombination equation the easy way, *Discr. Cont. Dyn. Syst. A* 36 (2016), 6645–6656.
- [5] E. Baake and M. Baake, Ancestral lines under recombination, in: *Probabilistic Structures in Evolution*, E. Baake and A. Wakolbinger (eds.), EMS Press, Berlin (2021), 365–382.
- [6] E. Baake, M. Baake and M. Salamat, The general recombination equation in continuous time and its solution, *Discr. Cont. Dyn. Syst. A* 36 (2016), 63–95.
- [7] E. Baake, F. Cordero and S. Hummel, A probabilistic view on the deterministic mutation-selection equation: Dynamics, equilibria, and ancestry via individual lines of descent, *J. Math. Biol.* 77 (2018), 795–820.
- [8] E. Baake, M. Esser and S. Probst, Partitioning, duality, and linkage disequilibria in the Moran model with recombination, *J. Math. Biol.* 73 (2016), 161–197.
- [9] E. Baake, U. Lenz and A. Wakolbinger, The common ancestor type distribution of a Λ -Wright–Fisher process with selection and mutation, *Electron. Commun. Probab.* 21 (2016), 1–16.
- [10] F. B. Christiansen, *Population Genetics of Multiple Loci*, Wiley (1999).
- [11] E. Baake and A. Wakolbinger, Lines of descent under selection, *J. Stat. Phys.* 172 (2018), 156–174.
- [12] A. Bhaskar and Y. S. Song, Closed-form asymptotic sampling distributions under the coalescent with recombination for an arbitrary number of loci, *Adv. Appl. Probab.* 44 (2012), 391–407.

- [13] S. Bossert and P. Pfaffelhuber, The fixation probability and time for a doubly beneficial mutant, *Stoch. Proc. Appl.* 128 (2018), 4018–4050.
- [14] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation*, Wiley, Chichester (2000).
- [15] F. Cordero, The deterministic limit of the Moran model: a uniform central limit theorem, *Markov Proc. Relat. Fields* 23 (2017), 313–324.
- [16] F. Cordero, Common ancestor type distribution: A Moran model and its deterministic limit, *Stoch. Proc. Appl.* 127 (2017), 590–621.
- [17] P. Donnelly and T.G. Kurtz, Genealogical processes for Fleming–Viot models with selection and recombination, *Ann. Appl. Probab.* 9 (1999), 1091–1148.
- [18] R. Durrett, *Probability Models for DNA Sequence Evolution*, 2nd ed., Springer, New York (2008).
- [19] M. Esser, *Recombination Models Forward and Backward in Time*, Dissertation, Bielefeld University (2016); urn:nbn:de:0070-pub-29102790.
- [20] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, Hoboken (1986), reprint (2005).
- [21] W.J. Ewens and G. Thomson, Properties of equilibria in multi-locus genetic systems, *Genetics* 87 (1977), 807–819.
- [22] W. Feller, *An Introduction to Probability Theory and its Applications, Vol. I*, 3rd ed., Wiley, New York (1968).
- [23] P. Fearnhead, Ancestral processes for non-neutral models of complex diseases, *Theor. Popul. Biol.* 63 (2003), 115–130.
- [24] R. C. Griffiths and P. Marjoram, Ancestral inference from samples of DNA sequences with recombination, *J. Comput. Biol.* 3 (1996), 479–502.
- [25] R. C. Griffiths and P. Marjoram, An ancestral recombination graph. in: *Progress in Population Genetics and Human Evolution*, P. Donnelly and S. Tavaré (eds.), Springer, New York (1997), pp. 257–270.
- [26] A. Hastings, Stable cycling in discrete-time genetic models, *Proc. Natl. Acad. Sci. USA* 78 (1981), 7224–7225.
- [27] J. Hofbauer, The selection-mutation equation, *J. Math. Biol.* 23 (1985), 41–53.
- [28] R. R. Hudson, Properties of a neutral allele model with intragenic recombination, *Theor. Popul. Biol.* 23 (1983), 183–201.

- [29] P. A. Jenkins, P. Fearnhead, and Y. S. Song, Tractable diffusion and coalescent processes for weakly correlated loci, *Electron. J. Probab.* 20 (2015), 1–26.
- [30] H. S. Jennings, The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage, *Genetics* 2 (1917), 97–154.
- [31] S. Karlin, General two-locus selection models: Some objectives, results and interpretations, *Theor. Pop. Biol.* 7 (1975), 364–398.
- [32] M. Kimura, A model of a genetic system which leads to closer linkage by natural selection, *Evolution* 10 (1956), 278–287.
- [33] V. Kirzhner and Y. Lyubich, Multilocus dynamics under haploid selection, *J. Math. Biol.* 35 (1997), 391–408.
- [34] S. M. Krone and C. Neuhauser, Ancestral processes with selection, *Theor. Popul. Biol.* 51 (1997), 210–237.
- [35] S. Jansen and N. Kurt, On the notion(s) of duality for Markov processes, *Probab. Surveys* 11 (2014), 59–120.
- [36] A. Lambert, V. Miró Pina, and E. Schertzer, Chromosome painting: How recombination mixes ancestral colors, *Ann. Appl. Probab.* 31 (2021), 826–864.
- [37] S. Lessard and A. R. Kermany, Fixation probability in a two-locus model by the ancestral recombination-selection graph, *Genetics* 190 (2012), 691–707.
- [38] T. M. Liggett, *Continuous Time Markov Processes: An Introduction*, Amer. Math. Soc., Providence, RI (2010).
- [39] Y. I. Lyubich, *Mathematical Structures in Population Genetics*, Springer, Berlin (1992).
- [40] D. McHale and G. A. Ringwood, Haldane linearisation of baric algebras, *J. London Math. Soc.* 28 (1983), 17–26.
- [41] M. Möhle, The concept of duality and applications to Markov processes arising in neutral population genetics models, *Bernoulli* 5 (1999), 761–777.
- [42] M. Möhle, Duality and cones of Markov processes and their semigroups, *Markov Process. Relat. Fields* 19 (2013), 149–162.
- [43] T. Nagylaki, J. Hofbauer, and P. Brunovský, Convergence of multilocus systems under weak epistasis or weak selection, *J. Math. Biol.* 38 (1999), 103–133.

- [44] J. Maynard Smith and J. Haigh, The hitch-hiking effect of a favourable gene, *Gen. Res.* 23 (1974), 23–35.
- [45] J.R. Norris, *Markov Chains*, Cambridge University Press, New York (1997), reprint (2005).
- [46] M. Pontz, J. Hofbauer and R. Bürger, Evolutionary dynamics in the two-locus two-allele model with weak selection, *J. Math. Biol.* 76 (2018), 151–203.
- [47] R. B. Robbins, Some applications of mathematics to breeding problems III, *Genetics* 3 (1918), 375–389.
- [48] H. A. Simon, On a class of skew distribution functions, *Biometrika* 43 (1955), 423–440.
- [49] W. Stephan, Y. S. Song and C. H. Langley, The hitchhiking effect on linkage disequilibrium between linked neutral loci, *Genetics* 172 (2006), 2647–2663.
- [50] A. Stuart and K. J. Ord, *Kendall's Advanced Theory of Statistics*, Vol. 1 (Distribution Theory), 5th ed., Wiley, Chichester (1994).
- [51] G. U. Yule, A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S., *Philos. Trans. Roy. Soc. London B* 13 (1924), 21–87.

Frederic Alberti
Faculty of Mathematics
Bielefeld University
Postbox 100131
33501 Bielefeld
Germany
falberti@math.uni-
bielefeld.de

Ellen Baake
Faculty of Technology
Bielefeld University
Postbox 100131
33501 Bielefeld
Germany
ebaake@math.uni-
bielefeld.de

