

The fixation time of a strongly beneficial allele in a structured population

Andreas Greven¹ Peter Pfaffelhuber² Cornelia Pokalyuk³
Anton Wakolbinger⁴

Abstract

For a beneficial allele which enters a large unstructured population and eventually goes to fixation, it is known that the time to fixation is approximately $2 \log(\alpha)/\alpha$ for a large selection coefficient α . For a population that is distributed over finitely many colonies, with migration between these colonies, we detect various regimes of the migration rate μ for which the fixation times have different asymptotics as $\alpha \rightarrow \infty$.

If μ is of order α , the allele fixes (as in the spatially unstructured case) in time $\sim 2 \log(\alpha)/\alpha$. If μ is of order α^γ , $0 \leq \gamma \leq 1$, the fixation time is $\sim (2 + (1 - \gamma)\Delta) \log(\alpha)/\alpha$, where Δ is the number of migration steps that are needed to reach all other colonies starting from the colony where the beneficial allele appeared. If $\mu = 1/\log(\alpha)$, the fixation time is $\sim (2 + S) \log(\alpha)/\alpha$, where S is a random time in a simple epidemic model.

The main idea for our analysis is to combine a new moment dual for the process conditioned to fixation with the time reversal in equilibrium of a spatial version of Neuhauser and Krone's ancestral selection graph.

Keywords: interacting Wright–Fisher diffusions; ancestral selection graph; branching process approximation.

AMS MSC 2010: Primary 92D15, Secondary 60J80; 60J85; 92D10.

Submitted to EJP on March 2, 2014, final version accepted on September 19, 2016.

Supersedes arXiv:1402.1769.

¹ Department Mathematik, Universität Erlangen-Nürnberg, Cauerstr. 11, 91058 Erlangen, Germany.
E-mail: greven@math.uni-erlangen.de

² Abteilung für Mathematische Stochastik, Albert-Ludwigs University of Freiburg, Eckerstraße 1,
79104 Freiburg, Germany. E-mail: p.p@stochastik.uni-freiburg.de

³ Institut für Mathematik, Goethe-Universität, Fachbereich 12, 60054 Frankfurt am Main, Germany.
E-mail: pokalyuk@math.uni-frankfurt.de

⁴ Institut für Mathematik, Goethe-Universität, Fachbereich 12, 60054 Frankfurt am Main, Germany.
E-mail: wakolbin@math.uni-frankfurt.de

1 Introduction

The goal of this paper is the asymptotic analysis of the time which it takes for a single strongly beneficial mutant to eventually go to fixation in a spatially structured population. The beneficial allele and the wildtype will be denoted by \mathcal{B} and \mathcal{b} , respectively. The evolution of type frequencies is modelled by a $[0, 1]^d$ -valued diffusion process $\mathcal{X} = (\underline{X}(t))_{t \geq 0}$, $\underline{X}(t) = (X_i(t))_{i=1, \dots, d}$, where $d \in \{2, 3, \dots\}$ denotes the number of colonies and $X_i(t)$ stands for the frequency of the beneficial allele \mathcal{B} in colony i at time t . The dynamics accounts for resampling, selection and migration. The process \mathcal{X} is started at time 0 by an entrance law from $\underline{0} := (0, \dots, 0)$ and is conditioned to eventually hit $\underline{1} := (1, \dots, 1)$.

Models of this kind are building blocks for more complex ones that are used to obtain predictions for genetic diversity patterns under various forms of selection. Indeed, together with the strongly beneficial allele, neutral alleles at physically linked genetic loci also have the tendency to go to fixation, provided these loci are not too far from the selective locus under consideration. This so-called genetic hitchhiking was first modelled by Maynard Smith and Haigh [16]. A synonymous notion is that of a selective sweep, which alludes to the fact that, after fixation of the beneficial allele \mathcal{B} , neutral variation has been swept from the population. Important tools were developed from these patterns to locate targets of selection in a genome and quantify the role of selection in evolution, see e.g. reviews in Sabeti et al. [22], Nielsen [19], Thornton et al. [28].

The process of fixation of a strongly beneficial mutant in the *panmictic* (i.e. unstructured) case has been studied using a combination of techniques from diffusion processes and coalescent processes in a random background; see e.g. Kaplan et al. [12], Stephan et al. [27], Schweinsberg and Durrett [23], Etheridge et al. [6]. However, since the analytical tools applied in these papers rely on the theory of *one-dimensional* diffusion processes, the extension of these results to a spatially structured situation is far from straight-forward.

The starting point for the tools developed in this paper is the ancestral selection graph (ASG) of Neuhauser and Krone [18]. This process has been introduced in order to study the genealogy under models including selection. Although the ASG can in principle be used for an arbitrary strength of selection, it has been employed mainly for models of weak selection, since then the resulting genealogy is close to a neutral one. However, Wakeley and Sargsyan [29] have used the ASG for strong balancing selection and Pfaffelhuber and Pokalyuk [21] have shown how to use the ASG in order to re-derive classical results for selective sweeps in a panmictic population. In our present work a *spatial version of the ASG* is the tool of choice which carries over from the panmictic to the structured case, thus extending the techniques developed in [21] and leading to new results for the spatially structured case. The key idea here is to employ the *equilibrium ASG* in a “paintbox representation” of the (fixed time) distributions of the type frequency process conditioned to eventual fixation, and then use *time reversal* of the equilibrium ASG to obtain an object accessible to the asymptotic analysis.

The fixation process in a structured population under selection has been the object of study before. Slatkin [25] and Whitlock [30] give heuristic results and comparisons to the panmictic case. While the former paper only gives results for strong selection but very weak migration, the latter study gives a comparison to the panmictic case and studies the question which parameters should be used in the panmictic setting in order to approximate fixation probabilities and fixation times for structured populations. In Kim and Maruki [13] the above studies are extended by analysing in addition the expected heterozygosity of linked neutral loci in the case of frequent migration for populations structured according to a circular stepping-stone model, see also Remark 2.7

below. Hartfield [11] gives a more thorough analysis of the fixation times for large selection/migration ratios in general stepping-stone populations based on the assumption that in each colony the beneficial mutation spreads before migrating.

Our investigation will provide rigorous results on fixation times for structured populations, and will detect the corresponding *regimes of relative migration/selection speed*.

Outline of the paper. After introducing the model in Section 2 we formulate our main results. These concern the existence of solutions and the structure of the set of solutions of the system of SDEs specified in our model (Theorem 1) and the asymptotics of the fixation times for a strongly beneficial allele \mathcal{B} in a structured population (Theorem 2). For the panmictic case (i.e. $d = 1$), it is well-known that the fixation time, for a large selection coefficient α , is approximately $2 \log(\alpha)/\alpha$. As it turns out, the time-scale of $\log(\alpha)/\alpha$ applies in our spatial setting as well. However, population structure may slow down the fixation process. We study this deceleration for various regimes of the migration rate μ . A spatial version of the ancestral selection graph is introduced in Section 3, and its role in the analysis of the fixation probability and the fixation time by the method of duality is clarified. This leads to a proof of Theorem 1 in Sec. 3.10, and to the key Proposition 3.1 which relates the asymptotic distribution of the fixation time of the Wright-Fisher system to that of a marked particle system. Based on the latter, the proof of Theorem 2 is completed in Sec. 4.

2 Model and main results

We consider solutions $\mathcal{X} = (\underline{X}(t))_{t \geq 0}$, $\underline{X}(t) = (X_1(t), \dots, X_d(t)) \in [0, 1]^d$, of the system of *interacting Wright-Fisher diffusions*

$$dX_i = \left(\alpha X_i(1 - X_i) + \mu \sum_{j=1}^d b(i, j)(X_j - X_i) \right) dt + \sqrt{\frac{1}{\rho_i} X_i(1 - X_i)} dW_i, \quad i = 1, \dots, d \quad (2.1)$$

for independent Brownian motions W_1, \dots, W_d . Here, α and μ are positive constants (the *selection* and *migration coefficient*), and $b(i, j)$, $i, j = 1, \dots, d$, $i \neq j$, are non-negative numbers (the *backward migration rates*) that constitute an irreducible rate matrix \underline{b} whose unique equilibrium distribution has the weights ρ_1, \dots, ρ_d (which stand for the relative population sizes of the colonies). It is well-known (see e.g. [4]) that the system (2.1) has a unique weak solution.

Equation (2.1) models the evolution of the relative frequencies of the beneficial allele at the various colonies, assuming a *migration equilibrium* between the colonies. The “gene flow” from colony i to colony j is $\rho_i \mu a(i, j) = \rho_j \mu b(j, i)$; here, $\underline{a} = (a(i, j))$ with

$$a(i, j) = \frac{\rho_j}{\rho_i} b(j, i) \quad (2.2)$$

is the matrix of *forward migration rates*.

Remark 2.1 (Limit of Moran models). We note in passing that the process \mathcal{X} arises as the weak limit (as $N \rightarrow \infty$) of a sequence of structured two-type Moran models with N individuals. The dynamics of this Moran model is local pairwise resampling with rates $1/\rho_i$, selection with coefficient α (i.e. offspring from every beneficial line in colony i replaces some line in the same colony at rate α ; note that this is the same as selection events which occur at rate $s := \frac{\alpha}{N}$ for each (ordered) pair of particles) and migration with rates $\mu a(i, j)$ per line. Considering now the relative frequencies of the beneficial type at the various colonies and letting $N \rightarrow \infty$ gives (2.1). Here, our assumption that

(ρ_i) constitutes an equilibrium for the migration ensures that we are in a *demographic equilibrium* with asymptotic colony sizes $\rho_i N$ (otherwise the ρ_i, ρ_j in the formulas would have to be replaced by time-dependent intensities).

We define the fixation time of \mathcal{X} as

$$T_{\text{fix}} := \inf\{t > 0 : \underline{X}(t) = \underline{1}\}. \tag{2.3}$$

The fixation probability of the system (2.1), started in $\underline{X}(0) = \underline{x}$, is well-known (see [17]). In Corollary 3.10 we will provide a new proof for the formula

$$\mathbf{P}_{\underline{x}}(T_{\text{fix}} < \infty) = \frac{1 - e^{-2\alpha(x_1\rho_1 + \dots + x_d\rho_d)}}{1 - e^{-2\alpha}}. \tag{2.4}$$

Since fixation of the beneficial allele, $\{T_{\text{fix}} < \infty\}$, is an event in the terminal σ -algebra of \mathcal{X} , conditioning on this event leads to an h -transform of (2.1) which turns out to be given by the system of SDEs

$$dX_i^* = \left(\alpha X_i^*(1 - X_i^*) \coth\left(\alpha \sum_{j=1}^d X_j^* \rho_j\right) + \mu \sum_{j=1}^d b(i, j)(X_j^* - X_i^*) \right) dt + \sqrt{\frac{1}{\rho_i} X_i^*(1 - X_i^*)} dW_i \tag{2.5}$$

for $i = 1, \dots, d$, with $\coth(x) = \frac{e^{2x} + 1}{e^{2x} - 1}$. The uniqueness of the solution of (2.1) carries over to that of (2.5) as long as $\underline{x} \neq \underline{0}$. For $\underline{x} = \underline{0}$, the right hand side of (2.5) is not defined, and we have to talk about *entrance laws from $\underline{0}$* for solutions of (2.5) in this case.

Definition 2.2 (Entrance law from $\underline{0}$). *Let $((\underline{X}^*(t))_{t>0}, \mathbf{P})$ with $\underline{X}^*(t) = (X_1^*(t), \dots, X_d^*(t))$ be a solution of (2.5) such that $\underline{X}^*(t) \neq \underline{0}$ for $t > 0$ and $\underline{X}^*(t) \xrightarrow{t \rightarrow 0} \underline{0}$ in probability. Then, the law of \underline{X}^* under \mathbf{P} is called an entrance law from $\underline{0}$ for the dynamics (2.5).*

The following is shown in Section 3.10.

Theorem 1. a) For $\underline{x} \in [0, 1]^d \setminus \{\underline{0}\}$, the system (2.5) has a unique weak solution.
 b) Every entrance law from $\underline{0}$ is a convex combination of d extremal entrance laws from $\underline{0}$, which we denote by $\mathbf{P}_{\underline{0}}^i(\mathcal{X}^* \in \cdot)$, with $(\mathcal{X}^*, \mathbf{P}_{\underline{0}}^i)$ arising as the limit in distribution of $(\mathcal{X}^*, \mathbf{P}_{\varepsilon \underline{e}_i})$ as $\varepsilon \rightarrow 0$, where \underline{e}_i is the vector whose i -th component is 1 and whose other components are 0.

Remark 2.3 (Interpretation of the extremal solutions). We call $(\mathcal{X}^*, \mathbf{P}_{\underline{0}}^i)$ the *solution with the founder in colony i* . In intuitive terms the case $\underline{x} = \underline{0}$ corresponds to the beneficial allele \mathcal{B} being present in a copy number which is too low to be seen in a very large population, i.e. on a macroscopic level. In this case, since the process is conditioned on fixation, there is exactly one individual – called founder – which will be the ancestor of all individuals at the time of fixation. This intuition is made precise in a picture involving duality, see Section 3.8. The d different entrance laws from $\underline{0}$ belonging to (2.5) correspond to the d different possible geographic locations of the founder.

Before stating our main result on the fixation time of the system (2.5) we fix some notation and formulate one more definition.

Remark 2.4 (Notation). To facilitate notation we will use Landau symbols. For functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$, we write (i) $f = \mathcal{O}(g)$ as $x \rightarrow x_0 \in \overline{\mathbb{R}}$ if $\limsup_{x \rightarrow x_0} |f(x)/g(x)| < \infty$, (ii) $f \in \Theta(g)$ if and only if $f \in \mathcal{O}(g)$ and $g \in \mathcal{O}(f)$ and (iii) $f \sim g$ as $x \rightarrow x_0$ if and only if $f(x)/g(x) \xrightarrow{x \rightarrow x_0} 1$, (iv) $f = o(g)$ as $x \rightarrow x_0 \in \overline{\mathbb{R}}$ if $\limsup_{x \rightarrow x_0} |f(x)/g(x)| = 0$. We write \Rightarrow for convergence in distribution and \rightarrow_p for convergence in probability.

In the case of a single colony ($d = 1$) we have $T_{\text{fix}} \sim 2 \log \alpha / \alpha$ as $\alpha \rightarrow \infty$. Indeed, it is well known that in this case the conditioned diffusion (2.5) can be separated into three phases [6]: the beneficial allele \mathcal{B} first has to increase up to a (fixed) small $\varepsilon > 0$. This phase lasts a time $\sim \log(\alpha) / \alpha$. In the second phase, the frequency increases to $1 - \varepsilon$ in time of order $1 / \alpha$ which is short as compared to the first and third phase. In the third phase, it takes still about time $\log(\alpha) / \alpha$ until the allele finally fixes in the population.

Definition 2.5 (Two auxiliary epidemic processes). *Let \underline{a} be the matrix of forward migration rates and let $G = (V, E)$ be the (connected) graph with vertex set $1, \dots, d$ and edge set $E := \{(i, j) : a(i, j) > 0\}$. We need two auxiliary processes in order to formulate our theorem.*

1. For $\gamma \in [0, 1]$ and $\iota \in \{1, \dots, d\}$, consider the (deterministic) process $\mathcal{I}^{\iota, \gamma} := \mathcal{I}^\iota = (\underline{I}^\iota(t))_{t \geq 0}$, $\underline{I}^\iota(t) = (I_1^\iota(t), \dots, I_d^\iota(t))$, with state space $\{0, 1\}^d$ defined as follows: The process starts in $I_j^\iota(0) = \delta_{\iota j}$, $j = 1, \dots, d$. As soon as one component (I_k^ι , say) reaches 1, then after the additional time $1 - \gamma$ all those components I_j^ι for which $a(k, j) > 0$ are set to 1. The fixation time of this process will be denoted by

$$S_{\mathcal{I}^{\iota, \gamma}} := \inf\{t \geq 0 : \underline{I}^\iota(t) = \underline{1}\}.$$

In other words, $S_{\mathcal{I}^{\iota, \gamma}} = (1 - \gamma)\Delta_\iota$, where Δ_ι is the number of steps that are needed to reach all other vertices of the graph G in a stepwise percolation starting from ι . An intuitive interpretation is as follows: State 1 of a component means that the colony is infected (by the beneficial type \mathcal{B}) and state 0 means that it is not infected. If a colony gets infected (at time t , say), then all the neighbouring (not yet infected) colonies get infected precisely at time $t + 1 - \gamma$.

2. For any $\iota \in \{1, \dots, d\}$, consider the (random) process $\mathcal{J}^\iota = (\underline{J}^\iota(t))_{t \geq 0}$, $\underline{J}^\iota(t) = (J_1^\iota(t), \dots, J_d^\iota(t))$, with state space $\{0, 1, 2\}^d$. In state 0, the colony is not infected, in state 1 it is infected but still not infectious, and in 2, it is infectious. The initial state is $J_\iota^\iota(0) = 1$ and $J_j^\iota(0) = 0$ for $j \neq \iota$, where ι is the founder colony. Transitions from state 1 to state 2 occur exactly one unit of time after entering state 1. For $j \neq \iota$, transitions from 0 to 1 occur at rate $2 \sum_k \rho_k a(k, j) \mathbf{1}_{\{J_k^\iota = 2\}}$. The fixation time of this process will be denoted by

$$S_{\mathcal{J}^\iota} := \inf\{t \geq 0 : \underline{J}^\iota(t) = \underline{2}\};$$

in particular, this time is larger than 1.

Infection in these epidemic processes indicates presence of the beneficial type. Our second main result quantifies in terms of these processes how various migration rates affect the spread and the fixation time of the beneficial type.

Theorem 2 (Fixation times of \mathcal{X}^*). For $\iota \in \{1, \dots, d\}$, let $\mathcal{X}^* = (\underline{X}^*(t))_{t \geq 0}$ be the solution of (2.5) with $\underline{X}^*(0) = \underline{0}$ and with the founder in colony ι , see Remark 2.3. Then, depending on the scaling ratio between μ and α as $\alpha \rightarrow \infty$, we have the following asymptotics for the fixation time T_{fix} defined in (2.3) (now for $\underline{\mathcal{X}}^*$ in place of $\underline{\mathcal{X}}$):

1. If $\mu \in \Theta(\alpha)$, then

$$\frac{\alpha}{\log \alpha} T_{\text{fix}} \xrightarrow{\alpha \rightarrow \infty} 2.$$

2. More generally, if $\mu \in \Theta(\alpha^\gamma)$ for some $\gamma \in [0, 1]$, then

$$\frac{\alpha}{\log \alpha} T_{\text{fix}} \xrightarrow{\alpha \rightarrow \infty} S_{\mathcal{I}^{\iota, \gamma}} + 2.$$

3. If $\mu = \frac{1}{\log \alpha}$, then

$$\frac{\alpha}{\log \alpha} T_{\text{fix}} \xrightarrow{\alpha \rightarrow \infty} S_{\mathcal{J}^\iota} + 1.$$

Remark 2.6. [Interpretation] Let us briefly give some heuristics for the three cases of the Theorem. The bottomline of our argument is this: Given a colony i is already “infected” by the beneficial mutant, the most probable scenario (as $\alpha \rightarrow \infty$) is that the beneficial type in colony i grows until migration exports the beneficial type to other colonies which can be reached from colony i . We argue with *successful lines*, which are – in a population undergoing Moran dynamics as in Remark 2.1 – individuals whose offspring are still present at the time of fixation.

For notational simplicity, we discuss here the situation $d = 2$ with the founder of the sweep being in colony $\iota = 1$. The three cases allow us to distinguish when the first successful migrant (carrying allele \mathcal{B} and still having offspring at the time of fixation) moves to colony 2.

1. $\mu \in \Theta(\alpha)$: Since in colony 1 the number of successful lines grows like a Yule process with branching rate α , migration of the first successful line will occur already while the Yule process has $\mathcal{O}(1)$ lines, i.e. at a time of order $1/\alpha$ if $\mu \in \Theta(\alpha)$. From here on, the beneficial allele has to fix in both colonies, which happens in time $2 \log(\alpha)/\alpha$ on each of the colonies.

We conjecture that this assertion is valid also for the case $\mu/\alpha \rightarrow \infty$, since intuitively a still higher migration rate should result in a panmictic situation due to an averaging effect. However, so far our techniques, and in particular our fundamental Lemma 4.1, do not cover this case.

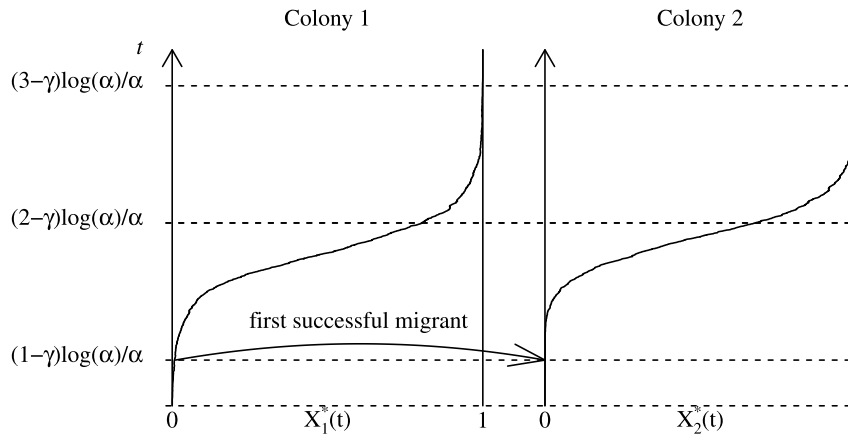
2. $\mu \in \Theta(\alpha^\gamma), 0 \leq \gamma < 1$: Again, the question is when the first successful migrant goes to colony 2. (In the epidemic model from Definition 2.5.1, this refers to infection of colony 2.) We will argue that this is the case after a time $(1 - \gamma) \log(\alpha)/\alpha$. Indeed, by this time, the Yule process approximating the number of successful lines in colony 1 has about $\exp(\alpha(1 - \gamma) \log(\alpha)/\alpha) = \alpha^{1-\gamma}$ lines, each of which travels to colony 2 at rate α^γ , so by that time the overall rate of migration to colony 2 is α . More generally, at time $x \log(\alpha)/\alpha$, the rate of successful migrants is $\alpha^{\gamma+x}$. So, if $\gamma + x < 1$, the probability that a successful migration happens up to time $x \log(\alpha)/\alpha$ is negligible, whereas if $\gamma + x > 1$, the probability that a successful migration happens up to time $x \log(\alpha)/\alpha$ is close to 1. By these arguments, the first successful migration must occur around time $(1 - \gamma) \log(\alpha)/\alpha$ and the time it then takes to fix in colony 2 is again $2 \log(\alpha)/\alpha$.
3. $\mu = 1/(\log \alpha)$: Here, migration is so rare that we have to wait until almost fixation in colony 1 before a successful migrant comes along. Consider the new timescale whose time unit is $\log \alpha/\alpha$, so that migration happens at rate $a(1, 2)/\alpha$ per individual on this timescale. Roughly, after time 1 (in the new timescale), the beneficial allele is almost fixed in colony 1.

For $N \gg \alpha$, a migrant is successful approximately with probability $2\alpha/N$, given by the survival probability of a supercritical branching process. So, if one of $N\rho_1$ lines on colony 1 migrates, each at rate $a(1, 2)/\alpha$, and with the success probability being $2\alpha/N$, the rate of successful migrants is $N\rho_1 \frac{a(1,2)}{\alpha} \frac{2\alpha}{N} = 2\rho_1 a(1, 2)$. At this rate, the second colony obtains a successful copy of the beneficial allele. Thus, in terms of the epidemic model from 2. in Definition 2.5, the first colony is infectious if allele \mathcal{B} is almost fixed there. From the time of the first successful migrant on, it takes again time 1 (in the new timescale) until the beneficial allele almost fixes in colony 2. This is when the state of colony 2 in the epidemic model changes from 1 (infected) to 2 (infectious).

The proof of Theorem 2 is given in Section 4.

The fixation time of a strongly beneficial allele

(A) $\mu \in \Theta(\alpha^\gamma)$



(B) $\mu = 1/(\log(\alpha))$

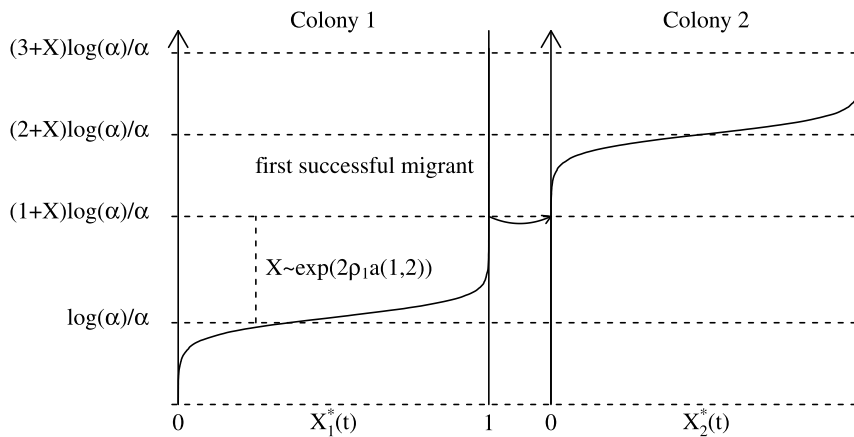


Figure 1: Two examples of a sweep in a structured population of $d = 2$ colonies. (A) For $\mu \in \Theta(\alpha^\gamma)$, the epidemic model $\mathcal{I}^{1,\gamma}$ from Theorem 2 starts with $\underline{I}^1(0) = (1, 0)$. The first successful migrant transports the beneficial allele to colony 2 at time $1 - \gamma$ (on the time-scale $\log(\alpha)/\alpha$). Hence, fixation occurs approximately at time $(3 - \gamma) \log(\alpha)/\alpha$. (B) For $\mu = 1/(\log \alpha)$, the epidemic model \mathcal{J}^1 from Theorem 2 starts with $\underline{J}^1(0) = (1, 0)$. The first successful migrant transports the beneficial allele to colony 2 at time $1 + X$, where X is an $\exp(2\rho_1 a(1, 2))$ distributed waiting time. Then, $\underline{J}^1(1 + X) = (2, 1)$ and thus $S_{\mathcal{J}^1} = 2 + X$. From here on, fixation in colony 2 takes one more unit of time. In total, fixation occurs approximately at time $(1 + S_{\mathcal{J}^1}) \log(\alpha)/\alpha = (3 + X) \log(\alpha)/\alpha$. For both figures we simulated a Wright-Fisher model, distributed on two colonies of equal size, i.e. $a(1, 2) = a(2, 1) = b(1, 2) = b(2, 1) = 1$ and $\rho_1 = \rho_2 = 1/2$. In (A), we used the following parameters: Each colony has size $N = 10^4$, $m = 0.001$ is the chance that an individual chooses its ancestor from the other colony, and $s = 0.01$ is the (relative) fitness advantage of beneficials, per generation. This amounts to $\gamma = \log(N \cdot s)/\log(N \cdot m) = 2/3$. In (B), we used $N = 10^5$, $s = 0.1$ and $N \cdot m = 1/(\log N \cdot s)$.

Remark 2.7. In [13] (see also [26]), it is derived in a heuristic manner that for $s \ll 1$ and $sN = \alpha > \mu = mN \gg 1$ the time to the first successful migrant is $\sim \frac{1}{\alpha} \log(1 + \frac{\alpha}{\mu})$. At least for $\mu \in \Theta(\alpha^\gamma)$, $0 \leq \gamma \leq 1$, this is confirmed by our Theorem 2.

Remark 2.8 (Different strengths of migration). The key argument mentioned at the beginning of Remark 2.6 continues to hold if the migration intensity between colonies is not of the same order of magnitude. More precisely, assume that the asymptotics of the gene flows as $\alpha \rightarrow \infty$ is of the form $\mu\rho_i a(i, j) = \mu\rho_j b(j, i) \in \Theta(\alpha^{\gamma_{ij}})$, where the exponents $(\gamma_{ij})_{i,j=1,\dots,d} \in [0, 1]^{d \times d}$ may vary with i, j (possibly also due to a strongly varying colony size).

Then colony j can become infected from neighbouring colonies only if one of the neighbouring colonies (i) is infected and (ii) carries enough beneficial mutants in order to infect colony j . So again the fixation time of the beneficial allele can be computed from taking the minimal time it takes to infect all colonies across the graph G , plus the final phase of fixation of the beneficial allele. Consequently, the epidemic process $\mathcal{I}^t := \mathcal{I}^{t,\gamma}$ from Definition 2.5 can be changed to $\mathcal{I}^{t,\underline{\gamma}}$ as follows: As soon as for some i the process I_i^t reaches the value 1, then after an additional fixed time of length $1 - \gamma_{ij}$ all of the I_j^t for which $a(i, j) > 0$ are set to 1.

In the sequel we focus on the case $\gamma_{ij} \equiv \gamma$ of a spatially homogeneous asymptotics in order to keep the presentation transparent. We emphasise however, that our proofs are designed in a way which makes the described generalization feasible.

3 The ancestral selection graph

A principal tool for the analysis of interacting Wright–Fisher diffusions with selection is their duality with the ancestral selection graph (ASG) of Krone and Neuhauser, which we recall in detail below. The main idea for the proof of Theorems 1 and 2 is

- to obtain via the ASG a duality relationship and a Kingman paintbox representation also for the diffusion process \mathcal{X}^* (i.e. the process conditioned to get absorbed at $\underline{1}$), and to represent T_{fix} via duality,
- to show how the equilibrium ASG and its time-reversal can be employed for asymptotic calculations as $\alpha \rightarrow \infty$.

This structure allows us to use the techniques of (multidimensional) birth-death processes in order to perform the asymptotic analysis using bounds based on sub- and supercritical branching processes.

In the present section we will focus on the two bullet points, while the asymptotic analysis of the birth-death processes is in Section 4, with the basic heuristics in Section 4.1. To carry out this program we proceed as follows:

In Section 3.1 we will give an informal description of the ASG and present some of the central ideas of the subsequent proofs. We will also state a key proposition (Proposition 3.1) which gives a connection between the fixation time and a two-dimensional birth-and-death process that describes the percolation of the beneficial type within the equilibrium ASG. We give a formal definition of the structured ASG via a particle representation in Section 3.2 and derive a time-reversal property in Section 3.3, which will be important in the proof of Proposition 3.1. In the subsequent sections we will derive paintbox representations for the solutions of (2.1) and (2.5) using the duality relationships from above, and complete the proofs of Proposition 3.1 and Theorem 1.

3.1 Outline of proof strategy and a key proposition

The basic tool for proving Theorems 1 and 2 will be a representation of $\underline{X}^*(\tau)$ (the solution of (2.5) at a fixed time τ) in terms of an exchangeable particle system. This representation is first achieved for initial conditions $\underline{x} \in [0, 1]^d \setminus \{\underline{0}\}$, and then also for the entrance laws from $\underline{0}$. At the heart of the construction is a *conditional duality* which

extends the classical duality between the (unconditioned) \mathcal{X} (the solution of (2.1)) and the *structured ancestral selection graph*.

The latter is constructed in terms of a branching-coalescing-migrating system $\mathcal{A} = (\mathcal{A}_r)_{r \geq 0}$ of particles, where each pair of particles in colony i

- coalesces at rate $1/\rho_i$, $i = 1, \dots, d$,

and each particle in colony i

- branches (i.e. splits into two) at rate α ,

- migrates (i.e. jumps) to colony j at rate $\mu b(i, j)$.

When the starting configuration of \mathcal{A} consists of k_i particles in colony i , $i = 1, \dots, d$, we will speak of a \underline{k} -ASG, where for brevity we write $\underline{k} := (k_i)_{i=1, \dots, d}$. A more refined definition of \mathcal{A} , which will also allow to speak of a *connectedness relation* between particles at different times, will be given in Sections 3.2 and 3.4. With this refined definition, each particle in \mathcal{A}_r is represented as a point in $\{1, \dots, d\} \times [0, 1]$, the first component referring to the colony in which the particle is located, and the second component being a *label* which is assigned independently and uniformly at each branching, coalescence and migration event. The ASG then records the trajectories of all the particles in \mathcal{A} , see Figure 2(a) for an illustration.

Writing $K_r^{\underline{k}}(i)$ for the number of particles in the \underline{k} -ASG in colony i at time r and using the notation

$$(\underline{1} - \underline{y})^{\underline{\ell}} := \prod_{i=1}^d (1 - y_i)^{\ell_i}, \quad \underline{y} = (y_1, \dots, y_d) \in [0, 1]^d, \quad \underline{\ell} = (\ell_1, \dots, \ell_d) \in \mathbb{N}_0^d, \quad (3.1)$$

we have a moment duality between $\underline{K} = (K(i))_{i=1, \dots, d}$ and the solution \mathcal{X} of (2.1):

$$\mathbf{E}_{\underline{x}}[(\underline{1} - \underline{X}(\tau))^{\underline{k}}] = \mathbb{E}[(\underline{1} - \underline{x})^{\underline{K}_{\tau}^{\underline{k}}}], \quad \underline{x} \in [0, 1]^d, \quad \underline{k} \in \mathbb{N}_0^d, \quad \tau \geq 0. \quad (3.2)$$

Here and in the following, we denote the probability measure that underlies the particle process \mathcal{A} (and processes related to it) by \mathbb{P} (and thus distinguish it from the probability measure $\mathbf{P}_{\underline{x}}$ that underlies the diffusion process \mathcal{X} appearing in (2.1) as well as the corresponding processes, like \mathcal{X}^*). Analogously, we use these notation types for the corresponding expectations and variances. The proof of the *basic duality relationship* (3.2) will be recalled in Lemma 3.7.

Eq.(3.2) has a conceptual interpretation in population genetics terms: We know that $\underline{X}(\tau)$ is the vector whose i -th coordinate is the frequency of the beneficial type \mathcal{B} in colony i at time τ when $\underline{X}(0) = \underline{x}$. Thus, the left hand side of (3.2) is the probability that nobody in a \underline{k} -sample drawn from the population (with k_i individuals drawn from colony i , $i = 1, \dots, d$) is of type \mathcal{B} , given that τ time units ago the type frequencies were \underline{x} . In the light of a Moran model with selection (whose diffusion limit yields the process \mathcal{X}), the particles' trajectories in the ASG can be interpreted as *potential ancestral lineages* of the \underline{k} -sample. The type of a particle in the sample can be recovered by a simple rule: it is the beneficial type \mathcal{B} if and only if at least one of its potential ancestors carries type \mathcal{B} . In other words, the beneficial type *percolates upwards* along the lineages of the ASG; see Fig. 2(b) for an illustration.

Consequently, the event that nobody in the \underline{k} -sample is of type \mathcal{B} equals the event that nobody of the sample's potential ancestors is of type \mathcal{B} . The probability of this event, however, is just the right hand side of (3.2). Thus, Eq. (3.2) expresses the probability of one and the same event in two different ways.

We will argue in Sec. 3.6 that the process \mathcal{A} can be started with infinitely many particles in each colony, with the number of particles immediately coming down from infinity.

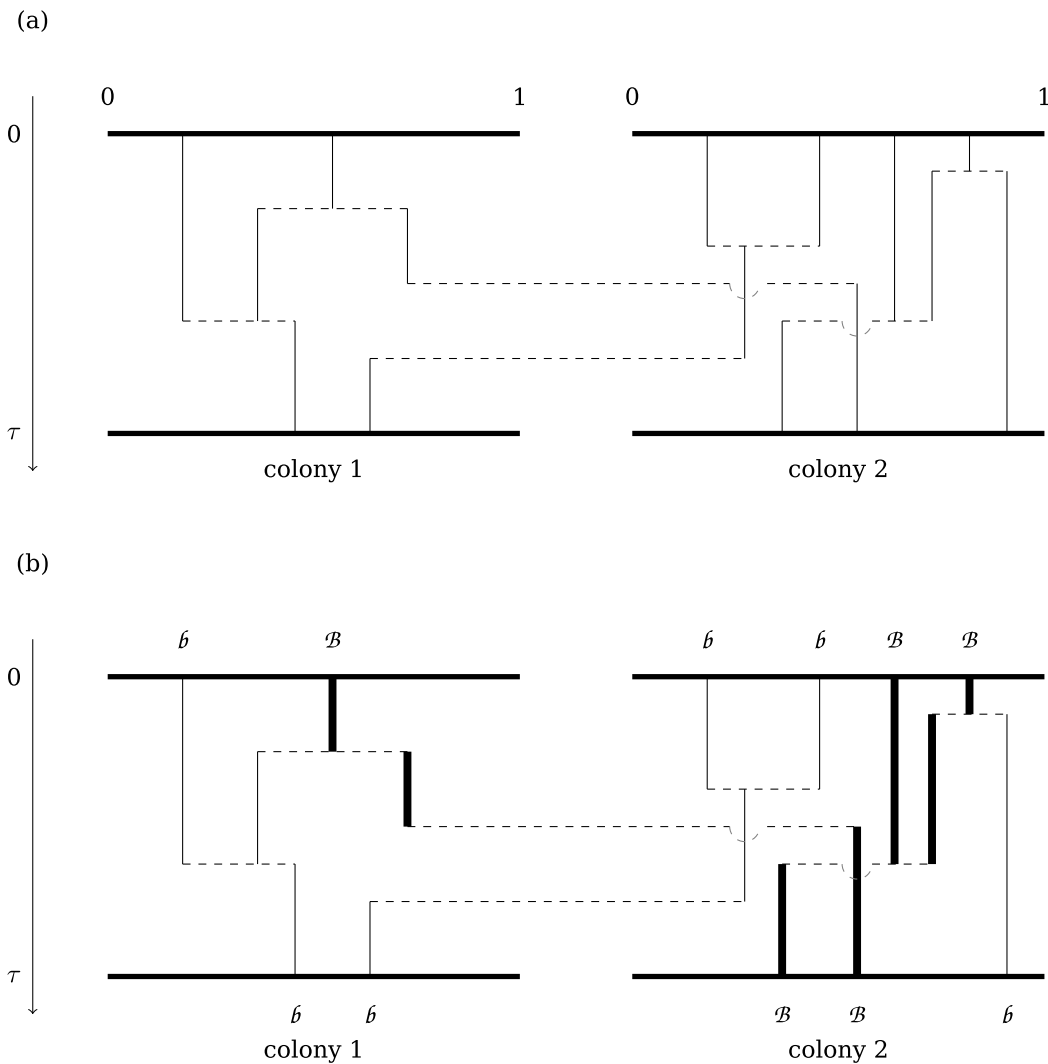


Figure 2: (a) A realisation of the k -ASG in the time interval $[0, \tau]$ with 2 colonies, and $k = (2, 4)$. Initially and at each coalescence, branching and migration event, independent and uniform $[0, 1]$ -distributed labels are assigned to the particles, and the *genealogical connections* of particles are recorded (visualised by the horizontal dashed lines).

(b) The same realisation of the ASG as in Figure 2(a), now showing the particle's types. Two of the five particles in \mathcal{A}_τ are marked with B . Percolation of type B happens "upwards" along the ASG: all those particles in the $(2, 4)$ -sample \mathcal{A}_0 are assigned type B which are connected to a type B -particle in \mathcal{A}_τ .

This process will be denoted by \mathcal{A}^∞ . If one marks the particles in \mathcal{A}_τ^∞ independently with probabilities given by \underline{x} and lets the types percolate upwards along the ASG, then one obtains for each $i \in \{1, \dots, d\}$ an exchangeable marking of the particles in \mathcal{A}_0^∞ that are located in colony i . Let us denote by $F_i^{\underline{x}, \tau}$ the relative frequency of the marked particles within all particles of \mathcal{A}_0^∞ that are located in colony i ; due to de Finetti's theorem, for each i , the quantity $F_i^{\underline{x}, \tau}$ exists a.s. Based on the duality relationship (3.2) we will show in Lemma 3.9 that

$$\mathbf{P}_{\underline{x}}(\underline{X}(\tau) \in (\cdot)) = \mathbf{P}(F^{\underline{x}, \tau} \in (\cdot)), \quad \underline{x} \in [0, 1]^d \setminus \{0\}, \tau \geq 0.$$

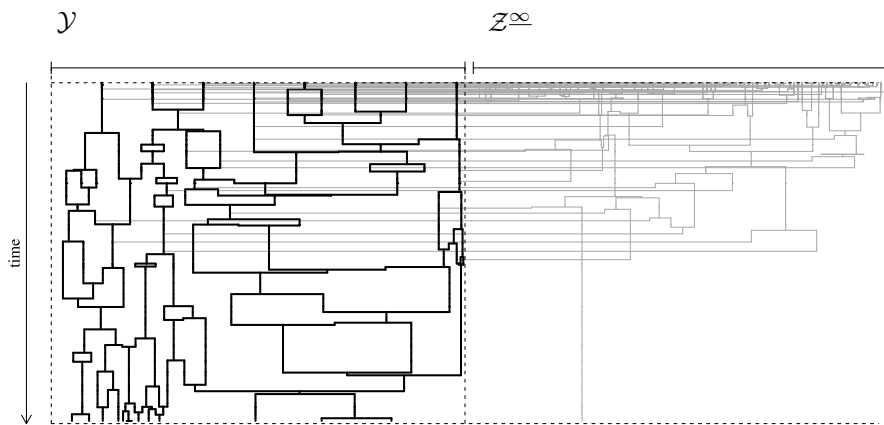


Figure 3: The paintbox representations constructed in Section 3.8 uses two particle systems that are coupled to each other. Initially, these two systems are disjoint, and the coupling consists in a (local) coalescence between the two ASG’s as illustrated in the figure. The potential ancestors of the sample on top of the figure are found at the bottom of the figure. The bold lines (in the left part of the figure) belong to \mathcal{Y} , the gray lines belong to $\mathcal{Z} \setminus \mathcal{Y}$.

Following Aldous’ terminology (see e.g. p. 88 in [1]) we will call this a “Kingman paintbox” representation of $\underline{X}(\tau)$.

In order to find a similar representation for $\underline{X}^*(\tau)$, we will use a coupling of two processes, denoted $\mathcal{Z} := \mathcal{Z}^\infty$ and \mathcal{Y} , which both follow the same dynamics as \mathcal{A} . Here, \mathcal{Z}^∞ starts with $\mathcal{Z}_0^\infty = \infty$ and \mathcal{Y}_0 is an equilibrium configuration of the coalescence-branching-migration dynamics described above. (As we will prove in Proposition 3.3, the particle numbers in equilibrium constitute a Poisson configuration with intensity measure $(2\alpha\rho_1, \dots, 2\alpha\rho_d)$, conditioned to be non-zero.) Since \mathcal{Z} and \mathcal{Y} follow the same exchangeable dynamics, we can embed both in a single particle system \mathcal{A} which starts in the a.s. disjoint union $\mathcal{A}_0 := \mathcal{Y}_0 \cup \mathcal{Z}_0$ and follows the coalescence-branching-migration dynamics. Then, \mathcal{Y} arises by following particles within \mathcal{Y}_0 along \mathcal{A} and \mathcal{Z} arises by following particles within \mathcal{Z}_0 along \mathcal{A} .

Let $\mathcal{A}_\tau^{(\underline{x})}$ denote the subsystem of marked particles of $\mathcal{A}_\tau = \mathcal{Y}_\tau \cup \mathcal{Z}_\tau$ which arises by an independent marking with probabilities \underline{x} . We will prove in Lemma 3.11 that

$$\mathbf{E}_{\underline{x}}[(1 - \underline{X}^*(\tau))^{\underline{k}}] = \mathbb{P}(\mathcal{Z}_\tau^{\underline{k}} \cap \mathcal{A}_\tau^{(\underline{x})} = \emptyset | \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset), \quad \underline{x} \in [0, 1]^d \setminus \{0\}, \underline{k} \in \mathbb{N}_0^d, \tau \geq 0,$$

with $\mathcal{Z}^{\underline{k}}$ started in \underline{k} particles. This *conditional duality relationship* will be crucial for deriving the paintbox representation for $\underline{X}^*(\tau)$. With the notation $\underline{F}^{\underline{x}, \tau}$ introduced above for the vector of frequencies of the marked particles we will prove in Lemma 3.12 that

$$\mathbf{P}_{\underline{x}}(\underline{X}^*(\tau) \in (\cdot)) = \mathbb{P}(\underline{F}^{\underline{x}, \tau} \in (\cdot) | \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset), \quad \underline{x} \in [0, 1]^d \setminus \{0\}, \tau \geq 0.$$

Let us emphasize that the conditioning under the event $\{\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset\}$ affects the distribution of \mathcal{Y} , i.e. takes it out of equilibrium and changes its dynamics between times 0 and τ . We will denote the vector of particle numbers in \mathcal{Y}_r by \underline{N}_r , $r \geq 0$.

Now consider, for some $\iota \in \{1, \dots, d\}$ and $0 < \varepsilon < 1$, the vector $\underline{x} = \varepsilon e_\iota$, meaning that initially a fraction ε of the particles in colony ι is of beneficial type while all the other colonies carry only the inferior type b . In the limit $\varepsilon \rightarrow 0$ the conditioning under the event $\{\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\varepsilon e_\iota)} \neq \emptyset\}$ amounts to changing the distribution of \underline{N}_τ from its

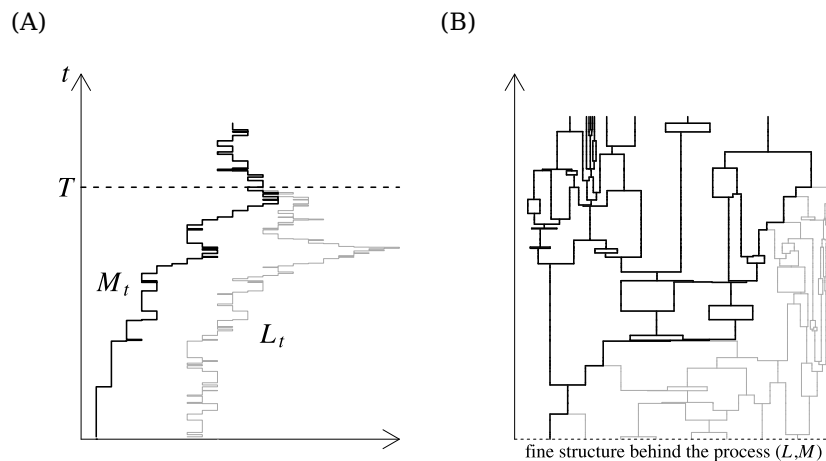


Figure 4: (A) A realisation of the processes $(\underline{M}_t)_{t \geq 0}$ and $(\underline{L}_t)_{t \geq 0}$ for the case of one colony. The joint distribution of these two processes is given in Proposition 3.1. T is the first time t when $\underline{M}_t = \underline{L}_t$. (B) The pair $(\underline{L}, \underline{M})$ has an underlying structure in terms of the particle system $\hat{\mathcal{Y}}$, where \underline{L} arises as the counting process of *all* particles in $\hat{\mathcal{Y}}$, and $(\underline{M}_t)_{t \geq 0}$ is the counting process of the *marked* particles in $\hat{\mathcal{Y}}$.

equilibrium distribution to the distribution of $\underline{\Pi} + e_\iota$, where $\underline{\Pi}$ is $\text{Poi}(2\alpha\rho)$ -distributed, see Remark 3.13. This will result in a paintbox representation for the distribution of $\mathcal{X}^*(\tau)$ under the measure \mathbf{P}'_0 which appears in Theorem 1, see Corollary 3.15 a). The event that, in the system (2.5), fixation of the beneficial type has occurred by time τ can then be reexpressed as the event that the (one) marked particle in \mathcal{Y}_τ is among the potential ancestors of *all* the infinitely many particles in \mathcal{Z}_0^∞ , see Corollary 3.15 c).

We will show in Lemma 3.18 and in Corollary 3.19 that frequencies within \mathcal{Y} and \mathcal{Z} are very close, such that for the distribution of the fixation time on the $\log(\alpha)/\alpha$ -timescale it will suffice to study the probability that the marking of a single particle in colony ι at time τ percolates “upwards” through \mathcal{Y} in the time interval $[0, \tau]$. This analysis is most conveniently carried through in the *time reversal* $\hat{\mathcal{Y}}$ of \mathcal{Y} , whose migration rates are reversed as given by Equation 2.2. The event $\{\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(e_\iota)} \neq \emptyset\}$ is the same as $\{\hat{\mathcal{Y}}_0 \cap \mathcal{A}_0^{(x)} \neq \emptyset\}$; thus the conditioning changes the initial condition of $\hat{\mathcal{Y}}$ but not its dynamics (whereas, as mentioned above, the dynamics of \mathcal{Y} , is changed by the conditioning).

We will write $(\underline{M}_t)_{t \geq 0}$ for the counting process of the marked particles in $(\hat{\mathcal{Y}}_t)_{t \geq 0}$, and $(\underline{L}_t)_{t \geq 0}$ for the counting process of *all* particles in $(\hat{\mathcal{Y}}_t)_{t \geq 0}$. The dynamics of the bivariate process $(\underline{L}_t)_{t \geq 0}, (\underline{M}_t)_{t \geq 0}$ is described next, together with the key result how to use the ASG for approximating the fixation time under strong selection. Its proof is given in Section 3.9 and an illustration is given in Figure 4.

Proposition 3.1 (An approximation of T_{fix}). *Let $(\underline{L}_t, \underline{M}_t)$, $\underline{L}_t = (L_t^1, \dots, L_t^d)$, $\underline{M}_t = (M_t^1, \dots, M_t^d)$, be defined as follows: For fixed $\iota \in \{1, \dots, d\}$, let Π_1, \dots, Π_d be independent and $\text{Poi}(2\alpha\rho_i)$ -distributed, and put $\underline{L}_0 = \underline{\Pi} + e_\iota$, $\underline{M}_0 = e_\iota$. The process $(\underline{L}, \underline{M})$ jumps from $(\underline{\ell}, \underline{m})$ to*

$$\begin{aligned} &(\underline{\ell} + e_i, \underline{m} + e_i) \text{ at rate } \alpha m_i, \\ &(\underline{\ell} + e_i, \underline{m}) \text{ at rate } \alpha(\ell_i - m_i), \end{aligned}$$

The fixation time of a strongly beneficial allele

$$\begin{aligned}
 (\underline{\ell} - \underline{e}_i, \underline{m} - \underline{e}_i) & \text{ at rate } \frac{1}{\rho_i} \binom{m_i}{2}, \\
 (\underline{\ell} - \underline{e}_i, \underline{m}) & \text{ at rate } \frac{1}{\rho_i} (\ell_i - m_i) m_i + \frac{1}{\rho_i} \binom{\ell_i - m_i}{2}, \\
 (\underline{\ell} - \underline{e}_i + \underline{e}_j, \underline{m} - \underline{e}_i + \underline{e}_j) & \text{ at rate } \mu a(i, j) m_i, \\
 (\underline{\ell} - \underline{e}_i + \underline{e}_j, \underline{m}) & \text{ at rate } \mu a(i, j) (\ell_i - m_i).
 \end{aligned}$$

Moreover, let

$$T := \inf\{t \geq 0 : \underline{M}_t = \underline{L}_t\}, \quad (3.3)$$

and let T_{fix} be the fixation time of \mathcal{X}^* , where \mathcal{X}^* is a solution of the SDE (2.5) as described in Theorem 1. Assume that the limiting distribution of $\frac{\alpha}{\log \alpha} T$ exists as $\alpha \rightarrow \infty$. Then

$$\lim_{\alpha \rightarrow \infty} \mathbf{P}'_0 \left(\frac{\alpha}{\log \alpha} T_{\text{fix}} \leq t \right) = \lim_{\alpha \rightarrow \infty} \mathbf{P} \left(\frac{\alpha}{\log \alpha} T \leq t \right), \quad (3.4)$$

in each continuity point of the limiting distribution function. Here, $\mu = \mu(\alpha)$ can depend on α in an arbitrary way.

Remark 3.2 (Existence of limiting distribution). Our proof of Theorem 2 in Sec. 4 will reveal in particular that the limiting distribution of $\frac{\alpha}{\log \alpha} T$ exists as $\alpha \rightarrow \infty$, at least if $\mu = \mu(\alpha)$ falls in one of the three cases of Theorem 2.

3.2 The structured ancestral selection graph as a particle system

We will define a Markov process $\mathcal{A} = (\mathcal{A}_r)_{r \geq 0}$ that takes its values with probability 1 in the set of finite subsets of $\{1, \dots, d\} \times [0, 1]$. We shall refer to the elements of \mathcal{A}_r as *particles*. For each particle $(i, u) \in \mathcal{A}_r$, we call i the particle's *location* and u the particle's *label*. Recall that we denote the probability measure that underlies \mathcal{A} by \mathbf{P} . It will sometimes be convenient to annotate the configuration of locations of the initial state as a superscript of \mathcal{A} or \mathcal{Z} . Specifically, for $\underline{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, we put

$$\mathcal{A}_0^{\underline{k}} = \bigcup_{i=1}^d \{(i, U_{ig}) : 1 \leq g \leq k_i\}, \quad (3.5)$$

where the U_{ig} are independent and uniformly distributed on $[0, 1]$.

We now specify the Markovian dynamics of \mathcal{A} in terms of its jump kernel \mathcal{D}^b for some migration kernel \underline{b} on $\{1, \dots, d\}$. Here we distinguish three kinds of events (see Figure 5 for an illustration):

- (1) **Coalescence:** for all $i = 1, \dots, d$, every pair of particles in colony i is replaced at rate $1/\rho_i$ by one particle in colony i with a label that is uniformly distributed on $[0, 1]$ and independent of everything else.
- (2) **Branching:** for all $i = 1, \dots, d$, every particle in colony i is replaced at rate α by two particles in colony i with labels that are uniformly distributed on $[0, 1]$ and independent of each other and of everything else.
- (3) **Migration:** for all $i = 1, \dots, d$, every particle in colony i is replaced at rate $\mu b(i, j)$, $j \in \{1, \dots, d\}, j \neq i$, by a particle in colony j with a label that is uniformly distributed on $[0, 1]$ and independent of everything else.

We will refer to $\mathcal{A} = (\mathcal{A}_r)_{r \geq 0}$ also as the *structured ancestral selection graph* (or ASG for short). The vector of *particle numbers* at time r is $\underline{K}_r = (K_r(1), \dots, K_r(d))$ with

$$K_r(i) := \#(\mathcal{A}_r \cap (\{i\} \times [0, 1])), \quad r \geq 0, \quad i = 1, \dots, d. \quad (3.6)$$

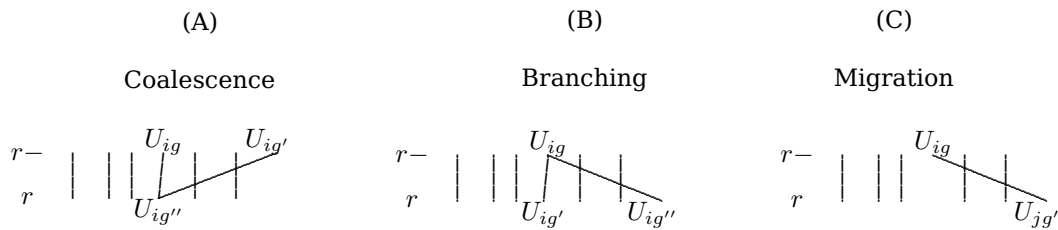


Figure 5: If a coalescing event (1), a branching event (2) or a migration event (3) occurs by time r , we connect the lines within the ASG according to the rules as given in Section 3.2. In all cases, labels U_{ig} are uniformly distributed on $[0, 1]$, and are updated upon any event for the affected lines.

$\underline{K} := (\underline{K}_r)_{r \geq 0}$ is a Markov process whose jump rates (based on the migration kernel \underline{b}) are for $\underline{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d \setminus \{\underline{0}\}$ given by

$$\begin{aligned} q_{\underline{k}, \underline{k} - \underline{e}_i}^b &:= q_{\underline{k}, \underline{k} + \underline{e}_i} := \frac{1}{\rho_i} \binom{k_i}{2}, \\ q_{\underline{k}, \underline{k} + \underline{e}_i}^b &:= q_{\underline{k}, \underline{k} - \underline{e}_i} := \alpha k_i, \\ q_{\underline{k}, \underline{k} - \underline{e}_i + \underline{e}_j}^b &:= \mu b(i, j) k_i, \\ q_{\underline{k}, \underline{\ell}}^b &:= q_{\underline{k}, \underline{\ell}} := 0 \quad \text{otherwise.} \end{aligned} \tag{3.7}$$

By analogy with the notation \mathcal{A}^k , we write $(\underline{K}_r^k)_{r \geq 0}$ for the process with initial state \underline{k} .

3.3 Equilibrium and time reversal of the ASG

Proposition 3.3 (Equilibrium for \mathcal{D}^b).

1. The unique equilibrium distribution π for the dynamics \mathcal{D}^b is the law π of a Poisson point process on $\{1, \dots, d\} \times [0, 1]$ with intensity measure $2\alpha \underline{\rho} \otimes \lambda$, conditioned to be non-zero (where $\underline{\rho} = (\rho_1, \dots, \rho_d)$ and λ stands for the Lebesgue measure on $[0, 1]$.)
2. The jump kernel $\hat{\mathcal{D}}$ of the time reversal of \mathcal{A} in its equilibrium π is again of the form (1),(2),(3), with the only difference that the migration rates $b(i, j)$ are replaced by the migration rates $a(i, j)$ as defined in (2.2), i.e. $\hat{\mathcal{D}} = \mathcal{D}^a$.

Proof. We will prove the duality relation

$$\pi(d\underline{z}) \mathcal{D}^b(\underline{z}, d\underline{z}') = \pi(d\underline{z}') \mathcal{D}^a(\underline{z}', d\underline{z}), \tag{3.8}$$

which by well known results about time reversal of Markov chains in equilibrium (see e.g. [20]) proves both assertions of the Proposition at once. Since, given the particles' locations, their labels are independent and uniformly distributed on $[0, 1]$ and since this is propagated in each of the (coalescence, branching and migration) events, it will be sufficient to consider the process \underline{K} . Indeed, defining $q_{\underline{k}, \underline{\ell}}^a$ as in (3.7) and putting

$$\pi_{(k_1, \dots, k_d)} = \frac{e^{-2\alpha}}{1 - e^{-2\alpha}} \frac{(2\alpha)^{k_1 + \dots + k_d}}{k_1! \dots k_d!} \rho_1^{k_1} \dots \rho_d^{k_d}, \quad \underline{k} \in \mathbb{N}_0^d \setminus \{\underline{0}\},$$

one readily checks for all $\underline{k} \in \mathbb{N}_0^d \setminus \{\underline{0}\}$

$$\pi_{\underline{k}} \cdot q_{\underline{k}, \underline{k} - \underline{e}_i} = \pi_{\underline{k} - \underline{e}_i} \cdot q_{\underline{k} - \underline{e}_i, \underline{k}}, \quad \pi_{\underline{k}} \cdot q_{\underline{k}, \underline{k} - \underline{e}_i + \underline{e}_j}^b = \pi_{\underline{k} - \underline{e}_i + \underline{e}_j} \cdot q_{\underline{k} - \underline{e}_i + \underline{e}_j, \underline{k}}^a.$$

This can be summarized as

$$\pi_{\underline{k}} q_{\underline{k}, \underline{\ell}}^b = \pi_{\underline{\ell}} q_{\underline{\ell}, \underline{k}}^a, \quad \underline{k}, \underline{\ell} \in \mathbb{N}_0^d \setminus \{\underline{0}\},$$

which by definition of \mathcal{D}^b and \mathcal{D}^a lifts to (3.8), and thus proves the Proposition. \square

3.4 Genealogical relationships in the ASG

Thanks to the labelling of the particles it makes sense to speak about genealogical relationships within \mathcal{A} . Doing so will facilitate the interpretation of the duality relationships in the proofs of Proposition 3.1 and Theorem 1.

Definition 3.4 (Connections between particles in \mathcal{A}). *Let \mathcal{A} follow the dynamics \mathcal{D}^b described in Section 3.2. We say that a particle (i', u') replaces a particle (i, u) if either of the following relations hold:*

- *there is a migration event in which (i, u) is replaced by (i', u') ,*
- *there is a coalescence event for which (i, u) belongs to the pair which is replaced by (i', u') ,*
- *there is a branching event for which (i', u') belongs to the pair which replaces (i, u) .*

(Note that in the 2nd and 3rd case we have necessarily $i = i'$.) For $r, s \geq 0$ we say that two particles $(i, u) \in \mathcal{A}_{r \wedge s}$, $(i', u') \in \mathcal{A}_{r \vee s}$ are connected if either $(i, u) = (i', u')$ or there exists an $n \in \mathbb{N}$ and $(i_0, u_0), \dots, (i_n, u_n)$ such that $(i_0, u_0) = (i, u)$, $(i_n, u_n) = (i', u')$, and (i_ℓ, u_ℓ) replaces $(i_{\ell-1}, u_{\ell-1})$ for $\ell = 1, \dots, n$. For any subset S_r of \mathcal{A}_r , let

$$\mathcal{C}_s(S_r) := \bigcup_{(i,u) \in S_r} \{(i', u') \in \mathcal{A}_s : (i, u) \text{ and } (i', u') \text{ are connected}\}$$

be the collection of all those particles in \mathcal{A}_s that are connected with at least one particle in S_r . We briefly call $\mathcal{C}_s(S_r)$ the subset of \mathcal{A}_s that is connected with S_r .

3.5 Basic duality relationship

We recall a basic duality result for the ASG for a structured population in Lemma 3.7, as can e.g. be found in [3, equation (1.5)]. For this purpose we use a marking procedure of the process \mathcal{A} .

Definition 3.5 (A marking of particles). *Let \mathcal{A} follow the dynamics \mathcal{D}^b described in Section 3.2, and fix a time $\tau > 0$. Take $\underline{x} = (x_1, \dots, x_d) \in [0, 1]^d$, and mark independently all particles in colony i at time τ with probability x_i . Denote by*

$$\mathcal{A}_\tau^{(\underline{x})} := \{(i, u) \in \mathcal{A}_\tau : (i, u) \text{ is marked}\}$$

the collection of all marked particles in \mathcal{A}_τ .

Remark 3.6 (Connectedness and marks). In the sequel we will use the following observation: for any subset S_0 of \mathcal{A}_0 ,

$$S_0 \cap \mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) = \emptyset \text{ if and only if } \mathcal{C}_\tau(S_0) \cap \mathcal{A}_\tau^{(\underline{x})} = \emptyset.$$

For $S_0 = \mathcal{A}_0$, we find that $\mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) = \emptyset$ if and only if $\mathcal{A}_\tau^{(\underline{x})} = \emptyset$.

In words: no particle in S_0 is marked (i.e. of “beneficial type”), if and only if no potential ancestral particle of S_0 is marked.

Lemma 3.7 (Basic duality relationship). *Let $\mathcal{X} = (\underline{X}(t))_{t \geq 0}$ be the solution of (2.1) with $\underline{X}(0) = \underline{x} \in [0, 1]^d$, and let \mathcal{A} follow the dynamics \mathcal{D}^b . Then, for all $\underline{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, we have, using the notation (3.1) and (3.6)*

$$\mathbf{E}_{\underline{x}}[(1 - \underline{X}(\tau))^{\underline{k}}] = \mathbf{E}[(1 - \underline{x})^{K_\tau^{\underline{k}}}] = \mathbb{P}(\mathcal{A}_\tau^{(\underline{x})} = \emptyset | \#\mathcal{A}_0 = \underline{k}). \tag{3.9}$$

Proof. The generator of the Markov process \mathcal{X} is given by

$$G_{\mathcal{X}}f(\underline{x}) = \frac{1}{2} \sum_{j=1}^d \frac{1}{\rho_j} x_j(1 - x_j) \frac{\partial^2 f(\underline{x})}{\partial^2 x_j} + \alpha \sum_{i=1}^d x_i(1 - x_i) \frac{\partial f(\underline{x})}{\partial x_i} + \mu \sum_{i,j=1}^d b(i, j)(x_j - x_i) \frac{\partial f(\underline{x})}{\partial x_i}$$

for functions $f \in \mathcal{C}^2([0, 1]^d)$. Hence, for $f_{\underline{k}}(\underline{x}) := (1 - \underline{x})^{\underline{k}}$ and $g_{\underline{x}}(\underline{k}) := (1 - \underline{x})^{\underline{k}}$,

$$\begin{aligned} G_{\mathcal{X}} f_{\underline{k}}(\underline{x}) &= \sum_{i=1}^d \frac{1}{\rho_i} x_i \binom{k_i}{2} (1 - \underline{x})^{k - \varepsilon_i} + \alpha \sum_{i=1}^d k_i (-x_i) (1 - \underline{x})^k \\ &\quad + \mu \sum_{i,j=1}^d b(i, j) k_i ((1 - x_j) - (1 - x_i)) (1 - \underline{x})^{k_i - \varepsilon_i} \\ &= \sum_{i=1}^d \frac{1}{\rho_i} \binom{k_i}{2} ((1 - \underline{x})^{k - \varepsilon_i} - (1 - \underline{x})^k) \\ &\quad + \alpha \sum_{i=1}^d k_i ((1 - \underline{x})^{k + \varepsilon_i} - (1 - \underline{x})^k) \\ &\quad + \mu \sum_{i,j=1}^d b(i, j) k_i ((1 - \underline{x})^{k - \varepsilon_i + \varepsilon_j} - (1 - \underline{x})^k) \\ &= G_{\underline{K}} g_{\underline{x}}(\underline{k}), \end{aligned}$$

where $G_{\underline{K}}$ is the generator of \underline{K} . Now, the first equality in the duality relationship (3.9) is straightforward; see [7, Section 4.4]. The second equality in (3.9) is immediate from the definition of the marking procedure in Definition 3.5. \square

3.6 A paintbox representation of $\underline{X}(\tau)$

Our next aim is a *de Finetti–Kingman paintbox representation* of the distribution of $\underline{X}(\tau)$ under $\mathbf{P}_{\underline{x}}$ in terms of the dual process \underline{K}^{∞} . In order to achieve this, we need to be able to start the ASG with infinitely many lines and define frequencies of marked particles.

Remark 3.8 (Asymptotic frequencies).

1. The process \mathcal{A} can be started from

$$\mathcal{A}_0^{\infty} = \bigcup_{i=1}^d \{(i, U_{ig}) : 1 \leq g < \infty\}, \tag{3.10}$$

where $(U_{ig})_{i=1, \dots, d, g=1, 2, \dots}$ is an independent family of uniformly distributed random variables on $[0, 1]$. Indeed, the quadratic death rates of the process \underline{K} (recall this process from (3.6)) ensure that the number of particles comes down from infinity. In order to see this, consider the process $(K_r^1 + \dots + K_r^d)_{r \geq 0}$ and note that given $K_r^1 + \dots + K_r^d = k$ it increases at rate αk and its rate of decrease is minimal if colony i carries $\rho_i k$ lines, $i = 1, \dots, d$, hence is bounded from below by

$$\sum_{i=1}^d \frac{1}{\rho_i} \binom{k_i}{2} \geq \frac{1}{2} \left(\sum_{i=1}^d k_i^2 - k \right) \geq \frac{1}{2} \left(\frac{1}{d} k^2 - k \right) \geq \frac{k(k - d)}{2d}, \tag{3.11}$$

where we have used the Cauchy–Schwarz inequality in the second “ \geq ”. Using the same bounds as in Proposition 6.9 of [5], we see that $\#\mathcal{A}_{\varepsilon}^{\infty} = O(\varepsilon^{-1})$ as $\varepsilon \rightarrow 0$.

2. For $i = 1, \dots, d$, let $(J_{i1}, J_{i2}, \dots) := ((i, U_{i1}), (i, U_{i2}), \dots)$ be the (numbered) collection of particles in \mathcal{A}_0^{∞} that are located in colony i . Then by definition of the dynamics of \mathcal{A}^{∞} , the sequence

$$(\mathbb{1}_{\{J_{i1} \in \mathcal{C}_0(\mathcal{A}_{\tau}^{\underline{x}})\}}, \mathbb{1}_{\{J_{i2} \in \mathcal{C}_0(\mathcal{A}_{\tau}^{\underline{x}})\}}, \dots) \tag{3.12}$$

The fixation time of a strongly beneficial allele

is exchangeable. Thus, by de Finetti's theorem, the asymptotic frequency of ones in this sequence exists a.s., which we denote by $\underline{F}^{\underline{x},\tau} = (F_i^{\underline{x},\tau})_{i=1,\dots,d}$ with

$$F_i^{\underline{x},\tau} := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{J_{ij} \in \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}})\}} \quad (3.13)$$

Lemma 3.9 (Asymptotic frequencies and the solution of (2.1)). *For $\underline{x} \in [0, 1]^d \setminus \{0\}$, let $\underline{F}^{\underline{x},\tau}$ be as in (3.13). Then, for the solution \underline{X} of (2.1) and $\tau \geq 0$,*

$$\mathbb{P}(\underline{F}^{\underline{x},\tau} \in (\cdot)) = \mathbf{P}_{\underline{x}}(\underline{X}(\tau) \in (\cdot)). \quad (3.14)$$

Proof. From (3.10), for all $\underline{k} \in \mathbb{N}_0^d \setminus \{0\}$, the process $\mathcal{A}^{\underline{k}}$ can be seen as embedded in \mathcal{A}^∞ , if we write

$$\mathcal{A}_0^{\underline{k}} := \bigcup_{i=1}^d \{(i, U_{ig}) : 1 \leq g \leq k_i\} \subset \mathcal{A}_0^\infty. \quad (3.15)$$

By exchangeability of the sequence (3.12) and de Finetti's theorem (cf. Remark 3.8) we obtain

$$\mathbb{E}[(\underline{1} - \underline{F}^{\underline{x},\tau})^{\underline{k}}] = \mathbb{P}(\mathcal{A}_0^{\underline{k}} \cap \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}}) = \emptyset). \quad (3.16)$$

Since the right-hand sides of (3.16) and (3.9) are equal, we conclude from Lemma 3.7 that

$$\mathbb{E}[(\underline{1} - \underline{F}^{\underline{x},\tau})^{\underline{k}}] = \mathbf{E}_{\underline{x}}[(\underline{1} - \underline{X}(\tau))^{\underline{k}}]$$

which shows (3.14), since $\underline{k} \in \mathbb{N}_0^d \setminus \{0\}$ was arbitrary. \square

Under \mathbb{P} we have $\underline{F}^{\underline{x},\tau} = \underline{1}$ a.s. if and only if for all $i = 1, \dots, d$ the sequences $(\mathbb{1}_{\{J_{i1} \in \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}})\}}, \mathbb{1}_{\{J_{i2} \in \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}})\}}, \dots)$ consist of ones a.s. Hence the events $\{\underline{F}^{\underline{x},\tau} = \underline{1}\}$ and $\{\mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}}) = \mathcal{A}_0^\infty\}$ are a.s. equal under \mathbb{P} . A fortiori we have

$$\mathbf{P}_{\underline{x}}(T_{\text{fix}} \leq \tau) = \mathbf{P}_{\underline{x}}(\underline{X}(\tau) = 1) = \mathbb{P}(\mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}}) = \mathcal{A}_0^\infty). \quad (3.17)$$

This equality allows to compute the probability of eventual fixation.

Corollary 3.10 (Eventual fixation). *The probability for eventual fixation of the beneficial type,*

$$h(\underline{x}) := \mathbf{P}_{\underline{x}}(T_{\text{fix}} < \infty) \quad (3.18)$$

can be represented as (using the notation introduced in Lemma 3.7)

$$h(\underline{x}) = 1 - \mathbb{E}[(\underline{1} - \underline{x})^\underline{\Psi}], \quad (3.19)$$

where

$$\underline{\Psi} \in \mathbb{N}_0^d \setminus \{0\} \text{ is Poisson}(2\alpha\rho)\text{-distributed, conditioned to be non-zero.} \quad (3.20)$$

In other words, $\underline{\Psi}$ counts the number of particles in colonies $1, \dots, d$ of the Poisson point process from Proposition 3.3. In particular, $h(\underline{x})$ is given by formula (2.4).

Proof. Since $\mathbf{P}_{\underline{x}}(T_{\text{fix}} < \infty) = \lim_{\tau \rightarrow \infty} \mathbf{P}_{\underline{x}}(T_{\text{fix}} \leq \tau)$, we can apply the representation (3.17). We have that $\underline{K}_\tau^\infty \xrightarrow{\tau \rightarrow \infty} \underline{\Psi}$, and the probability that $(\underline{K}_r^\infty)_{r \geq 0}$ between times $r = 0$ and $r = \tau$ has a "bottleneck" at which the total number of lines equals 1 converges to one; this was called the *ultimate ancestor* in [14]. Thus, as $\tau \rightarrow \infty$, the r.h.s. of (3.17) converges to the probability that at least one particle in the configuration $\underline{\Psi}$ is marked (provided all the particles at colony i are marked independently with probability x_i).

This latter probability equals the r.h.s. of (3.19). To evaluate this explicitly, we write for independent $L_i \sim \text{Poi}(2\alpha\rho_i)$, $i = 1, \dots, d$ and $\underline{L} = (L_1, \dots, L_d)$, $L = L_1 + \dots + L_d$ (see Proposition 3.3)

$$\begin{aligned} (1 - e^{-2\alpha})h(\underline{x}) &= (1 - e^{-2\alpha})(1 - \mathbb{E}[(1 - \underline{x})^\Psi]) \\ &= (1 - e^{-2\alpha}) - \mathbb{E}[(1 - \underline{x})^{\underline{L}}, \underline{L} \neq 0] \\ &= (1 - e^{-2\alpha}) - \mathbb{E}[(1 - \underline{x})^{\underline{L}}] + \mathbb{P}(L = 0) \\ &= 1 - \prod_{i=1}^d \mathbb{E}[(1 - x_i)^{L_i}] \\ &= 1 - \prod_{i=1}^d e^{-2\alpha\rho_i} e^{2\alpha\rho_i(1-x_i)} = 1 - e^{-2\alpha(x_1\rho_1 + \dots + x_d\rho_d)}, \end{aligned}$$

i.e. we have shown (2.4). □

3.7 A duality conditioned on fixation

The next lemma is the analogue of Lemma 3.7 for the conditioned diffusion \mathcal{X}^* in place of \mathcal{X} . Here, for $\underline{k} \in \mathbb{N}_0^d \setminus \{0\}$, we will use the process \mathcal{A} , which follows the dynamics \mathcal{D}^b and has the initial state $\mathcal{Y}_0 \cup \mathcal{Z}_0^{\underline{k}}$, where $\mathcal{Z}_0^{\underline{k}}$ is as in the right hand side of (3.5) and \mathcal{Y}_0 is an equilibrium state for the dynamics \mathcal{D}^b (as described in Proposition 3.3) which is independent of $\mathcal{Z}_0^{\underline{k}}$. Note that this independence guarantees that, with probability one, all labels are distinct, and hence \mathcal{Y}_0 is a.s. disjoint from $\mathcal{Z}_0^{\underline{k}}$.

In terms of \mathcal{A} , we define two processes \mathcal{Y} and $\mathcal{Z} = \mathcal{Z}^{\underline{k}}$, which follow the dynamics \mathcal{D}^b with initial states \mathcal{Y}_0 and \mathcal{Z}_0 , by setting

$$\mathcal{Z}_s = \mathcal{C}_s(\mathcal{Z}_0^{\underline{k}}) \text{ and } \mathcal{Y}_s = \mathcal{C}_s(\mathcal{Y}_0), s \geq 0.$$

We emphasize that $\mathcal{Z}^{\underline{k}} \stackrel{d}{=} \mathcal{A}^{\underline{k}}$ and $\mathcal{Y} \stackrel{d}{=} \mathcal{A}^\Psi$ due to exchangeability of particles, hence $\mathcal{Z}^{\underline{k}}$ and \mathcal{Y} constitute a coupling of $\mathcal{A}^{\underline{k}}$ and \mathcal{A}^Ψ (with disjoint initial states).

Lemma 3.11 (Duality conditioned on fixation). *Under $\mathbf{P}_{\underline{x}}$ let $\mathcal{X}^* = (\underline{X}^*(t))_{t \geq 0}$ be the solution of (2.5), started in $\underline{X}^*(0) = \underline{x}$. Under \mathbb{P} and for $\underline{k} \in \mathbb{N}_0^d \setminus \{0\}$, let \mathcal{A} , \mathcal{Y} and $\mathcal{Z} = \mathcal{Z}^{\underline{k}}$ be as described above. Then*

$$\begin{aligned} \mathbf{E}_{\underline{x}}[(1 - \underline{X}^*(\tau))^{\underline{k}}] &= \mathbb{P}(\mathcal{Z}_0^{\underline{k}} \cap \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}}) = \emptyset \mid \mathcal{Y}_0 \cap \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}}) \neq \emptyset) \\ &= \mathbb{P}(\mathcal{Z}_\tau^{\underline{k}} \cap \mathcal{A}_\tau^{\underline{x}} = \emptyset \mid \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{\underline{x}} \neq \emptyset). \end{aligned} \tag{3.21}$$

Proof. In view of Remark 3.6 the fixation probability (3.18) can be expressed as

$$h(\underline{x}) = \mathbb{P}(\mathcal{Y}_0 \cap \mathcal{C}_0(\mathcal{A}_\tau^{\underline{x}}) \neq \emptyset) = \mathbb{P}(\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{\underline{x}} \neq \emptyset), \quad \tau \geq 0. \tag{3.22}$$

The second equality in (3.21) follows right away from Remark 3.6. To show the first equality, we set out by writing the Markovian semigroup of \underline{X}^* as the h -transform of the semigroup of \underline{X} ,

$$\mathbf{E}_{\underline{x}}[(1 - \underline{X}^*(\tau))^{\underline{k}}] = \frac{\mathbf{E}_{\underline{x}}[(1 - \underline{X}(\tau))^{\underline{k}}, T_{\text{fix}} < \infty]}{\mathbf{P}_{\underline{x}}(T_{\text{fix}} < \infty)} = \frac{\mathbf{E}_{\underline{x}}[(1 - \underline{X}(\tau))^{\underline{k}} h(\underline{X}(\tau))]}{h(\underline{x})}. \tag{3.23}$$

The numerator of the right-hand side of (3.23) equals

$$\begin{aligned} \mathbf{E}_{\underline{x}}[(1 - \underline{X}(\tau))^{\underline{k}} (1 - \mathbb{E}[(1 - \underline{X}(\tau))^\Psi])] \\ = \mathbf{E}_{\underline{x}}[(1 - \underline{X}(\tau))^{\underline{k}}] - \mathbb{E} \otimes \mathbf{E}_{\underline{x}}[(1 - \underline{X}(\tau))^{\Psi + \underline{k}}]. \end{aligned} \tag{3.24}$$

Writing $(\underline{K}_r^k)_{r \geq 0}$, $(\underline{N}_r)_{r \geq 0}$ and $(\underline{G}_r)_{r \geq 0}$ for the processes of particle numbers in \mathcal{Z}^k , \mathcal{Y} and \mathcal{A} , respectively, we obtain from the duality relation (3.9) that

$$\begin{aligned} \mathbb{E} \otimes \mathbb{E}_{\underline{x}}[(\underline{1} - \underline{X}(\tau))^{\underline{N}_0 + k}] &= \mathbb{E}[\mathbb{E}[\mathbb{E}_{\underline{x}}(\underline{1} - \underline{X}(\tau))^{\underline{N}_0 + k} | \underline{N}_0]] \\ &= \mathbb{E}[\mathbb{E}[(\underline{1} - \underline{x})^{\underline{G}_\tau} | \underline{G}_0]] = \mathbb{E}[(\underline{1} - \underline{x})^{\underline{G}_\tau}]. \end{aligned}$$

Hence, again by the duality relation (3.9) and by Remark 3.6, the right hand side of (3.24) is equal to

$$\begin{aligned} \mathbb{E}[(\underline{1} - \underline{x})^{K_\tau^k}] - \mathbb{E}[(\underline{1} - \underline{x})^{\underline{G}_\tau}] &= \mathbb{P}(\mathcal{Z}_\tau^k \cap \mathcal{A}_\tau^{(\underline{x})} = \emptyset) - \mathbb{P}(\mathcal{A}_\tau^{(\underline{x})} = \emptyset) \\ &= \mathbb{P}(\mathcal{Z}_0^k \cap \mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) = \emptyset) - \mathbb{P}((\mathcal{Z}_0^k \cup \mathcal{Y}_0) \cap \mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) = \emptyset) \\ &= \mathbb{P}(\{\mathcal{Z}_0^k \cap \mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) = \emptyset\} \cap \{\mathcal{Y}_0 \cap \mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) \neq \emptyset\}). \end{aligned}$$

Combining this with (3.23), (3.24) and (3.22), we arrive at the first equality in (3.21). \square

3.8 A paintbox representation for $\underline{X}^*(\tau)$

We now lift the assertion from Lemma 3.9 about the paintbox construction of $\underline{X}(\tau)$ to $\underline{X}^*(\tau)$. For this, let the process \mathcal{A} follow the dynamics \mathcal{D}^b and have the initial state $\mathcal{Y}_0 \cup \mathcal{Z}_0^\infty$, where \mathcal{Z}_0^∞ is as in (3.10) and \mathcal{Y}_0 is an equilibrium state for the dynamics \mathcal{D}^b (as described in Proposition 3.3) which is independent of \mathcal{Z}_0^∞ . Recall from (3.13), the definition of the asymptotic frequencies $\underline{F}^{\underline{x}, \tau} = (F_i^{\underline{x}, \tau})_{i=1, \dots, d}$ of $\mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})})$ within \mathcal{A}_0 .

Lemma 3.12 (A paintbox for $\underline{X}^*(\tau)$). *Under $\mathbb{P}_{\underline{x}}$ let $\mathcal{X}^* = (\underline{X}^*(t))_{t \geq 0}$ be the solution of (2.5), started in $\underline{X}^*(0) = \underline{x}$. Under \mathbb{P} , let the process \mathcal{A} and the frequencies $\underline{F}^{\underline{x}, \tau}$ be as above. Then,*

$$\mathbb{P}_{\underline{x}}(\underline{X}^*(\tau) \in (\cdot)) = \mathbb{P}(\underline{F}^{\underline{x}, \tau} \in (\cdot) \mid \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset). \quad (3.25)$$

Proof. For $\mathcal{Z}_0^\infty = \{J_{ig} := (i, U_{ig}) : i = 1, \dots, d, g = 1, 2, \dots\}$, we observe that the sequence (3.12) is exchangeable under the measure $\mathbb{P}(\cdot \mid \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset)$, which guarantees the a.s. existence of $\underline{F}^{\underline{x}, \tau}$. We now parallel the argument in the proof of Lemma 3.9: For each $\underline{k} \in \mathbb{N}_0^d \setminus \{\underline{0}\}$, with $\mathcal{Z}_0^{\underline{k}}$ is as in the right hand side of (3.5), we have because of exchangeability

$$\mathbb{E}[(\underline{1} - \underline{F}^{\underline{x}, \tau})^{\underline{k}} \mid \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset] = \mathbb{P}(\mathcal{Z}_0^{\underline{k}} \cap \mathcal{C}_0(\mathcal{A}_\tau^{(\underline{x})}) = \emptyset \mid \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset).$$

Combining this with Lemma 3.11, and since \underline{k} was arbitrary, we obtain the assertion. \square

We are interested in the limit of (3.25) as $\underline{x} = \underline{x}(\varepsilon) \sim \varepsilon \underline{e}_\iota$ and $\varepsilon \rightarrow 0$ for a fixed $\iota \in \{1, \dots, d\}$. For brevity we write

$$\mathbb{P}^{\underline{x}, \tau}(\cdot) := \mathbb{P}(\cdot \mid \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset). \quad (3.26)$$

Remark 3.13 (Limit of small frequencies). Let \mathcal{P} be a Poisson point process on $\{1, \dots, d\} \times [0, 1]$ with intensity measure $2\alpha\rho \otimes \lambda$. (Compare with Proposition 3.3.) For $\iota \in \{1, \dots, d\}$ and $\underline{x} = \underline{x}(\varepsilon) = \varepsilon \underline{e}_\iota$, the conditional distribution of $(\mathcal{Y}_\tau, \mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x}(\varepsilon))})$ given $\{\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x}(\varepsilon))} \neq \emptyset\}$ converges, as $\varepsilon \rightarrow 0$, to the distribution of $(\mathcal{P}^{(\iota)}, \{(\iota, U)\})$, with $\mathcal{P}^{(\iota)} := \mathcal{P} \cup \{(\iota, U)\}$, and U independent of \mathcal{P} and uniformly distributed on $[0, 1]$. In particular, under the limit of $\mathbb{P}^{\varepsilon \underline{e}_\iota, \tau}$ as $\varepsilon \rightarrow 0$, with probability 1 there is exactly one marked particle in \mathcal{Y}_τ , with the location of this particle being ι . Indeed, (using the same

notation as in the proof of Corollary 3.10),

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \mathbb{P}^{\underline{x}(\varepsilon), \tau}(\#\mathcal{Y}_\tau \cap (\{\iota\} \times [0, 1])) = k &= \lim_{\varepsilon \rightarrow 0} \frac{e^{-2\alpha\rho_\iota} (2\alpha\rho_\iota)^k (1 - (1 - \varepsilon)^k) / k!}{1 - \sum_{\ell=0}^{\infty} e^{-2\alpha\rho_\iota} (2\alpha\rho_\iota)^\ell (1 - \varepsilon)^\ell / \ell!} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{e^{-2\alpha\rho_\iota} (2\alpha\rho_\iota)^k k\varepsilon / k!}{1 - e^{-2\alpha\rho_\iota} \varepsilon} \\ &= e^{-2\alpha\rho_\iota} \frac{(2\alpha\rho_\iota)^{k-1}}{(k-1)!}, \end{aligned} \tag{3.27}$$

the weight of a Poisson($2\alpha\rho_\iota$)-distribution at $k - 1$. A similar calculation shows that this also equals the limit of $\mathbb{P}^{\underline{x}(\varepsilon), \tau}(\#\mathcal{Y}_\tau \cap (\{\iota\} \times [0, 1])) = k, \#\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} = 1$) as $\varepsilon \rightarrow 0$, explaining the additional particle (ι, U) in \mathcal{Y}_τ under $\mathbb{P}^{\iota, \tau}$.

Definition 3.14 (The process \mathcal{A} with small marking probability).

- The weak limit of $\mathbb{P}^{\varepsilon e_\iota, \tau}(\mathcal{A} \in (\cdot))$ as $\varepsilon \rightarrow 0$ will be denoted by

$$\mathbb{P}^{\iota, \tau}(\mathcal{A} \in (\cdot)).$$

From the previous remark, under $\mathbb{P}^{\iota, \tau}$, there is a.s. exactly one marked particle in \mathcal{Y}_τ , with the location of this particle being ι . This particle will be denoted by \bullet .

- For each colony i , consider the configuration $\mathcal{C}_0(\{\bullet\}) \cap \mathcal{Z}_0^\infty \cap (\{i\} \times [0, 1])$, i.e. the configuration of all particles in \mathcal{Z}_0^∞ that are located in colony i and are connected with $\{\bullet\}$. By exchangeability, the relative frequency of this configuration within $\mathcal{Z}_0^\infty \cap (\{i\} \times [0, 1])$ exists, $i = 1, \dots, d$, cf. Remark 3.7.2. As before, we denote the vector of these relative frequencies by $\underline{F}^{\iota, \tau} := (F_1^{\iota, \tau}, \dots, F_d^{\iota, \tau})$.

Corollary 3.15 (Entrance laws for (2.5)). There exists a weak limit of the distribution of \mathcal{X}^* under $\mathbb{P}_{\varepsilon e_\iota}$ as $\varepsilon \rightarrow 0$, which we denote by $\mathbb{P}_0^\iota(\mathcal{X}^* \in (\cdot))$. In particular, $((\underline{X}_t^*)_{t>0}, \mathbb{P}_0^\iota)$ defines an entrance law from $\underline{0}$ for the dynamics (2.5).

Proof. As a consequence of (3.25) and the reasoning in Remark 3.13 we have

$$\mathbb{P}_{\varepsilon e_\iota}(\underline{X}^*(\tau) \in (\cdot)) = \mathbb{P}^{\varepsilon e_\iota, \tau}(\underline{F}^{\varepsilon e_\iota, \tau} \in (\cdot)) \xrightarrow{\varepsilon \rightarrow 0} \mathbb{P}^{\iota, \tau}(\underline{F}^{\iota, \tau} \in (\cdot)). \tag{3.28}$$

Together with the Markov property, this shows that there exists a weak limit of the distribution of \mathcal{X}^* under $\mathbb{P}_{\varepsilon e_\iota}$ as $\varepsilon \rightarrow 0$. Hence the result follows. \square

Remark 3.16 (Asymptotic expected frequencies). For the asymptotic frequencies, we have that $\rho_\iota \mathbb{E}^{\iota, t}[F_j^{\iota, t}] / t \xrightarrow{t \rightarrow 0} \delta_{\iota, j}$. Indeed, $\mathbb{E}^{\iota, t}[F_j^{\iota, t}]$ is the probability that a particle from \mathcal{Z}_0^∞ located on colony j belongs to $\mathcal{C}_0(\{\bullet\})$. In order for the particle to be connected to \bullet , a coalescence event within time t must occur. For small t , and up to linear order in t , this can only happen if the particle is located on the same colony, i.e. $\iota = j$. In this case, since the coalescence rate on colony ι is $1/\rho_\iota$, the result follows.

Remark 3.17 (A correction of [21]). In [21] the case of a single colony ($d = 1$) is studied. Lemma 2.4 of [21] can be seen as an analogue of our Lemma 3.12 (together with Remark 3.13). However, Lemma 2.4 of [21] neglects the effect which the conditioning on the event $\{\mathcal{Y}_\tau \cap \mathcal{A}_\tau^{(\underline{x})} \neq \emptyset\}$ has on the distribution of $\underline{\Psi}$, and works right away with the time-reversal of \mathcal{Y} in equilibrium. Our analysis shows that, in spite of this imprecision, the conclusions of the main results of [21] remain true.

As a consequence of (3.25) and (3.28) we obtain

$$\mathbb{P}_0^\iota(T_{\text{fix}} \leq \tau) = \mathbb{P}_0^\iota(\underline{X}^*(\tau) = 1) = \mathbb{P}^{\iota, \tau}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})). \tag{3.29}$$

3.9 Proof of Proposition 3.1

From (3.29) we now derive a result on how to approximate T_{fix} as $\alpha \rightarrow \infty$. The idea is that in this limit the time which it takes for \mathcal{Z}^∞ to coalesce with \mathcal{Y} is essentially negligible on the $\frac{\log \alpha}{\alpha}$ -timescale. This is captured by the following lemma, whose proof we defer to the end of the section.

Lemma 3.18 (Approximating T_{fix}). *For $\delta, \tau > 0$, let $\delta_\alpha := \delta \frac{\log \alpha}{\alpha}$ and $\tau_\alpha := \tau \frac{\log \alpha}{\alpha}$, and let \bullet be as in Definition 3.14. Then,*

$$\mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})) \leq \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})) \text{ for all } \alpha > 0, \quad (3.30)$$

$$\liminf_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})) \leq \liminf_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha + \delta_\alpha}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})). \quad (3.31)$$

The next corollary follows by combining (3.29) and Lemma 3.18.

Corollary 3.19. *For $\alpha > 0$ let S_α be a random variable with distribution function $\tau \mapsto \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\}))$, where $\tau_\alpha = \tau \frac{\log \alpha}{\alpha}$. (In the subsequent proof of Proposition 3.1 we will see that S_α has a natural interpretation as the rescaled fixation time of \bullet in the time-reversal of \mathcal{Y} .) If S_α converges in distribution as $\alpha \rightarrow \infty$ and if τ is a point of continuity of the limiting distribution function, we have*

$$\lim_{\alpha \rightarrow \infty} \mathbf{P}'_0(T_{\text{fix}} \leq \tau_\alpha) = \lim_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})). \quad (3.32)$$

Proof. The limit in the right hand side exists by assumption. If $\tau - \delta$ is a continuity point of the limiting distribution function F , then we have by (3.31) (with τ replaced by $\tau - \delta$) and again abbreviating $\delta_\alpha = \delta \frac{\log \alpha}{\alpha}$

$$\lim_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha - \delta_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})) \leq \liminf_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})).$$

Hence, working along a sequence of continuity points $\tau - \delta$ of F with $\delta \downarrow 0$, we have

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})) &= \lim_{\delta \rightarrow 0} \lim_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha - \delta_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})) \\ &\leq \liminf_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})) = \liminf_{\alpha \rightarrow \infty} \mathbf{P}'_0(T_{\text{fix}} \leq \tau_\alpha) \\ &\leq \limsup_{\alpha \rightarrow \infty} \mathbf{P}'_0(T_{\text{fix}} \leq \tau_\alpha) = \limsup_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})) \\ &\leq \lim_{\alpha \rightarrow \infty} \mathbb{P}^{\ell, \tau_\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})). \quad \square \end{aligned}$$

The preceding corollary shows that, in order to study the asymptotic distribution of T_{fix} on the $\frac{\log \alpha}{\alpha}$ -timescale, it suffices to analyse the asymptotics of the percolation probabilities of the marked particles *within the equilibrium ASG* under the (conditional) probability $\mathbb{P}^{\ell, \tau_\alpha}$. As already explained in Sec. 3.1, the link to Proposition 3.1 is now given by a time reversal argument.

Proof of Proposition 3.1. In view of (3.32), we are done once we show that, for $\tau > 0$,

$$\mathbb{P}(T \leq \tau) = \mathbb{P}^{\ell, \tau}(\mathcal{C}_0(\{\bullet\}) \supseteq \mathcal{Y}_0), \quad (3.33)$$

where T is defined in (3.3). For this, we bring the time reversal $\widehat{\mathcal{Y}}$ of $\mathcal{Y} = (\mathcal{Y}_r)_{0 \leq r \leq \tau}$ into play, which is defined by

$$\widehat{\mathcal{Y}}_s := \mathcal{Y}_{\tau-s}, \quad 0 \leq s \leq \tau.$$

Analogously, we define $\widehat{\mathcal{C}}_s(\{\bullet\}) := \mathcal{C}_{\tau-s}(\{\bullet\})$. Then, our assertion (3.33) is equivalent to

$$\mathbb{P}(T \leq \tau) = \mathbb{P}^{\ell, \tau}(\widehat{\mathcal{C}}_\tau(\{\bullet\}) \supseteq \widehat{\mathcal{Y}}_\tau). \quad (3.34)$$

We recall that the dynamics of $\widehat{\mathcal{Y}}$ in equilibrium is given by \mathscr{D}^a ; see Proposition 3.3. While for \mathcal{Y} the conditioning (3.26) is at the *terminal* time τ (and thus modifies the dynamics \mathscr{D}), the same conditioning expressed for $\widehat{\mathcal{Y}}$ happens at the *initial* time 0 and thus does effect the initial state but not the dynamics \mathscr{D}^a . The distribution of $\widehat{\mathcal{Y}}_0$ which results from this conditioning is described in Remark 3.13. Thus we observe that under $\mathbb{P}^{\iota, \tau}$, the time-reversed process $\widehat{\mathcal{Y}}$ follows the dynamics \mathscr{D}^a and has initial state $\widehat{\mathcal{Y}}_0 = \mathscr{P}^{(\iota)} = \mathscr{P} \cup \{(\iota, U)\}$, with \mathscr{P} defined in Remark 3.13 and $\bullet := (\iota, U)$.

We now put for $i = 1, \dots, d$ and $t \geq 0$

$$\widehat{N}_t^i := \# \left(\widehat{\mathcal{Y}}_t \cap (\{i\} \times [0, 1]) \right), \quad \widehat{H}_t^i := \# \left(\widehat{\mathcal{C}}_t(\{\bullet\}) \cap (\{i\} \times [0, 1]) \right). \quad (3.35)$$

Under $\mathbb{P}^{\iota, \tau}$ the process $(\widehat{N}_t, \widehat{H}_t)_{0 \leq t \leq \tau}$ with $\widehat{N}_t = (\widehat{N}_t^1, \dots, \widehat{N}_t^d)$ and $\widehat{H}_t = (\widehat{H}_t^1, \dots, \widehat{H}_t^d)$, then has the same law as the process $(\underline{L}_t, \underline{M}_t)_{0 \leq t \leq \tau}$ defined in Proposition 3.1. In particular, (3.34) is shown. \square

We prepare the proof of Lemma 3.18 by two estimates and include their (simple) proofs for convenience.

Remark 3.20 (Comparing $\underline{\Pi}$ and $\underline{\Pi} + \underline{e}_\iota$). Recall that $\underline{\Pi} = (\Pi_1, \dots, \Pi_d)$ is distributed according to d independent Poisson distributions, where $\Pi_i \sim \text{Poi}(2\alpha\rho_i)$. As above, $\underline{\Psi}$ is distributed as $\underline{\Pi}$, conditioned to be positive (compare with (3.20)) and $\underline{\Pi} + \underline{e}_\iota$ is as in Proposition 3.1. Then, (d_{TV} denoting the total variation distance)

$$\begin{aligned} d_{TV}(\underline{\Pi}, \underline{\Psi}) &= o(1), \\ d_{TV}(\underline{\Pi}, \underline{\Pi} + \underline{e}_\iota) &= o(1) \end{aligned} \quad (3.36)$$

as $\alpha \rightarrow \infty$.

Indeed: The first result is immediate since $\mathbb{P}[\underline{\Pi} = 0] = e^{-2\alpha}$. For the second result, by a second moment calculation, we have that $\Pi_\iota / (2\alpha\rho_\iota) \xrightarrow{\alpha \rightarrow \infty} 1$ in L^2 and therefore, as $\alpha \rightarrow \infty$,

$$\begin{aligned} d_{TV}(\underline{\Pi}, \underline{\Pi} + \underline{e}_\iota) &= e^{-2\alpha\rho_\iota} \sum_{k=1}^{\infty} \left| \frac{(2\alpha\rho_\iota)^k}{k!} - \frac{(2\alpha\rho_\iota)^{k-1}}{(k-1)!} \right| + o(1) \\ &= e^{-2\alpha\rho_\iota} \sum_{k=1}^{\infty} \frac{(2\alpha\rho_\iota)^k}{k!} \left| 1 - \frac{k}{2\alpha\rho_\iota} \right| + o(1) \\ &= \mathbb{E} \left[\left| 1 - \frac{\Pi_\iota}{2\alpha\rho_\iota} \right| \right] + o(1) = o(1). \end{aligned}$$

We are now ready for the

Proof of Lemma 3.18. For proving (3.30) it suffices to show that, for each $\alpha > 0$, $\mathbb{P}^{\iota, \tau\alpha}((I, U) \notin \mathcal{C}_0(\{\bullet\}), \mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})) = 0$ for a particle (I, U) taken uniformly from \mathcal{Y}_0 . To show this equality, we will prove that for all $i = 1, \dots, d$

$$\mathbb{P}^{\iota, \tau\alpha}((I, U) \notin \mathcal{C}_0(\{\bullet\}), \mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\}), I = i) = 0. \quad (3.37)$$

We write $\mathbf{p} := (I, U)$, and note that

$$R_{\mathbf{p}} := \inf\{r > 0 : \mathcal{C}_r(\{\mathbf{p}\}) \not\subseteq \{i\} \times [0, 1]\} > 0 \quad \mathbb{P}^{\iota, \tau\alpha} \text{ a.s.}$$

The idea is now that with probability 1 we will find particles in \mathcal{Z} which coalesce with $\bigcup_{r \geq 0} \mathcal{C}_r(\{\mathbf{p}\})$, without being affected by an earlier branching or coalescence with $\mathcal{Y} \setminus \bigcup_{r \geq 0} \mathcal{C}_r(\{\mathbf{p}\})$, and hence on the event $\{\mathbf{p} \notin \mathcal{C}_0(\{\bullet\})\}$ never connect to the particle \bullet . In order to achieve this, we recall that under $\mathbb{P}^{\iota, \tau\alpha}$ the dynamics of \mathcal{Z} is given by \mathscr{D}^b ,

and this also applies conditional under \mathcal{Y} for the particles in \mathcal{Z} up to the time of their possible coalescence with particles in \mathcal{Y} .

We now consider the subsystem of particles in \mathcal{Z} which initiates from all those \mathcal{Z} particles that are located in colony i at time 0, and remove from it all those particles that undergo a migration or a branching event, or coalesce with some particle in $\mathcal{Y}_r \setminus \mathcal{C}_r(\mathbf{p})$ at some time $r \geq 0$. The system of particles of \mathcal{Z} at time r which remain after this pruning (and all of which are located in colony i by construction) will be denoted by $\mathcal{Z}_r^{(i)}$.

Given \mathcal{Y} , the process $(\#\mathcal{Z}_r^{(i)})$ is up to time $R_{\mathbf{p}}$ stochastically bounded from below by a death process $(K_r)_{r \geq 0}$ entering from infinity with death rate $\frac{1}{\rho_i} \binom{k}{2} + (\alpha + \mu \sum_{j \neq i} b(i, j) + M)k$, where $M := \max\{\#\mathcal{Y}_r \cap (\{i\} \times [0, 1])\} : 0 \leq r \leq R_{\mathbf{p}}\}$. Hence, the essentially quadratic death rate guarantees that for any $c > 0$ $\lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^c K_r dr = \infty$ a.s. Indeed, $\rho_i r K_r \xrightarrow{r \rightarrow 0} 2$ a.s. by a second moment calculation, and $K_r \approx \frac{2}{\rho_i r}$ is not integrable at $r = 0$. Consequently, also $\lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^c \#\mathcal{Z}_r^{(i)} dr = \infty$ a.s., and thus with probability 1 there will be a coalescence between $\mathcal{Z}_r^{(i)}$ and $\mathcal{C}_r(\{\mathbf{p}\})$ for some $r < R_{\mathbf{p}}$.

Since on the event $\{\mathbf{p} \notin \mathcal{C}_0(\{\bullet\})\}$ the set $\mathcal{C}_r(\{\mathbf{p}\})$ is contained in the complement of $\mathcal{C}_r(\{\bullet\})$, we conclude the existence of particles in $\mathcal{Z}_0^{(i)}$ (and hence in \mathcal{Z}_0^{∞}) that belong to the complement of $\mathcal{C}_0(\{\bullet\})$. This shows (3.37).

To prove (3.31), we first note that the particle \bullet specified in Definition 3.14 is (because of the random marking) a uniform choice from the particles in $\mathcal{Y}_{\tau_{\alpha}} \cap (\{\iota\} \times [0, 1])$ under $\mathbb{P}^{\iota, \tau_{\alpha}}$, and a uniform choice from the particles in $\mathcal{Y}_{\tau_{\alpha} + \delta_{\alpha}} \cap (\{\iota\} \times [0, 1])$ under $\mathbb{P}^{\iota, \tau_{\alpha} + \delta_{\alpha}}$.

However, as noted already after formula (3.34), the conditioning at time τ_{α} , which is inherent in $\mathbb{P}^{\iota, \tau_{\alpha}}$, destroys the time-homogeneity of the dynamics of \mathcal{Y} between times 0 and τ_{α} ; consequently, under $\mathbb{P}^{\iota, \tau_{\alpha}}$ the marking probabilities in \mathcal{Z}_0^{∞} will be different from those in \mathcal{Y}_0 . In order to account for this, the strategy of our proof will be to define under the *unconditioned* probability measure \mathbb{P} particles \circ and \circ' whose distributions will turn out to be close in variation distance to that of \bullet under $\mathbb{P}^{\iota, \tau_{\alpha}}$ and under $\mathbb{P}^{\iota, \tau_{\alpha} + \delta_{\alpha}}$, respectively, and which lead to the same marking probabilities in \mathcal{Y}_0 and \mathcal{Z}_0^{∞} .

To be specific, let \circ result from a uniform pick from $(\mathcal{Y}_{\tau_{\alpha}} \cup \mathcal{Z}_{\tau_{\alpha}}^{\infty}) \cap (\{\iota\} \times [0, 1])$ provided that this set is not empty; otherwise we pick \circ uniformly from $\mathcal{Y}_{\tau_{\alpha}} \cup \mathcal{Z}_{\tau_{\alpha}}^{\infty}$. Similarly, we pick \circ' uniformly from $(\mathcal{Y}_{\tau_{\alpha} + \delta_{\alpha}} \cup \mathcal{Z}_{\tau_{\alpha} + \delta_{\alpha}}^{\infty}) \cap (\{\iota\} \times [0, 1])$ provided that this set is not empty; otherwise we pick \circ' uniformly from $\mathcal{Y}_{\tau_{\alpha} + \delta_{\alpha}} \cup \mathcal{Z}_{\tau_{\alpha} + \delta_{\alpha}}^{\infty}$.

This construction immediately implies that for any fixed $i = 1, \dots, d$, the family of events $(\{(i, U_{ig}) \in \mathcal{C}_0(\circ)\})_{(i, U_{ig}) \in \mathcal{Y}_0 \cup \mathcal{Z}_0^{\infty}}$, is exchangeable conditional under $\mathcal{Y}_0 \cup \mathcal{Z}_0^{\infty}$. We will show five properties (A)–(E) of the joint distribution of \mathcal{A} , \mathcal{Y} , \mathcal{Z}^{∞} and \circ , proceeding in two main steps proving first (A) and then (B)–(E).

(A) the total variation distance between the distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^{\infty}, \circ)_{0 \leq t \leq \tau_{\alpha}}$ under \mathbb{P} and the distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^{\infty}, \bullet)_{0 \leq t \leq \tau_{\alpha}}$ under $\mathbb{P}^{\iota, \tau_{\alpha}}$ converges to 0 as $\alpha \rightarrow \infty$. Likewise, the total variation distance between the distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^{\infty}, \circ')_{0 \leq t \leq \tau_{\alpha} + \delta_{\alpha}}$ under \mathbb{P} and the distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^{\infty}, \bullet)_{0 \leq t \leq \tau_{\alpha} + \delta_{\alpha}}$ under $\mathbb{P}^{\iota, \tau_{\alpha} + \delta_{\alpha}}$ converges to 0 as $\alpha \rightarrow \infty$.

Having achieved this, we will construct a process $(\mathcal{Z}'_r)_{0 \leq r \leq \delta_{\alpha}}$ under \mathbb{P} with the following properties:

- (B) $\mathcal{Z}'_r \subseteq \mathcal{Z}_r^{\infty}$ for all $r \in [0, \delta_{\alpha}]$,
- (C) $\{\mathcal{Z}'_{\delta_{\alpha}} \subseteq \mathcal{C}_{\delta_{\alpha}}(\circ')\} \subseteq \{\mathcal{Z}_0^{\infty} \subseteq \mathcal{C}_0(\circ')\}$;
- (D) for any $i = 1, \dots, d$, $\mathcal{Z}'_{\delta_{\alpha}}(\{i\} \times [0, 1]) = \mathcal{O}(\alpha / \log(\alpha))$ with high probability as $\alpha \rightarrow \infty$,
- (E) for any $i = 1, \dots, d$, the family of events $(\{(i, U_{ig}) \in \mathcal{C}_{\delta_{\alpha}}(\circ')\})_{(i, U_{ig}) \in \mathcal{Y}_{\delta_{\alpha}} \cup \mathcal{Z}'_{\delta_{\alpha}}}$ is exchangeable conditional under $\mathcal{Y}_{\delta_{\alpha}} \cup \mathcal{Z}'_{\delta_{\alpha}}$.

The proof of the first assertion of (A) will be achieved in several steps.

(i) We first note that because of Remark 3.20 the total variation distance between the distributions of $\mathcal{Y}_{\tau_{\alpha}}$ under $\mathbb{P}^{\iota, \tau_{\alpha}}$ and under \mathbb{P} converges to 0 as $\alpha \rightarrow \infty$.

- (ii) Now a crucial observation is that the time-reversed dynamics of $(\mathcal{Y}_t)_{0 \leq t \leq \tau_\alpha}$ under \mathbb{P} and under $\mathbb{P}^{\ell, \tau_\alpha}$ both are given by the dual jump kernel \mathcal{D}^a . Consequently, the conditional distribution of $(\mathcal{Y}_t)_{0 \leq t \leq \tau_\alpha}$ given $\mathcal{Y}_{\tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ equals that under \mathbb{P} . This shows that the variational distance between the distributions of $(\mathcal{Y}_t)_{0 \leq t \leq \tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ and under \mathbb{P} equals the variational distance between the distributions of $\mathcal{Y}_{\tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ and under \mathbb{P} .
- (iii) Next note that the conditional distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^\infty)_{0 \leq t \leq \tau_\alpha}$ given $(\mathcal{Y}_t)_{0 \leq t \leq \tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ equals that under \mathbb{P} . Hence the variational distance between the distributions of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^\infty)_{0 \leq t \leq \tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ and under \mathbb{P} equals the variational distance between the distributions of $(\mathcal{Y}_t)_{0 \leq t \leq \tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ and under \mathbb{P} .
- (iv) Combining (i)–(iii) we see that the total variation distance between the distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^\infty)_{0 \leq t \leq \tau_\alpha}$ under \mathbb{P} and the distribution of $(\mathcal{A}_t, \mathcal{Y}_t, \mathcal{Z}_t^\infty)_{0 \leq t \leq \tau_\alpha}$ under $\mathbb{P}^{\ell, \tau_\alpha}$ converges to 0 as $\alpha \rightarrow \infty$.
- (v) According to Definition 3.14, due to the random marking under $\mathbb{P}^{\ell, \tau_\alpha}$ the particle \bullet arises by a uniform choice from $\mathcal{Y}_{\tau_\alpha} \cap (\{\ell\} \times [0, 1])$. We now claim that under \mathbb{P} , on an event whose probability converges to 1 as $\alpha \rightarrow \infty$, the particle \circ constitutes a uniform choice from $\mathcal{Y}_{\tau_\alpha} \cap (\{\ell\} \times [0, 1])$. We will prove in the next section a key lemma, Lemma 4.1, which will tell us that under \mathbb{P} the number of particles in \mathcal{Y}_t in colony i , $i = 1, \dots, d$, is with high probability as $\alpha \rightarrow \infty$ concentrated around $2\rho_i\alpha$, uniformly in $t \in [0, \tau_\alpha]$. Hence our claim holds if $\#(\mathcal{Z}_{\tau_\alpha}^\infty \setminus \mathcal{Y}_{\tau_\alpha}) = o(\alpha)$ with high probability as $\alpha \rightarrow \infty$. To see this, we note that the probability of the event

$$\{\#(\mathcal{Y}_t \cap (\{i\} \times [0, 1])) \geq 2\alpha\rho_i(1 - \varepsilon) \text{ for some } \varepsilon > 0 \text{ and for all } i; 0 \leq t \leq \tau_\alpha\}$$

tends to 1 as $\alpha \rightarrow \infty$ because of Lemma 4.1. On this event, however, the process $\#(\mathcal{Z}_t^\infty \setminus \mathcal{Y}_t)_{0 \leq t \leq \tau_\alpha}$ under \mathbb{P} is stochastically bounded from above by a birth-death process which in state (k_1, \dots, k_d) with $k = k_1 + \dots + k_d$ has birth rate αk and death rate at least $\sum_{i=1}^d \frac{1}{\rho_i} \binom{k_i}{2} + \frac{1}{\rho_i} 2\alpha\rho_i(1 - \varepsilon)k_i \geq \frac{k(k-d)}{2d} + 2\alpha(1 - \varepsilon)k$, see (3.11). Hence a second moment calculation shows that, with high probability as $\alpha \rightarrow \infty$, $\#(\mathcal{Z}_{\tau_\alpha}^\infty \setminus \mathcal{Y}_{\tau_\alpha}) = O(\frac{\alpha}{\log \alpha}) = o(\alpha)$. Together with (iv), this shows the first part of the assertion of (A); the arguments for the second part of (A) are the same, with τ being replaced by $\tau + \delta$.

For (B)–(E), we define the particle system $(\mathcal{Z}'_t)_{0 \leq t \leq \delta_\alpha}$ as a subsystem of $(\mathcal{Z}_t^\infty)_{0 \leq t \leq \delta_\alpha}$ (from which property (B) is automatic). As its initial state we take $\mathcal{Z}'_0 := \mathcal{Z}_0^\infty$. We then impose the rule that the particles in \mathcal{Z}' perform all coalescence and migration events dictated by \mathcal{Z} , but follow only a single one of the two particles in \mathcal{Z} upon a branching event. More formally,

- if $(i, U_{ig}), (i, U_{ig'}) \in \mathcal{Z}_{r-}^\infty$ coalesce, i.e. are replaced by $(i, U_{ig''}) \in \mathcal{Z}_{r-}^\infty$, and if $(i, U_{ig}), (i, U_{ig'}) \in \mathcal{Z}'_{r-}$, then the same replacement happens in \mathcal{Z}'_{r-} ,
- if $(i, U_{ig}), (i, U_{ig'}) \in \mathcal{Z}_{r-}^\infty \cup \mathcal{Y}_{r-}$ coalesce, i.e. are replaced by $(i, U_{ig''}) \in \mathcal{Z}_{r-}^\infty$, and if only $(i, U_{ig}) \in \mathcal{Z}'_{r-}$ but $(i, U_{ig'}) \notin \mathcal{Z}'_{r-}$, then $(i, U_{ig}) \in \mathcal{Z}'_{r-}$ is replaced by $(i, U_{ig''})$ in \mathcal{Z}'_{r-} ,
- if $(i, U_{ig}) \in \mathcal{Z}_{r-}^\infty$ migrates to j , i.e. is replaced by $(j, U_{jg'})$ in \mathcal{Z}_{r-}^∞ , and if $(i, U_{ig}) \in \mathcal{Z}'_{r-}$, the particle also migrates to j in \mathcal{Z}'_{r-} , i.e. (i, U_{ig}) is replaced by $(j, U_{jg'})$ in \mathcal{Z}'_{r-} ,
- if $(i, U_{ig}) \in \mathcal{Z}_{r-}^\infty$ branches, i.e. is replaced by $(i, U_{ig'}), (i, U_{ig''}) \in \mathcal{Z}_{r-}^\infty$, and if $(i, U_{ig}) \in \mathcal{Z}'_{r-}$, then (i, U_{ig}) is replaced by $(i, U_{ig'})$ in \mathcal{Z}'_{r-} .

Note that $\mathcal{Z}'_0 \subseteq \mathcal{C}_0(\mathcal{Z}'_{\delta_\alpha})$ by construction, so if $\mathcal{Z}'_{\delta_\alpha} \subseteq \mathcal{C}_{\delta_\alpha}(\circ')$ then $\mathcal{Z}'_0 \subseteq \mathcal{C}_0(\mathcal{C}_{\delta_\alpha}(\circ')) = \mathcal{C}_0(\circ')$, i.e. we have property (C). Since \mathcal{Z}' is a coalescing random walk, it is a death process which in state (k_1, \dots, k_d) with $k = k_1 + \dots + k_d$ has death rate (using (3.11)) $\sum_{i=1}^d \frac{1}{\rho_i} \binom{k_i}{2} \geq \frac{k(k-d)}{2d}$. A second moment calculation then shows (D). Finally, the exchangeability claimed in (E) holds by construction.

Based on properties (A)–(E) we can now prove (3.31). Indeed, because of (A)

$$\mathbb{P}^{\iota, \tau\alpha}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\bullet\})) = \mathbb{P}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\circ\})) + o(1) \text{ as } \alpha \rightarrow \infty. \quad (3.38)$$

From the stationarity of \mathcal{Y} under \mathbb{P} together with property (A),

$$\mathbb{P}(\mathcal{Y}_0 \subseteq \mathcal{C}_0(\{\circ\})) = \mathbb{P}(\mathcal{Y}_{\delta\alpha} \subseteq \mathcal{C}_{\delta\alpha}(\{\circ'\})) + o(1) \text{ as } \alpha \rightarrow \infty. \quad (3.39)$$

For all fixed $i \in \{1, \dots, d\}$, consider the event

$$E_{i,\alpha} := \{(\mathcal{Y}_{\delta\alpha}(\{i\} \times [0, 1]) \geq \rho_i\alpha, \mathcal{Z}'_{\delta\alpha}(\{i\} \times [0, 1]) \leq \rho_i\alpha\}.$$

Then because of the exchangeability property (E) we have

$$\mathbb{P}(\mathcal{Y}_{\delta\alpha} \cap (\{i\} \times [0, 1]) \subseteq \mathcal{C}_{\delta\alpha}(\{\circ'\}) \mid E_{i,\alpha}) \leq \mathbb{P}(\mathcal{Z}'_{\delta\alpha} \cap (\{i\} \times [0, 1]) \subseteq \mathcal{C}_{\delta\alpha}(\{\circ'\}) \mid E_{i,\alpha}).$$

Because of property (D) we have $\mathbb{P}(E_{i,\alpha}) \rightarrow 1$ as $\alpha \rightarrow \infty$, and consequently

$$\liminf_{\alpha \rightarrow \infty} \mathbb{P}(\mathcal{Y}_{\delta\alpha} \subseteq \mathcal{C}_{\delta\alpha}(\{\circ'\})) \leq \liminf_{\alpha \rightarrow \infty} \mathbb{P}(\mathcal{Z}'_{\delta\alpha} \subseteq \mathcal{C}_{\delta\alpha}(\{\circ'\})). \quad (3.40)$$

Property (C) yields

$$\mathbb{P}(\mathcal{Z}'_{\delta\alpha} \subseteq \mathcal{C}_{\delta\alpha}(\{\circ'\})) \leq \mathbb{P}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\circ'\})) \quad (3.41)$$

and property (A) implies

$$\mathbb{P}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\circ'\})) = \mathbb{P}^{\iota, \tau\alpha + \delta\alpha}(\mathcal{Z}_0^\infty \subseteq \mathcal{C}_0(\{\bullet\})) + o(1) \text{ as } \alpha \rightarrow \infty. \quad (3.42)$$

Combining (3.38)–(3.42) we arrive at (3.31). \square

3.10 Proof of Theorem 1

Let $\underline{x} \neq \underline{0}$. Then equation (3.21) shows that the one-dimensional distributions of \mathcal{X}^* are determined. This shows the uniqueness (see Theorem 4.4.2 of [7]).

Now let $(\mathcal{X}^*, \mathbf{P})$ with $\mathcal{X}^* = (\underline{X}^*(t))_{t \geq 0}$ be an entrance law from $\underline{0}$ for the dynamics (2.5). For fixed $t > 0$ and $0 < \delta < t$ we can represent $\mathbf{P}(\underline{X}^*(t) \in (\cdot))$ by means of (3.25), putting $\tau := t - \delta$ and using the “random paintbox” \underline{X}_δ^* instead of the deterministic \underline{x} figuring in (3.25). More specifically, we have by the Markov property of \mathcal{X}^*

$$\begin{aligned} \mathbf{P}(\underline{X}^*(t) \in (\cdot)) &= \mathbf{E}[\mathbf{P}_{\underline{X}^*(\delta)}(\underline{X}^*(t - \delta) \in (\cdot))] \\ &= \mathbf{E}[\mathbf{P}(\underline{F}^{\underline{X}^*(\delta), t - \delta} \in (\cdot) \mid \mathcal{Y}_{t - \delta} \cap \mathcal{A}_{t - \delta}^{(\underline{X}^*(\delta))} \neq \emptyset \mid \underline{X}^*(\delta))]. \end{aligned} \quad (3.43)$$

Now consider the random vector $\mathcal{N}_\delta := (\mathcal{Y}_{t - \delta}(\{i\} \times [0, 1]))_{i=1, \dots, d}$, and write $\nu_\delta^{\underline{X}^*(\delta)}$ for the distribution of \mathcal{N}_δ conditioned under the event $\{\mathcal{Y}_{t - \delta} \cap \mathcal{A}_{t - \delta}^{(\underline{X}^*(\delta))} \neq \emptyset\}$ for given $\underline{X}^*(\delta)$. We recall that the unconditional distribution of $\mathcal{Y}_{t - \delta}$ is the distribution π described in Proposition 3.3. Thus we are faced with a Poisson coloring, where the coloring is rare (due to the assumption that $\underline{X}^*(\delta) \rightarrow \underline{0}$ in probability as $\delta \rightarrow 0$) but conditioned to produce at least one colored particle. Using the notation $\underline{\Pi}$ for a Poisson vector as in Proposition 3.3, we infer that there exist $\{1, \dots, d\}$ -valued random variables J_δ independent of $\underline{\Pi}$ such that the total variation distance between $\nu_\delta^{\underline{X}^*(\delta)}$ and the distribution of $\underline{\Pi} + \underline{e}_{J_\delta}$ converges to 0 as $\delta \rightarrow 0$. We thus obtain from (3.43) for all $t > 0$

$$\mathbf{P}(\underline{X}^*(t) \in (\cdot)) = \mathbf{E}[\mathbf{P}^{J_\delta, t - \delta}(\underline{F}^{J_\delta, t - \delta} \in (\cdot))] + o(1) \quad \text{as } \delta \rightarrow 0. \quad (3.44)$$

Because of compactness, there is a sequence $\delta_n \rightarrow 0$, and an $\{1, \dots, d\}$ -valued random variable J such that $J_{\delta_n} \xrightarrow{n \rightarrow \infty} J$. By continuity, we thus obtain from (3.44) the representation

$$\mathbf{P}(\underline{X}^*(t) \in (\cdot)) = \mathbf{E}[\mathbf{P}^{J, t}(\underline{F}^{J, t} \in (\cdot))], \quad t > 0. \quad (3.45)$$

We claim that this representation is unique. Indeed, let J' be a $\{1, \dots, d\}$ -valued random variable whose distribution is different from that of J , and which obeys

$$\mathbf{E}[\mathbb{P}^{J,t}(\underline{F}^{J,t} \in (.))] = \mathbf{E}[\mathbb{P}^{J',t}(\underline{F}^{J',t} \in (.))], \quad t > 0. \quad (3.46)$$

Then there must exist an $i \in \{1, \dots, d\}$ such that $\mathbf{P}(J = i) < \mathbf{P}(J' = i)$. On the other hand, from Remark 3.16,

$$\limsup_{t \rightarrow 0} \frac{\mathbf{E}[\mathbb{E}^{J,t}[F_i^{J,t}]]}{\mathbf{E}[\mathbb{E}^{J',t}[F_i^{J',t}]]} = \limsup_{t \rightarrow 0} \frac{\sum_{j=1}^d \mathbf{P}(J = j) \mathbb{E}^{j,t}[F_i^{j,t}]}{\sum_{j=1}^d \mathbf{P}(J' = j) \mathbb{E}^{j,t}[F_i^{j,t}]} = \frac{\mathbf{P}(J = i)}{\mathbf{P}(J' = i)} < 1, \quad (3.47)$$

which contradicts (3.46).

From (3.45) and (3.28) we obtain the representation

$$\mathbf{P}(\underline{X}^*(t) \in (.)) = \mathbf{E}[\mathbb{P}^{J,t}(\underline{F}^{J,t} \in (.))] = \mathbf{E}[\mathbf{P}_0^J(\underline{X}^*(t) \in (.))], \quad t > 0,$$

which shows that every entrance law from $\underline{0}$ is a convex combination of the entrance laws $\mathbf{P}_0^i(\underline{X}^* \in (.))$, $i = 1, \dots, d$. To see the extremality of the latter, note that by the same reasoning which led to the contradiction of (3.46) and (3.47), the equality

$$\mathbf{P}_0^i(\underline{X}^*(t) \in (.)) = \mathbf{E}[\mathbf{P}_0^J(\underline{X}^*(t) \in (.))], \quad t > 0$$

is impossible unless $\mathbf{P}(J = i) = 1$. This completes the proof of Theorem 1.

4 Proof of Theorem 2

4.1 Heuristics

Before we come to the formal proofs, we give a summary of all three cases. Some basic ideas will be formalised in a few lemmas that are collected in Section 4.2. The basis of our proof is the ancestral selection graph and the approximate representation of the fixation time in Proposition 3.1. Moreover, by our interpretation of the d extremal entrance laws (see Remark 2.3) and symmetry, we can consider the situation when the ASG has a single marked particle in colony 1. Recall from Definition 3.14 that this marked particle \bullet is of the form $(1, U)$ for a $[0,1]$ -uniformly distributed U .

It is important to note that at all times during the sweep, L_t^i from Proposition 3.1 (which is the same as the number of particles in \mathcal{Y} with jump kernel \mathcal{D}^a from Section 3.2, started in $\mathcal{P} \cup \{\bullet\}$) in colony i is about $2\alpha\rho_i$ with high probability, see Lemma 4.1. Within \mathcal{Y} , we distinguish between marked particles (comprising $\underline{M}_t = (M_t^1, \dots, M_t^d)$ with $M_t^i := \#(\mathcal{C}_t(\{\bullet\}) \cap (\{i\} \times [0, 1]))$) and wildtype particles; see also (3.35).

Let us turn to case 1. Here, migration happens at rate of order α . Since splitting events of marked particles in $(\underline{M}_t)_{t \geq 0}$ happen at rate α as well, marked particles are present quickly (i.e. after time of order $1/\alpha$) in all colonies. More precisely, the number of particles of the \mathcal{B} allele $(M_1(t) + \dots + M_d(t))_{t \geq 0}$ is close to a pure branching process with branching rate α in this starting phase. Then, when the number of particles exceeds $\alpha\varepsilon$ (for some small ε), the particles start to coalesce and the process is not pure branching any more. The time when this happens is roughly $(\log(\varepsilon\alpha))/\alpha \approx \log(\alpha)/\alpha$; compare with Lemma 4.4. Rescaling time by a factor of α , we can see – using an ordinary differential equation – that the time the system needs to reach at least $2\alpha\rho_i(1 - \varepsilon)$ particles in colony i , $i = 1, \dots, d$, is of order $1/\alpha$ and hence is negligible for our claim. When there are $2\alpha\rho_i(1 - \varepsilon)$ marked particles in colony i , there are about $\varepsilon 2\alpha$ wildtype particles in total. Any wildtype line performs a subcritical branching process with splitting rate α (which is the splitting rate within the ASG) and death rate at least $\frac{1}{\rho_i} 2\alpha\rho_i(1 - \varepsilon) = 2\alpha(1 - \varepsilon)$ (which is the coalescence rate with one of the $2\alpha\rho_i(1 - \varepsilon)$ marked particles within the same

colony). The extinction time of such a subcritical branching process can be computed to be about $\log(\alpha)/\alpha$; see Lemma 4.7. In total, this gives a fixation time $2\log(\alpha)/\alpha$.

Now we come to case 2, where migration happens at rate of order α^γ . For simplicity let us consider the case of two colonies first. The number of marked particles increases exponentially at rate α in colony 1, so the number of particles at time $(1 - \gamma)\log(\alpha)/\alpha$ is $\exp((1 - \gamma)\log(\alpha)) = \alpha^{1-\gamma}$. Since the migration rate is of the order α^γ , the first migrant to colony 2 arises exactly by that time. Indeed, the total rate of migration is of order $\alpha^{1-\gamma}\alpha^\gamma = \alpha$, but at time $(1 - \gamma - \varepsilon)\log(\alpha)/\alpha$ the total migration rate was only $\alpha^{1-\gamma-\varepsilon}\alpha^\gamma = \alpha^{1-\varepsilon}$. Moreover, we note that at time $(1 - \gamma + \varepsilon)\log(\alpha)/\alpha$ there are already $\alpha^{1+\varepsilon}$ migrants, such that the first migrant occurs around time $(1 - \gamma)\log(\alpha)/\alpha$. After the first migrant arises, its offspring starts to expand exponentially at rate α in colony 2. After another time $x\log(\alpha)/\alpha$, it increased in frequency to α^x particles. Moreover, the number of migrants from colony 1 (in the case $x < \gamma$, i.e. during the exponential growth phase in colony 1) is $\int_0^{x\log(\alpha)/\alpha} \alpha^{1-\gamma} e^{\alpha t} \alpha^\gamma dt \approx \alpha^x$ which indicates that the number of marked particles in colony 2 is of order α^x by time $(1 - \gamma + x)\log(\alpha)/\alpha$ for $x < \gamma$; see also 2. in Lemma 4.4. After time $\log(\alpha)/\alpha$, the exponential growth phase in colony 1 is over and the marked particles in colony 2 still increase exponentially due to splitting events in colony 2. At time $(2 - \gamma)\log(\alpha)/\alpha$, the exponential growth phase in both colonies is over and – as in case 1 – it takes time of order $1/\alpha$ until there are at least $2\alpha\rho_i(1 - \varepsilon)$ particles in colony i , $i = 1, 2$. Again, we can consider the total number of wildtype particles and approximate it by a subcritical branching process which dies after time about $\log(\alpha)/\alpha$; see again Lemma 4.7. Hence, the fixation time is about $(3 - \gamma)\log(\alpha)/\alpha$.

For more than two colonies, it is clear that infection of a new colony happens if and only if a neighbouring colony has about $\alpha^{1-\gamma}$ marked particles, which happens some time $(1 - \gamma)\log(\alpha)/\alpha$ after this colony was infected. This leads to the first epidemic model.

For case 3, where migration happens at rate of order $1/(\log \alpha)$, observe that the total number of migration events between colonies in a time of order $\log(\alpha)/\alpha$ is of order 1 (since there are of order α particles per colony, each of which has a migration rate of order $1/\log \alpha$). Again, we start by considering two colonies, $\mu = c/(\log \alpha)$, and consider the process on the new time-scale $d\tau = \frac{\alpha}{\log \alpha} dt$. If the number of marked particles in colony 1 is smaller than α , migration of a marked particle is unlikely. At time $\tau = 1$, however, there are about $2\rho_1\alpha$ marked particles in colony 1, each of which migrates at rate c/α (on time-scale $d\tau$), leading to an effective rate $2c\rho_1$ of migration. This means we have to wait an exponential waiting time with rate $2c\rho_1$ for the first migrant. After that time, the marked particles have already fixed in colony 1, but colony 2 needs another 2 time-units (on the time-scale $d\tau$) before fixation.

For d colonies, note that a new colony k gets infected, if a migrant from another infected colony is successful. After time $\tau = 1$, enough particles have accumulated on this colony such that it can send migrants to its neighbouring colonies, hence becomes infectious. If it is infectious, it sends migrants at rate $2\rho_k a(k, j)$ to colony j , which is exactly the second epidemic model.

4.2 Some lemmas

We now state some general lemmas, which are used in the proof of Theorem 2. Recall that $\underline{\rho} = (\rho_1, \dots, \rho_d)$ constitutes the equilibrium distribution for the migration dynamics.

Lemma 4.1 (\underline{L} concentrated around $2\alpha\rho$). Assume $t_\alpha \downarrow 0$ and let $\underline{L} = (\underline{L}_t)_{t \geq 0}$ with $\underline{L}_t = (L_t^1, \dots, L_t^d)$ follow the same dynamics as in Proposition 3.1. (Recall that this process depends on the parameters α and μ .)

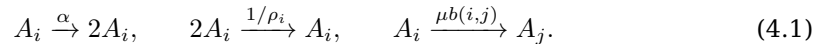
Let $\mu = O(\alpha)$, $\varepsilon_\alpha \downarrow 0$ be any sequence such that $t_\alpha/\varepsilon_\alpha \rightarrow 0$ and $\mathbb{P}\left(\left|\frac{L_0}{\alpha} - 2\underline{\rho}\right| > \varepsilon_\alpha^2\right) \rightarrow 0$.

Then,

$$\lim_{\alpha \rightarrow \infty} \mathbb{P} \left(\sup_{0 \leq r \leq t_\alpha} \left| \frac{L_r}{\alpha} - 2\rho \right| > \varepsilon_\alpha \right) = 0.$$

Before turning to the proof of this lemma, let us observe that a sequence $\varepsilon_\alpha \downarrow 0$, which fulfills the requirements of Lemma 4.1, exists iff $L_0/\alpha \xrightarrow{\alpha \rightarrow \infty} 2\rho$.

Remark 4.2 (A Lyapunov function for the limiting system). In the proof of the lemma, a function h arises; see (4.3). In order to understand the form of this function, consider a chemical reaction network for chemical species A_1, \dots, A_d , governed by



for $i, j = 1, \dots, d$. Here, the chemical species A_i refers to the particles in colony i . (We refer the reader to [9] for general notions of chemical reaction network theory.) For mass action kinetics, properly rescaled, the vector of concentrations $\underline{c} = (c_1, \dots, c_d)$ with c_i being the concentration of species A_i satisfies the dynamical system

$$\dot{c}_i = \alpha c_i - \frac{1}{2\rho_i} c_i^2 + \mu \sum_{j \neq i} c_j b(j, i) - c_i b(i, j), \quad i = 1, \dots, d. \quad (4.2)$$

Since the system (4.1) is weakly reversible and complex balanced, local asymptotic stability has been shown via the Lyapunov function $h(\underline{c}) = \sum_{i=1}^d ((\log(c_i/c_i^*) - 1)c_i + c_i^*)$, see Proposition 5.3 in [9], where (c_1^*, \dots, c_d^*) denotes the equilibrium value of (4.2). In fact, with $\kappa_i = c_i$ and $2\rho_i = c_i^*$, this is the function h appearing in (4.3) below.

Proof of Lemma 4.1. The generator of $L^\alpha := L/\alpha$ is

$$\begin{aligned} G_{L^\alpha} f(\underline{\kappa}) &= \alpha^2 \sum_{i=1}^d \left(\kappa_i (f(\underline{\kappa} + \underline{e}_i/\alpha) - f(\underline{\kappa})) \right. \\ &\quad \left. + \frac{\kappa_i(\kappa_i - 1/\alpha)}{2\rho_i} (f(\underline{\kappa} - \underline{e}_i/\alpha) - f(\underline{\kappa})) \right) \\ &\quad + \mu \alpha \sum_{i,j=1}^d b(i, j) \kappa_i (f(\underline{\kappa} + \underline{e}_j/\alpha - \underline{e}_i/\alpha) - f(\underline{\kappa})) \end{aligned}$$

for functions $f : \mathbb{R}_+^d \rightarrow \mathbb{R}$. Now, define

$$h(\underline{\kappa}) = \sum_{i=1}^d \left(\left(\log \left(\frac{\kappa_i}{2\rho_i} \right) - 1 \right) \kappa_i + 2\rho_i \right) = 2 + \sum_{i=1}^d \left(\log \left(\frac{\kappa_i}{2\rho_i} \right) - 1 \right) \kappa_i. \quad (4.3)$$

This function is strictly convex and vanishes if and only if $\underline{\kappa} = 2\rho$. Hence we are done once we show that $\sup_{0 \leq r \leq t_\alpha} h(L_r^\alpha) \xrightarrow{\alpha \rightarrow \infty} 0$ in probability. For this, we will make use of Doob's maximal inequality for sub-martingales and some calculations using the generator of L^α . Since $\log(x + \delta) \leq \log(x) + \frac{\delta}{x}$, for $i, j = 1, \dots, d$ and $i \neq j$,

$$\begin{aligned} h(\underline{\kappa} \pm \underline{e}_i/\alpha) - h(\underline{\kappa}) &= \left(\log \left(\frac{\kappa_i \pm 1/\alpha}{2\rho_i} \right) - \log \left(\frac{\kappa_i}{2\rho_i} \right) \right) (\kappa_i \pm \frac{1}{\alpha}) \\ &\quad \pm \frac{1}{\alpha} \left(\log \left(\frac{\kappa_i}{2\rho_i} \right) - 1 \right) \\ &= \pm \frac{1}{\alpha} \left(\log \left(\frac{\kappa_i \pm 1/\alpha}{2\rho_i} \right) - 1 \right) + \kappa_i \log \left(1 \pm \frac{1}{\alpha \kappa_i} \right) \\ &\leq \pm \frac{1}{\alpha} \log \left(\frac{\kappa_i \pm 1/\alpha}{2\rho_i} \right), \\ h(\underline{\kappa} + \underline{e}_j/\alpha - \underline{e}_i/\alpha) - h(\underline{\kappa}) &\leq \frac{1}{\alpha} \left(\log \left(\frac{\kappa_j + 1/\alpha}{2\rho_j} \right) - \log \left(\frac{\kappa_i - 1/\alpha}{2\rho_i} \right) \right). \end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{i,j=1}^d b(i,j) \left(\kappa_j \frac{\rho_i}{\rho_j} - \kappa_i \right) &= \sum_{j=1}^d \frac{\kappa_j}{\rho_j} \sum_{i=1}^d \rho_i b(i,j) - \sum_{i,j=1}^d \kappa_i b(i,j) \\ &= \sum_{j=1}^d \frac{\kappa_j}{\rho_j} \sum_{i=1}^d \rho_j b(j,i) - \sum_{i,j=1}^d \kappa_i b(i,j) \\ &= \sum_{i,j=1}^d \kappa_j b(j,i) - \kappa_i b(i,j) = 0, \end{aligned}$$

Hence, using that $\log(x) \leq x - 1$ and $(1 - x) \log(x) \leq 0$ for all $x \geq 0$, we obtain for sufficiently large α and for $\underline{\kappa} \in A := (\rho_1, 4\rho_1) \times \cdots \times (\rho_d, 4\rho_d)$

$$\begin{aligned} G_{\underline{L}^\alpha} h(\underline{\kappa}) &\leq \alpha \sum_{i=1}^d \kappa_i \log \left(\frac{\kappa_i + 1/\alpha}{2\rho_i} \right) - \frac{\kappa_i(\kappa_i - 1/\alpha)}{2\rho_i} \log \left(\frac{\kappa_i - 1/\alpha}{2\rho_i} \right) \\ &\quad + \mu \sum_{i,j=1}^d b(i,j) \kappa_i \underbrace{\left(\log \left(\frac{\kappa_j + 1/\alpha}{2\rho_j} \right) - \log \left(\frac{\kappa_i - 1/\alpha}{2\rho_i} \right) \right)}_{\leq \frac{(\kappa_j + 1/\alpha)\rho_i}{(\kappa_i - 1/\alpha)\rho_j} - 1} \\ &\leq \sum_{i=1}^d \alpha \kappa_i \underbrace{\left(\log \left(\frac{\kappa_i - 1/\alpha}{2\rho_i} \right) - \frac{\kappa_i - 1/\alpha}{2\rho_i} \log \left(\frac{\kappa_i - 1/\alpha}{2\rho_i} \right) \right)}_{\leq 0} + \frac{2\alpha\kappa_i}{\alpha(\kappa_i - 1/\alpha)} \tag{4.4} \\ &\quad + \mu \sum_{i,j=1}^d b(i,j) \left(\kappa_j \frac{\rho_i}{\rho_j} - \kappa_i \right) + C \frac{\mu}{\alpha} \sum_{i,j=1}^d b(i,j) \frac{(\kappa_i + \kappa_j)\rho_i}{\kappa_i\rho_j} \\ &\leq C' \end{aligned}$$

for some $C, C' > 0$ which are independent of all parameters; recall that $\mu = \mathcal{O}(\alpha)$ by assumption. Note that (4.4) shows that $(G_{\underline{K}_\alpha} h)^+$ is bounded uniformly by C' for all α on the set A . Now, consider the martingale (recall that $g = g^+ - g^-$ with $g^+ = g \vee 0$ and $g^- = (-g)^+ \geq 0$)

$$\begin{aligned} &\left(h(\underline{L}^\alpha(r \wedge T_A)) - \int_0^{r \wedge T_A} (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s))) ds \right)_{r \geq 0} \\ &= \left(h(\underline{L}^\alpha(r \wedge T_A)) + \int_0^{r \wedge T_A} (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s)))^- - (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s)))^+ ds \right)_{r \geq 0}, \end{aligned}$$

which is stopped when \underline{L}^α leaves the set A at the stopping time T_A . Clearly, since $h \geq 0$,

$$\left(h(\underline{L}^\alpha(r \wedge T_A)) + \int_0^{r \wedge T_A} (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s)))^- \right)_{r \geq 0}$$

is a positive submartingale. We restrict the initial state $\underline{L}^\alpha(0)$ to be in the set A (this event has probability converging to 1 as $\alpha \rightarrow \infty$). Note that, by assumption, we find some $C'' > 0$ such that $\mathbb{E}[h(\underline{L}^\alpha(0))] \leq C'' \varepsilon_\alpha^2$ and $\frac{t_\alpha}{\varepsilon_\alpha} \xrightarrow{\alpha \rightarrow \infty} 0$. By Doob's martingale inequality, for $t_\alpha \downarrow 0$ and if ε is small enough, for $\underline{L}^\alpha(0) \in A$,

$$\begin{aligned} \mathbb{P} \left(\sup_{0 \leq r \leq t_\alpha} h(\underline{L}^\alpha(r)) > \varepsilon_\alpha \right) &= \mathbb{P} \left(\sup_{0 \leq r \leq t_\alpha} h(\underline{L}^\alpha(r \wedge T_A)) > \varepsilon_\alpha \right) \\ &\leq \mathbb{P} \left(\sup_{0 \leq r \leq t_\alpha} h(\underline{L}^\alpha(r \wedge T_A)) + \int_0^{r \wedge T_A} (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s)))^- ds > \varepsilon_\alpha \right) \end{aligned}$$

The fixation time of a strongly beneficial allele

$$\begin{aligned} &\leq \frac{1}{\varepsilon_\alpha} \mathbb{E} \left[h(\underline{L}^\alpha(t_\alpha \wedge T_A)) + \int_0^{t_\alpha \wedge T_A} (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s)))^- ds \right] \\ &= \frac{1}{\varepsilon_\alpha} \mathbb{E} \left[h(\underline{L}^\alpha(0)) + \int_0^{t_\alpha \wedge T_A} (G_{\underline{L}^\alpha} h(\underline{L}^\alpha(s)))^+ ds \right] \\ &\leq \frac{C'' \varepsilon_\alpha^2 + C' t_\alpha}{\varepsilon_\alpha} \xrightarrow{\alpha \rightarrow \infty} 0 \end{aligned}$$

and the result follows. □

We also need a little refinement of the last lemma. Here, only bounds on the birth and death rates are assumed.

Corollary 4.3 (Particle-counting in a single colony concentrated around $2\alpha\rho$). *Let $\mathcal{V} = (V_r)_{r \geq 0}$ be a birth-death process with birth- and death rates b_k and d_k satisfying*

$$\alpha k \leq b_k \leq \alpha k + c\alpha^{1+\gamma}, \quad \frac{1}{\rho} \binom{k}{2} \leq d_k \leq \frac{1}{\rho} \binom{k}{2} + c\alpha^\gamma k$$

for some $\gamma \in [0, 1)$ and $c \geq 0, \rho > 0$. If $V_0/\alpha \xrightarrow{\alpha \rightarrow \infty} 2\rho$, then

$$\sup_{0 \leq r \leq t_\alpha} \left| \frac{V_r}{\alpha} - 2\rho \right| \xrightarrow{\alpha \rightarrow \infty} 0$$

for $t_\alpha \downarrow 0$.

Proof. For $c = 0$, the assertion would just be a special case of Lemma 4.1 for a single colony. For $c > 0$, we fix $\varepsilon > 0$ and take α large enough such that

$$\alpha k \leq b_k \leq (\alpha + c'\alpha^\gamma)k, \quad \frac{1}{\rho} \binom{k}{2} \leq d_k \leq \frac{1}{\rho}(1 + \varepsilon) \binom{k}{2}$$

for some $c' > 0$ whenever $k \in [\alpha\rho, 4\alpha\rho]$. Now consider the process $\mathcal{V}' = (V'_r)_{r \geq 0}$ ($\mathcal{V}'' = (V''_r)_{r \geq 0}$) with the lower (upper) bound of b_k and the upper (lower) bound of d_k as birth- and death rates. Clearly, the processes $\mathcal{V}, \mathcal{V}', \mathcal{V}''$ can be coupled such that $V'_r \leq V_r \leq V''_r$ for all r as long as $V_r, V'_r, V''_r \in [\alpha\rho, 4\alpha\rho]$ and conclude from Lemma 4.1 (by suitably modifying the proof and the value of α used there) that

$$\begin{aligned} &\sup_{0 \leq r \leq t_\alpha} \left| \frac{V'_r}{\alpha} - \frac{2\rho}{1 + \varepsilon} \right| \xrightarrow{\alpha \rightarrow \infty} 0, \\ &\sup_{0 \leq r \leq t_\alpha} \left| \frac{V''_r}{\alpha} - 2\rho \frac{\alpha + c'\alpha^\gamma}{\alpha} \right| \xrightarrow{\alpha \rightarrow \infty} 0. \end{aligned}$$

Combining the last two limits gives the result since $\varepsilon > 0$ was arbitrary. □

Since the processes M_1, \dots, M_d , which count the marked particles, are in their initial phases close to a supercritical branching process, we need bounds for this kind of processes. In the proof of Theorem 2 we will use the next lemma to control (i) the time until the number of marked particles in the first colony reaches the order α^p , (ii) the time until another colony is infected from the first colony (i.e. the occurrence of the first marked particle on this second colony), and (iii) the time until α^γ particles are marked in the infected colony, when the migration rate $\mu = c\alpha^\gamma$. These three asymptotics correspond to (4.5), (4.6) and (4.7) below. In Lemma 4.4 we will deliberately suppress the effects of back-migration. These effects are controlled in the course of the proof of Theorem 2 by comparison arguments.

Lemma 4.4 (Asymptotic hitting times of a bivariate birth-death process). *Let $c, c', c'' > 0$, $\gamma, p \in (0, 1]$; $\varepsilon_\alpha \downarrow 0, \varepsilon'_\alpha \downarrow 0$ with $\varepsilon_\alpha, \varepsilon'_\alpha > 1/(\log \alpha)$. Let $\mathcal{V} = (V_t)_{t \geq 0}$ be a birth-death process with birth rate $b_k = \alpha k$ and death rate $d_k \leq c\varepsilon_\alpha \alpha k$ for $k \leq \varepsilon_\alpha \alpha$, started in $V_0 = 1$. Moreover, conditional under \mathcal{V} let $\mathcal{W} = (W_t)_{t \geq 0}$ be a birth-death process with time-inhomogeneous birth rate $\mu V_t + \alpha W_t$ and death rate $d_k \leq c'\varepsilon'_\alpha \alpha^\gamma k$ for $k \leq \varepsilon'_\alpha \alpha^\gamma$, starting in $W_0 = 0$. Then we can conclude*

1. For $n \in \mathbb{N}$ let T_n be the first time when $V_t = n$. Then, $\mathbb{P}(T_{\varepsilon_\alpha \alpha^p} = \infty) \leq c\varepsilon_\alpha$ and for all $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p} - p\right| > \varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0. \tag{4.5}$$

2. For $n \in \mathbb{N}$ let S_n be the first time when $W_t = n$. Then, for $\mu = c''\alpha^\gamma$, and any $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} S_1 - (1 - \gamma)\right| > \varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0 \tag{4.6}$$

and

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} S_{\varepsilon'_\alpha \alpha^\gamma} - 1\right| > \varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0. \tag{4.7}$$

Proof. 1. We start with proving (4.5). First, let \mathcal{V}' be a pure branching process with branching rate α (i.e. $b_k' = \alpha k$ and $d_k' = 0$), started with $V'_0 = 1$ and T'_n its hitting time of $V'_t = n$. Then we observe that, as $\alpha \rightarrow \infty$,

$$\mathbb{E}[T'_{\varepsilon_\alpha \alpha^p}] = \sum_{i=1}^{\varepsilon_\alpha \alpha^p - 1} \frac{1}{\alpha i} = \frac{\log \alpha^p}{\alpha} + \mathcal{O}\left(\frac{\log(\varepsilon_\alpha)}{\alpha}\right), \quad \mathbb{V}[T'_{\varepsilon_\alpha \alpha^p}] = \sum_{i=1}^{\varepsilon_\alpha \alpha^p - 1} \frac{1}{\alpha^2 i^2} = \mathcal{O}\left(\frac{1}{\alpha^2}\right). \tag{4.8}$$

Hence by Chebyshev's inequality

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} T'_{\varepsilon_\alpha \alpha^p} - p\right| > \varepsilon\right) \leq \frac{\alpha^2 \mathbb{V}[T'_{\varepsilon_\alpha \alpha^p}]}{(\log \alpha)^2 \varepsilon^2} \xrightarrow{\alpha \rightarrow \infty} 0.$$

Since $T'_n \leq T_n$ stochastically for all n , this implies

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p} - p < -\varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0.$$

For the second bound in (4.5) we consider a process $\mathcal{V}'' = (V''_t)_{t \geq 0}$ with $b_k'' = \alpha k$ and $d_k'' = c\varepsilon_\alpha \alpha k$ with $V''_0 = 1$, and its hitting time T''_n of n . Within the branching process \mathcal{V}'' we consider the immortal lines, i.e. the process of those particles which have descendants at any later time. By classical theory [2, Chapter I.5], the probability that a single line will not be immortal equals the solution of $\frac{c\varepsilon_\alpha \alpha}{\alpha(1+c\varepsilon_\alpha)} + \frac{\alpha}{\alpha(1+c\varepsilon_\alpha)} x^2 = x$, which is smaller than 1, and hence equals $c\varepsilon_\alpha$. So, $\mathbb{P}(T_{\varepsilon_\alpha \alpha^p} < \infty) \geq 1 - c\varepsilon_\alpha$ follows and assuming $T_{\varepsilon_\alpha \alpha^p} < \infty$ we can restrict ourselves in the sequel to the event that the (single) initial particle of \mathcal{V}'' is immortal. Moreover, when an immortal particle splits in \mathcal{V}'' , the new particle has the chance $1 - c\varepsilon_\alpha$ to be immortal. So, every splitting event leads to a new immortal particle with probability $1 - c\varepsilon_\alpha$, so \mathcal{V}'' (given it starts with a single immortal particle) is bounded from below by a binary pure branching process \mathcal{V}''' with individual branching rate $\alpha(1 - c\varepsilon_\alpha)$. For $n \in \mathbb{N}$, let T'''_n be the time it takes \mathcal{V}''' to reach n . Then $T'''_n \geq T''_n$ stochastically for all n , on the event that \mathcal{V}'' starts with an immortal particle at

time 0. On the other hand it is clear that, for all $n \in \mathbb{N}$, $T_n'' \geq T_n$ stochastically. Hence we obtain by the same calculations as in (4.8), now applied to the process \mathcal{V}''' , the estimate

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p} - p > \varepsilon\right) &\leq \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p}''' - p > \varepsilon\right) \\ &= \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha(1 - \frac{c\varepsilon_\alpha}{2})}{\log(\alpha(1 - \frac{c\varepsilon_\alpha}{2}))} T_{\varepsilon_\alpha \alpha^p}''' - p > \varepsilon\right) \\ &\leq \lim_{\alpha \rightarrow \infty} \frac{\alpha^2 \mathbb{V}[T_{\varepsilon_\alpha \alpha^p}''']}{\log(\alpha)^2 \varepsilon^2} = 0. \end{aligned}$$

This completes the proof of (4.5).

2. For the proof of (4.6), we again use comparison arguments based on the processes \mathcal{V}' and \mathcal{V}''' defined in the first part of the proof. Having in mind that $V_t''' \leq V_t \leq V_t'$ stochastically as long as $V_t' \leq \varepsilon_\alpha \alpha$, we introduce the birth processes $\mathcal{W}' = (W_t')_{t \geq 0}$ and $\mathcal{W}''' = (W_t''')_{t \geq 0}$, whose birth rates, conditional on \mathcal{V}' resp. \mathcal{V}''' are $\mu V_t' + \alpha W_t'$ and $\mu(V_t''' - W_t''') + \alpha W_t'''$, respectively. Also, we assume $W_0' = W_0''' = 0$. Let S_1' and S_1''' be the first jump times of \mathcal{W}' and of \mathcal{W}''' (from 0 to 1). From this construction, it is clear that $S_1' \leq S_1 \leq S_1'''$ stochastically. We claim that, on the event $\{T_{\varepsilon_\alpha \alpha} < \infty\}$, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} S_1' - (1 - \gamma) < -2\varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0 \tag{4.9}$$

as well as

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} S_1''' - (1 - \gamma) > 2\varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0 \tag{4.10}$$

which together imply the assertion (4.6). For (4.9), let L' be the number of particles in \mathcal{V}' at the time when \mathcal{W}' reaches 1 for the first time. Then, L' is geometrically distributed with success parameter $\frac{c'' \alpha^\gamma}{\alpha + c'' \alpha^\gamma} = \frac{c''}{\alpha^{1-\gamma} + c''}$ and thus $\mathbb{P}(L' < \alpha^{1-\gamma-\varepsilon}) \xrightarrow{\alpha \rightarrow \infty} 0$. Recalling that T_n' is the first time when $V_t' = n$, we conclude by

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_1' - (1 - \gamma) < -2\varepsilon\right) &= \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_1' - (1 - \gamma) < -2\varepsilon, L' \geq \alpha^{1-\gamma-\varepsilon}\right) \\ &\leq \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\alpha^{1-\gamma-\varepsilon}}' - (1 - \gamma) < -2\varepsilon\right) = 0, \end{aligned}$$

where the last equality follows by a similar calculation as in 1. For (4.10), let L''' be the number of particles in \mathcal{V}''' at the time when \mathcal{W}''' reaches 1 for the first time. Then, L''' is geometrically distributed with success parameter $\frac{c' \alpha^\gamma}{\alpha(1-c\varepsilon_\alpha) + c' \alpha^\gamma} = \frac{c'}{\alpha^{1-\gamma}(1-c\varepsilon_\alpha) + c'}$ and thus $\mathbb{P}(L''' > \alpha^{1-\gamma+\varepsilon}) \xrightarrow{\alpha \rightarrow \infty} 0$. Similarly as above we observe that

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_1''' - (1 - \gamma) > 2\varepsilon\right) &= \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_1''' - (1 - \gamma) > 2\varepsilon, L''' \leq \alpha^{1-\gamma+\varepsilon}\right) \\ &\leq \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\alpha^{1-\gamma+\varepsilon}}''' - (1 - \gamma) > 2\varepsilon\right) = 0. \end{aligned}$$

This concludes the proof of (4.6).

Let us now turn to the proof of (4.7). Using (4.6) we can work on the event

$$\left\{ \left| \frac{\alpha}{\log \alpha} S_1 - (1 - \gamma) \right| < \varepsilon \right\} \cap \{T_{\varepsilon_\alpha \alpha} < \infty\}.$$

Then the time it takes to have $W_t = \varepsilon'_\alpha \alpha^\gamma$ is stochastically smaller than the waiting time until one particle starting at time $(1 - \gamma + 2\varepsilon) \frac{\log \alpha}{\alpha}$ has $\varepsilon'_\alpha \alpha^\gamma$ offspring if we take the birth rate to be αk and the death rate to be $c' \varepsilon'_\alpha \alpha^\gamma k$. This time, in turn, is smaller than the

time until the number of immortal lines \mathcal{U} in the latter process reaches $c'\varepsilon'_\alpha\alpha^\gamma$. (In fact, \mathcal{U} is a pure branching process with individual branching rate $(1 - c'\varepsilon'_\alpha\alpha^{\gamma-1})\alpha$.) Hence, by the same calculation as in the proof of part 1., now denoting by T'_n the first time when $U_t = n$

$$\begin{aligned} & \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{\varepsilon'_\alpha\alpha^\gamma} - 1 > 3\varepsilon\right) \\ &= \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{\varepsilon'_\alpha\alpha^\gamma} - 1 > 3\varepsilon, \frac{\alpha}{\log \alpha} S_1 < 1 - \gamma + 2\varepsilon\right) \\ &\leq \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} T'_{\varepsilon'_\alpha\alpha} - 1 > \varepsilon \mid \frac{\alpha}{\log \alpha} T'_1 = (1 - \gamma + 2\varepsilon)\right) = 0. \end{aligned}$$

This proves one of the bounds in (4.7). For the other bound we work again with \mathcal{V}' , the pure branching process with individual branching rate α started in $V'_0 = 1$, and note that $\mathbb{E}[V_t] \leq \mathbb{E}[V'_t] = e^{\alpha t}$. Again, conditional on \mathcal{V}' , let \mathcal{W}' be a birth-death process with time-inhomogeneous birth rate $\mu V'_t + \alpha W'_t$ and death rate 0, now starting at time $s = (1 - \gamma - 2\varepsilon)\frac{\log \alpha}{\alpha}$ with $W'_s = 1$, and recall $\mathbb{E}[V'_s] = e^{\alpha s} = \alpha^{1-\gamma-2\varepsilon}$. Then, the time it takes to have $W_t = \varepsilon'_\alpha\alpha^\gamma$ is stochastically larger than the hitting time of $\varepsilon'_\alpha\alpha^\gamma$ of the process \mathcal{W}' . We have that $\frac{d}{dt}\mathbb{E}[W'_t] = \mu\mathbb{E}[V'_t] + \alpha\mathbb{E}[W'_t]$, $t \geq s$, $W'_s = 1$, which is solved by

$$\mathbb{E}[W'_t] = \frac{e^{\alpha t}}{\alpha} (\alpha^{\gamma+2\varepsilon} + \alpha\mu t - \mu(1 - \gamma - 2\varepsilon) \log \alpha), \quad t \geq s.$$

Therefore, with $\mu = c''\alpha^\gamma$ and $t = (1 - 3\varepsilon)\frac{\log \alpha}{\alpha}$, using Markov's inequality,

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{\varepsilon'_\alpha\alpha^\gamma} < 1 - 3\varepsilon\right) &\leq \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(W'_{(1-3\varepsilon)\frac{\log \alpha}{\alpha}} > \varepsilon'_\alpha\alpha^\gamma\right) \\ &\leq \lim_{\alpha \rightarrow \infty} \frac{\alpha^{1-3\varepsilon}}{\varepsilon'_\alpha\alpha^{1+\gamma}} (\alpha^{\gamma+2\varepsilon} + c''\alpha^\gamma(\gamma - \varepsilon) \log \alpha) = 0, \end{aligned}$$

which completes the proof of (4.7). □

The following is a direct consequence of Lemma 4.4 in the case of d colonies.

Corollary 4.5. *Assume the birth-death process \mathcal{V} with the same rates as in Lemma 4.4 starts in $V_0 = k$ for $k \in \mathbb{N}$, and consider not a single birth-death process \mathcal{W} , but ℓ birth-death processes $\mathcal{W}^1, \dots, \mathcal{W}^\ell$, which, conditional under \mathcal{V} , have birth rate $\mu V_t + \alpha W_t^i$ for $i \in \{0, \dots, \ell\}$ and death rate $d_k \leq c\varepsilon'_\alpha\alpha^\gamma k$ for $k \leq \varepsilon'_\alpha\alpha^\gamma$ (again with the notation and assumptions from Lemma 4.4). Let $n \in \mathbb{N}$ and S_n^i be the first time when $W_t^i = n$. Then, for $\mu = c''\alpha^\gamma$ and any $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} S_1^i - (1 - \gamma)\right| > \varepsilon, i \in \{0, \dots, \ell\}\right) \xrightarrow{\alpha \rightarrow \infty} 0$$

and

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} S_{\varepsilon'_\alpha\alpha^\gamma}^i - 1\right| > \varepsilon, i \in \{0, \dots, \ell\}\right) \xrightarrow{\alpha \rightarrow \infty} 0.$$

We now complement Lemma 4.4 to cover also the case in which the process \mathcal{V} starts in $c'\alpha^\gamma$ for some $c' > 1$ instead of 1. This lemma will be used later to control the time until of order α particles are marked when one starts with $c'\alpha^\gamma$ marked particles.

Lemma 4.6 (Exponential growth of a near-exponential process). *Let $0 \leq \gamma < p \leq 1$, $c', c'' > 0$ and $c_\alpha, \varepsilon_\alpha \downarrow 0$ be sequences with $\varepsilon_\alpha > 1/(\log \alpha)$ and $\log c_\alpha \in o(\log \alpha)$. Let $\mathcal{V} = (V_t)_{t \geq 0}$ be a birth-death process with birth rate b_k with $\alpha k \leq b_k \leq \alpha k + c'\alpha^{1+\gamma}$ and death rate $d_k \leq c''\varepsilon_\alpha\alpha k$ for $k \leq \varepsilon_\alpha\alpha$, started in $V_0 = c_\alpha\alpha^\gamma$. Let T_n be the first time when $V_t = n$.*

Then, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p} - (p - \gamma)\right| > \varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0. \tag{4.11}$$

Proof. We need to take two bounds for the process \mathcal{V} . Let $\mathcal{V}' = (V'_t)_{t \geq 0}$ be a birth-death process with birth rate $b'_k = \alpha k + c' \alpha^{1+\gamma}$, death rate $d_k = 0$ and $V'_0 = c_\alpha \alpha^\gamma$. If $T'_{\varepsilon_\alpha \alpha^p}$ is the first time when $V'_t = \varepsilon_\alpha \alpha^p$, it is clear that $T'_{\varepsilon_\alpha \alpha^p} \leq T_{\varepsilon_\alpha \alpha^p}$ stochastically.

We define $\mathcal{W}' = (W'_t)_{t \geq 0}$ with $W'_t := \frac{\log V'_t \log(\alpha)/\alpha}{\log \alpha}$, i.e. $V'_t \log(\alpha)/\alpha = \alpha^{W'_t}$ and $W'_0 = \gamma + \frac{\log c_\alpha}{\log \alpha} \in \gamma + o(1)$. Note that $\frac{\alpha}{\log \alpha} T'_{\varepsilon_\alpha \alpha^p}$ is the time when \mathcal{W}' hits $p + \frac{\log \varepsilon_\alpha}{\log \alpha} \in p + o(1)$. Let G' be the generator of \mathcal{W}' . Then, for $x > \gamma$

$$G' f(x) = (\log \alpha)(\alpha^x + c\alpha^\gamma) \underbrace{\left(f\left(\frac{1}{\log \alpha} \log(\alpha^x + 1)\right) - f(x)\right)}_{\approx x + \frac{1}{\log \alpha} \alpha^{-x}} \xrightarrow{\alpha \rightarrow \infty} f'(x).$$

Consequently, and since W'_t quickly leaves its initial state $W'_0 \in \gamma + o(1)$, by Theorem 4.2.11 in [7] the process \mathcal{W}' converges as $\alpha \rightarrow \infty$ on the subsets $E_\alpha := \{\frac{\log k}{\log \alpha} : \log k \in \mathbb{N}, k \geq \gamma \log \alpha + \log c_\alpha\}$ to the (right continuous) process with semigroup $T(t)f(x) = f(x+t)$ for $x \geq \gamma$, growing linearly and deterministically at speed 1. Since $W'_0 \in \gamma + o(1)$, it hits $p + o(1)$ asymptotically as $\alpha \rightarrow \infty$ at time $p - \gamma$ and

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p} - (p - \gamma) < -\varepsilon\right) \leq \mathbb{P}\left(\frac{\alpha}{\log \alpha} T'_{\varepsilon_\alpha \alpha^p} - (p - \gamma) < -\varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0.$$

On the other hand, consider the process $\mathcal{V}'' = (V''_t)_{t \geq 0}$ with birth rate $b''_k = \alpha k$, death rate $d_k = c'' \varepsilon_\alpha \alpha k$ and $V''_0 = c_\alpha \alpha^\gamma$, as well as the time $T''_{\varepsilon_\alpha \alpha^p}$ when this process hits $\varepsilon_\alpha \alpha^p$. Again, consider $\mathcal{W}'' = (W''_t)_{t \geq 0}$ with $W''_t := \frac{\log V''_t \log(\alpha)/\alpha}{\log \alpha}$ and note that $\frac{\alpha}{\log \alpha} T''_{\varepsilon_\alpha \alpha^p}$ is the time when \mathcal{W}'' hits $p + \frac{\log \varepsilon_\alpha}{\log \alpha} \in p + o(1)$. Then, as above, if G'' is the generator of \mathcal{W}'' , for smooth f ,

$$G'' f(x) = (\log \alpha) \alpha^x \left(f\left(\frac{1}{\log \alpha} \log(\alpha^x + 1)\right) - f(x)\right) + c'' (\log \alpha) \varepsilon_\alpha \alpha^x \left(f\left(\frac{1}{\log \alpha} \log(\alpha^x - 1)\right) - f(x)\right) \xrightarrow{\alpha \rightarrow \infty} f'(x)$$

and, since $W''_0 = \gamma + \frac{\log c_\alpha}{\log \alpha} \in \gamma + o(1)$, the process \mathcal{W}'' hits $p + \frac{\log \varepsilon_\alpha}{\log \alpha} \in p + o(1)$ asymptotically at time $p - \gamma$ and

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{\varepsilon_\alpha \alpha^p} - (p - \gamma) > \varepsilon\right) \leq \mathbb{P}\left(\frac{\alpha}{\log \alpha} T''_{\varepsilon_\alpha \alpha^p} - (p - \gamma) > \varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0. \quad \square$$

While the last two lemmata were about supercritical branching processes, we also need the following result about the extinction time of a process which is close to a subcritical branching process.

Lemma 4.7 (Extinction time of a birth-death process). *Let $c > 0$ and $\varepsilon_\alpha \downarrow 0$. Let $\mathcal{V} = (V_t)_{t \geq 0}$ be a birth-death process with birth rate $b_k = \alpha k$ and death rate d_k such that $\alpha(2 - \varepsilon_\alpha)k \leq d_k \leq \alpha(2 + \varepsilon_\alpha)k$, started in $V_0 = z_\alpha \alpha$ with $z_\alpha \rightarrow z$ for some $z > 0$. Let $T_{z_\alpha \alpha}$ be the extinction time of \mathcal{V} , i.e. the first time when $V_t = 0$.*

Then, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{\alpha}{\log \alpha} T_{z_\alpha \alpha} - 1\right| > \varepsilon\right) \xrightarrow{\alpha \rightarrow \infty} 0.$$

Proof. As a first step, consider a sub-critical branching process $\mathcal{W} = (W_t)_{t \geq 0}$ with birth rate α and death rate $\alpha(1 + x_\alpha)$, where $x_\alpha \downarrow x$ with $x > 0$. Let $S_1^{x_\alpha}$ be the extinction time,

when the process is started in a single particle, $W_0 = 1$. Then, from classical theory (see e.g. [10, Chapter V (3.4)]) it follows, that

$$f(t) := \mathbb{P}(S_1^{x_\alpha} > t \mid W_0 = 1) = \frac{x_\alpha}{(1 + x_\alpha)e^{t\alpha x_\alpha} - 1}.$$

Now, consider the same branching process, but started in $W_0 = z_\alpha \alpha$ and denote its extinction time by $S_{z_\alpha \alpha}^{x_\alpha}$. Then, $g(t) := \mathbb{P}(S_{z_\alpha \alpha}^{x_\alpha} > t)$ satisfies

$$g(t) = 1 - (1 - f(t))^{z_\alpha \alpha}.$$

Hence, for $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{z_\alpha \alpha}^{x_\alpha} - \frac{1}{x_\alpha} > \varepsilon\right) &= g\left(\frac{\log \alpha}{\alpha} \left(\frac{1}{x_\alpha} + \varepsilon\right)\right) = 1 - \left(1 - \frac{x_\alpha}{(1 + x_\alpha)\alpha^{1+x_\alpha\varepsilon} - 1}\right)^{z_\alpha \alpha} \\ &\xrightarrow{\alpha \rightarrow \infty} 0, \\ \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{z_\alpha \alpha}^{x_\alpha} - \frac{1}{x_\alpha} < -\varepsilon\right) &= 1 - g\left(\frac{\log \alpha}{\alpha} \left(\frac{1}{x_\alpha} - \varepsilon\right)\right) = 1 - \left(1 - \frac{x_\alpha}{(1 + x_\alpha)\alpha^{1-x_\alpha\varepsilon}}\right)^{z_\alpha \alpha} \\ &\xrightarrow{\alpha \rightarrow \infty} 0. \end{aligned} \tag{4.12}$$

Stochastically, $S_{z_\alpha \alpha}^{1+\varepsilon_\alpha} \leq T_{z_\alpha \alpha} \leq S_{z_\alpha \alpha}^{1-\varepsilon_\alpha}$ and hence,

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{z_\alpha \alpha} - 1 < -2\varepsilon\right) \leq \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{z_\alpha \alpha}^{1+\varepsilon_\alpha} - \frac{1}{1 + \varepsilon_\alpha} < -2\varepsilon + \frac{\varepsilon_\alpha}{1 + \varepsilon_\alpha}\right) \xrightarrow{\alpha \rightarrow \infty} 0$$

as well as, by (4.12),

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{z_\alpha \alpha} - 1 > 2\varepsilon\right) \leq \mathbb{P}\left(\frac{\alpha}{\log \alpha} S_{z_\alpha \alpha}^{1-\varepsilon_\alpha} - \frac{1}{1 - \varepsilon_\alpha} > 2\varepsilon - \frac{\varepsilon_\alpha}{1 - \varepsilon_\alpha}\right) \xrightarrow{\alpha \rightarrow \infty} 0,$$

and we are done. □

While Lemma 4.4 dealt with the initial phase in which allele \mathcal{B} is established in a colony, and Lemmata 4.6 and 4.7 are good for the final phase of fixation, the following lemma links up these two phases.

Lemma 4.8 (Fast middle phase of local sweep). *Let $\mathcal{V} = (V_t)_{t \geq 0}$ be a birth-death process with birth rate $b_k \geq \alpha k$ and death rate $d_k \leq \frac{1}{\rho} \binom{k}{2} + c\alpha^\gamma k$ for some $\gamma \in (0, 1)$ and $c \geq 0, \rho > 0$. Moreover, let T_n be the first time when $V_t = n$. Then there exists a sequence $\varepsilon_\alpha \downarrow 0$ with $\varepsilon_\alpha > 1/(\log \alpha)$ such that for all $\varepsilon'_\alpha \downarrow 0$ with $\varepsilon'_\alpha \geq \varepsilon_\alpha$ and for all $\varepsilon > 0$*

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{(1-\varepsilon'_\alpha)2\alpha\rho} > \varepsilon \mid V_0 = \varepsilon'_\alpha \alpha\right) \xrightarrow{\alpha \rightarrow \infty} 0. \tag{4.13}$$

Proof. We only need to consider the case $b_k = \alpha k$ and $d_k = \frac{1}{\rho} \binom{k}{2} + c\alpha^\gamma k$, since $T_{(1-\varepsilon'_\alpha)2\alpha\rho}$ is maximal in this case. It suffices to show that for all $\delta > 0$ small enough and for all $\varepsilon > 0$

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_{(1-\delta)2\alpha\rho} > \varepsilon \mid V_0 = \delta\alpha\right) \xrightarrow{\alpha \rightarrow \infty} 0. \tag{4.14}$$

We consider the generator of the process $(V_{t/\alpha}/\alpha)_{t \geq 0}$, which is given by

$$\begin{aligned} G_\alpha f(y) &= \alpha y(f(y + \frac{1}{\alpha}) - f(y)) + \left(\frac{1}{\rho} \frac{\alpha y(y - \frac{1}{\alpha})}{2} + c\alpha^\gamma y\right)(f(y - \frac{1}{\alpha}) - f(y)) \\ &\xrightarrow{\alpha \rightarrow \infty} y\left(1 - \frac{y}{2\rho}\right)f'(y). \end{aligned}$$

Using standard arguments, $(V_{t/\alpha}/\alpha)_{t \geq 0}$ converges weakly as $\alpha \rightarrow \infty$ to the solution of the ODE $y' = y(1 - y/(2\rho))$, and if $V_0 = \delta\alpha$, the limiting process starts in $y(0) = \delta$. This solution converges to 2ρ as $t \rightarrow \infty$ (from below since $\delta < 2\rho(1 - \delta)$) and its hitting time of $2\rho(1 - \delta)$ is finite. Consequently, $T_{(1-\delta)2\alpha\rho} \in O(1/\alpha)$ with high probability as $\alpha \rightarrow \infty$, and (4.14) follows. \square

4.3 Proof of Theorem 2

We are now in the position to prove our main result, Theorem 2. The proof will be structured in three main parts, corresponding to the three cases $\mu \in \Theta(\alpha)$, $\mu \in \Theta(\alpha^\gamma)$, and $\mu = \frac{1}{\log \alpha}$ in Theorem 2. Parts 2 and 3 will each be divided into subparts A and B, where A deals with the special case $d = 2$ and B with the general case $d \geq 2$. We feel that this is instructive, because most of the ideas and tools prepared in Sections 3 and 4.2 come into play already in the case $d = 2$. We will give the arguments in parts 1, 2.A and 3.A in detail, whereas we restrict to an outline of the main ideas in parts 2.B and 3.B. Parts 2.A and 2.B will additionally be structured into the cases (i) $\gamma \in (0, 1)$ and (ii) $\gamma = 0$.

The proof of all cases is based on an application of Proposition 3.1. In view of this result, it suffices to check that the fixation time T defined in (3.3) satisfies the properties claimed for T_{fix} in Theorem 2. In the sequel, T_k or T_k^V will always denote the hitting time of k (or of $\lfloor k \rfloor$ if k is not an integer) of a birth-death process V .

Convention. We will use the term *with high probability* or *whp* as a synonym for *with probability 1* as $\alpha \rightarrow \infty$.

Note that in cases 1 and 2 of Theorem 2 the right hand sides are deterministic, so that we have to show that for all $\varepsilon > 0$

$$\mathbb{P} \left(\left| \frac{\alpha}{\log(\alpha)} T - 2 - S_{T^v, \gamma} \right| > \varepsilon \right) \xrightarrow{\alpha \rightarrow \infty} 0.$$

As a prelude, we state two results which hold in all cases. Recall from Proposition 3.1 that the process $(\underline{L}, \underline{M})$ starts in $(\underline{\Pi} + \underline{e}_1, \underline{e}_1)$.

(a) Note that $\frac{L_0}{\alpha} \Rightarrow 2\rho$. Hence, by Lemma 4.1, for some large c , there exists a sequence $\varepsilon_\alpha \downarrow 0$ with

$$\mathbb{P} \left(L_r^i \in [2\alpha\rho_i(1 - \varepsilon_\alpha^2), 2\alpha\rho_i(1 + \varepsilon_\alpha^2)] \text{ for all } i, \text{ for all } 0 \leq r \leq cd \frac{\log \alpha}{\alpha} \right) \xrightarrow{\alpha \rightarrow \infty} 1. \quad (4.15)$$

(b) Let $\varepsilon_\alpha > 0$ be as in (a). For some $\tau > 0$, and $\tau_\alpha = \tau \frac{\log \alpha}{\alpha}$, consider the event

$$\begin{aligned} E_{\tau, i_0} &:= \{L_{\tau_\alpha}^i \in [2\alpha\rho_i(1 - \varepsilon_\alpha^2), 2\alpha\rho_i(1 + \varepsilon_\alpha^2)], M_{\tau_\alpha}^i \\ &\in [2\alpha\rho_i(1 - 4\varepsilon_\alpha^2), 2\alpha\rho_i(1 + \varepsilon_\alpha^2)], i = 1, \dots, d, \\ &M_{\tau_\alpha}^{i_0} = 2\alpha\rho_{i_0}(1 - 4\varepsilon_\alpha^2) \text{ for some } i_0\}. \end{aligned}$$

Now, consider $L^1 + \dots + L^d - M^1 - \dots - M^d$, which is a birth-death process with birth rate $b_n = \alpha n$ if $L^i = \ell_i, M^i = m_i, i = 1, \dots, d$ and $\ell_1 + \dots + \ell_d - m_1 - \dots - m_d = n$ and death rate

$$d_{\underline{\ell}, \underline{m}} := \sum_{i=1}^d \frac{1}{\rho_i} \left(\binom{\ell_i - m_i}{2} + m_i(\ell_i - m_i) \right) = \sum_{i=1}^d \frac{\ell_i + m_i - 1}{2} \frac{\ell_i - m_i}{\rho_i}.$$

(Note that the birth and death rates are independent of μ .) By Lemma 4.1, for all $i = 1, \dots, d$, and since the dynamics of \underline{M} and of \underline{L} coincide, whp on the event E_{τ, i_0} , M^i stays in $[2\alpha\rho_i(1 - 2\varepsilon_\alpha), 2\alpha\rho_i(1 + 2\varepsilon_\alpha)]$ between the times τ_α and $\tau_\alpha + cd \frac{\log \alpha}{\alpha}$. Moreover, L^1, \dots, L^d are bounded as stated in (4.15). Hence, we find the bounds

$$2\alpha(1 - \mathcal{O}(\varepsilon_\alpha))n \leq d_{\underline{\ell}, \underline{m}} \leq 2\alpha(1 + \mathcal{O}(\varepsilon_\alpha))n.$$

By stopping at time $\tau_\alpha + cd \frac{\log \alpha}{\alpha}$ with $cd \geq 1$, we can apply Lemma 4.7 to conclude that, whp,

$$L^1 + \dots + L^d - M^1 - \dots - M^d \text{ hits } 0 \text{ at time in} \tag{4.16}$$

$$\left[\tau_\alpha + \frac{\log \alpha}{\alpha} (1 - \mathcal{O}(\varepsilon_\alpha)), \tau_\alpha + \frac{\log \alpha}{\alpha} (1 + \mathcal{O}(\varepsilon_\alpha)) \right].$$

1. Case $\mu \in \Theta(\alpha)$ Set $\mu = \alpha$ for simplicity. If $(M^i)_{i=1,\dots,d} = \underline{k} = (k_i)_{i=1,\dots,d}$ and $\ell := k_1 + \dots + k_d$, the process $M^1 + \dots + M^d$ has birth rate $b_\ell = \alpha\ell$ and death rate $\frac{1}{\rho_1} \binom{k_1}{2} + \dots + \frac{1}{\rho_d} \binom{k_d}{2} \leq \frac{1}{\min \rho_i} \binom{\ell}{2} =: d_\ell$. For any $\varepsilon_\alpha \downarrow 0$ with $\varepsilon_\alpha > 1/(\log \alpha)$, we can choose $c > 0$ such that $d_\ell \leq c\varepsilon_\alpha \alpha \ell$ for $\ell \leq \varepsilon_\alpha \alpha$. Then, Lemma 4.4, Assertion 1, (used for $M^1 + \dots + M^d$ in place of \mathcal{V} , and with $p = 1$) shows that $M^1 + \dots + M^d$ hits $\varepsilon_\alpha \alpha$ at some time $T_{\varepsilon_\alpha \alpha}$ in $\frac{\log(\alpha)}{\alpha} (1 + o(1))$ whp. Arguing as in the proof of Lemma 4.8, we see that for any $\varepsilon > 0$ small enough any potential limit $\underline{x} = (x_i)_{i=1,\dots,d}$ of the processes $\frac{1}{\alpha} (M_{T_{\varepsilon_\alpha \alpha} + t/\alpha}^1, \dots, M_{T_{\varepsilon_\alpha \alpha} + t/\alpha}^d)$ as $\alpha \rightarrow \infty$ solves for $t > 0$ the system of ODEs

$$\dot{x}_i = x_i - \frac{1}{2\rho_i} x_i^2 + \sum_{j=1}^d a(j, i) x_j - a(i, j) x_i,$$

starting at $t = 0$ in some state with $x_1 + \dots + x_d = \varepsilon$. These ODEs have equilibrium 2ρ and a state \underline{x} with $x_i \in [2\rho_i(1 - \varepsilon), 2\rho_i(1 + \varepsilon)]$, $i = 1, \dots, d$ and $x_{i_0} = 2\rho_{i_0}(1 - \varepsilon)$ for some i_0 is reached after time of order $o(\log(\alpha))$. Now we can – as in the proof of Lemma 4.8 – pass to a sequence ε_α , such that the conditions from above are fulfilled and so that at some time $t \in \frac{\log \alpha}{\alpha} (1 + o(1))$ a state (M_t^1, \dots, M_t^d) with $M_t^i \in [2\alpha\rho_i(1 - \varepsilon_\alpha), 2\alpha\rho_i(1 + \varepsilon_\alpha)]$, $i = 1, \dots, d$ and $M_t^{i_0} = 2\alpha\rho_{i_0}(1 - \varepsilon_\alpha)$ for some i_0 is reached whp. In summary, fixation in the sense of (3.3) occurs at time $t \in \frac{\log \alpha}{\alpha} (2 + o(1))$.

2. A.(i) Case $\mu \in \Theta(\alpha^\gamma)$ for $\gamma \in (0, 1)$, $d = 2$: In the first steps we will apply Lemma 4.4 a couple of times, with suitable choices of the process \mathcal{V} and \mathcal{W} in order to control the “initial phase” of the pair of processes (M^1, M^2) . Note that when (M^1, M^2) is in state (k, ℓ) , then the process M^1 has birth rate $b_k^1 = \alpha k + \mu a(2, 1)\ell$ and death rate $d_k^1 = \frac{1}{\rho_1} \binom{k}{2} + \mu a(1, 2)k$, whereas the process M^2 has birth rate $\alpha\ell + \mu a(1, 2)k$ and death rate $d_\ell^2 = \frac{1}{\rho_2} \binom{\ell}{2} + \mu a(2, 1)\ell$. Moreover, $M^1 + M^2$ is a birth-death process with birth rate $\alpha(k + \ell)$ and death rate $\frac{1}{\rho_1} \binom{k}{2} + \frac{1}{\rho_2} \binom{\ell}{2}$. Let $\varepsilon_\alpha, \varepsilon'_\alpha \downarrow 0$ be sequences with $\varepsilon_\alpha, \varepsilon'_\alpha > 1/(\log \alpha)$.

First, we are going to establish that M^1 hits $\varepsilon_\alpha \alpha^p$ by time

$$T_{\varepsilon_\alpha \alpha^p}^{M^1} \in \frac{\log \alpha}{\alpha} (p + o(1)) \text{ whp.}$$

On the one hand, this hitting time $T_{\varepsilon_\alpha \alpha^p}^{M^1}$ is stochastically larger than $T_{\varepsilon_\alpha \alpha^p}^{M^1 + M^2}$. For the latter, Assertion 1 of Lemma 4.4 (applied to with $V_t = M_t^1 + M_t^2$) ensures that $T_{\varepsilon_\alpha \alpha^p}^{M^1 + M^2} \in \frac{\log \alpha}{\alpha} (p + o(1))$ whp. On the other hand, $T_{\varepsilon_\alpha \alpha^p}^{M^1}$ is smaller than the hitting time of $\varepsilon_\alpha \alpha^p$ when only non-(im)migrated lines in M^1 are counted. This process of non-immigrated lines is a birth-death process \tilde{M}^1 with birth rate αk and death rate $\frac{1}{\rho_1} \binom{k}{2} + \mu a(1, 2)k$, and therefore fulfills the conditions of the process \mathcal{V} of Lemma 4.4 (with ε_α as above). Consequently, also $T_{\varepsilon_\alpha \alpha^p}^{\tilde{M}^1} \in \frac{\log \alpha}{\alpha} (p + o(1))$ whp. Taking these two comparisons together, we find that $T_{\varepsilon_\alpha \alpha^p}^{M^1} \in \frac{\log \alpha}{\alpha} (p + o(1))$ whp as well.

Second, we will show that the process M^2 hits 1 by time

$$T_1^{M^2} \in \frac{\log \alpha}{\alpha} (1 - \gamma + o(1)) \text{ whp.} \tag{4.17}$$

This hitting time is actually the same if we change the birth rate of M^1 (from $\alpha k + \mu\ell$) to αk , since $M^2 = 0$ before $T_1^{M^2}$. Hence, up to time $T_1^{M^2}$, the process (M^1, M^2) in place of $(\mathcal{V}, \mathcal{W})$ fulfills the conditions of Lemma 4.4, with the μ appearing there replaced by $\mu a(1, 2)$. This lemma can now be directly applied to obtain (4.17).

Third, we will argue that there exists a sequence $\varepsilon'_\alpha \downarrow 0$ with $\varepsilon'_\alpha > 1/(\log \alpha)$ such that M^2 hits $\varepsilon'_\alpha \alpha^\gamma$ by time

$$T_{\varepsilon'_\alpha \alpha^\gamma}^{M^2} \in \frac{\log \alpha}{\alpha}(1 + o(1)) \text{ whp.} \tag{4.18}$$

On the one hand, this hitting time is stochastically larger than the hitting time if migration from colony 2 to colony 1 is suppressed. For the thus modified process (\hat{M}^1, \hat{M}^2) , \hat{M}^1 has birth rate αk and therefore (\hat{M}^1, \hat{M}^2) fulfills the requirements of Lemma 4.4 (for the same combination of $\varepsilon_\alpha, \varepsilon'_\alpha$ as described above) and $T_{\varepsilon'_\alpha \alpha^\gamma}^{\hat{M}^2} \in \frac{\log \alpha}{\alpha}(1 + o(1))$ whp. On the other hand, this hitting time is stochastically smaller than the hitting time of $\varepsilon'_\alpha \alpha^\gamma$ if only a single migration event from colony 1 to colony 2 happens, i.e. the hitting time $T_{\varepsilon'_\alpha \alpha^\gamma}^V$ of a process V which starts at time $\frac{\log \alpha}{\alpha}(1 - \gamma + o(1))$ with $V = 1$, and has birth rate $\alpha\ell$ and death rate $\frac{1}{\rho_2} \binom{\ell}{2} + \mu a(2, 1)\ell$. By Lemma 4.4, Assertion 1, this time is $T_{\varepsilon'_\alpha \alpha^\gamma}^V \in \frac{\log \alpha}{\alpha}(1 - \gamma + o(1)) + \frac{\log \alpha}{\alpha}(\gamma + o(1)) = \frac{\log \alpha}{\alpha}(1 + o(1))$ whp and (4.18) follows. Moreover, we have shown that the pair (M^1, M^2) inherits the properties (4.5), (4.6), (4.7) proved in Lemma 4.4 for the pair $(\mathcal{V}, \mathcal{W})$.

In order to go further, we next observe that (as a consequence of the statement in the first step of this proof, with $p = 1$) we have that $T_{\varepsilon_\alpha \alpha}^{M^1} \in \frac{\log \alpha}{\alpha}(1 + o(1))$ and $T_{\varepsilon_\alpha \alpha^\gamma}^{M^2} \in \frac{\log \alpha}{\alpha}(1 + o(1))$ whp. By Lemma 4.8 (applied to the process M^1) we find a sequence ε_α decreasing sufficiently slow such that it takes time of at most order $o(\log(\alpha)/\alpha)$ until M^1 hits $2\alpha\rho_1(1 - \varepsilon_\alpha)$. Note, that the sequences $\varepsilon_\alpha, \varepsilon'_\alpha$ were arbitrary and only had to fulfill $\varepsilon_\alpha, \varepsilon'_\alpha > 1/(\log \alpha)$, hence there exist sequences $\varepsilon_\alpha, \varepsilon'_\alpha$ for which all assertions claimed so far are fulfilled. Also in the following we will if necessary replace the sequences by slower converging ones. We note that, due to Corollary 4.3, whp the process M^1 will not drop below $2\alpha\rho_1(1 - 2\varepsilon_\alpha)$ for the entire period remaining to fixation with ε_α again suitably adapted. Now, if $M^2 = \ell$, it has birth rate $\alpha\ell + \mu a(1, 2)M^1$, and since $M^1 \leq L^1 \leq 2\alpha\rho_1(1 + \varepsilon_\alpha)$, this is bounded above by $\alpha\ell + c\alpha^{1+\gamma}$ for some constant c . The death rate of M^2 is (for the same $\varepsilon_\alpha, \varepsilon'_\alpha$ as above) $\frac{1}{\rho_2} \binom{\ell}{2} + \mu a(2, 1)\ell \leq c\varepsilon'_\alpha \alpha\ell/2$ for $\ell \leq \varepsilon'_\alpha \alpha$ for some $c > 0$. In addition, the sequence ε_α fulfills the conditions on the sequence c_α in Lemma 4.6. Hence, Lemma 4.6 implies that M^2 hits $\varepsilon'_\alpha \alpha$ by time $T_{\varepsilon'_\alpha \alpha}^{M^2} \in T_{\varepsilon'_\alpha \alpha^\gamma}^{M^2} + \frac{\log(\alpha)}{\alpha}(1 - \gamma + o(1)) = \frac{\log(\alpha)}{\alpha}(2 - \gamma + o(1))$ when $\varepsilon_\alpha, \varepsilon'_\alpha$ are suitably adapted. Again, M^2 rises to $2\alpha\rho_2(1 - 2\varepsilon_\alpha)$ by some time of order $o(\log(\alpha)/\alpha)$ by Lemma 4.8 (applied to the process M^2), so by some time in $\frac{\log(\alpha)}{\alpha}(2 - \gamma + o(1))$, we find that $M^1 \geq 2\alpha\rho_1(1 - 2\varepsilon_\alpha)$ and $M^2 = 2\alpha\rho_2(1 - 2\varepsilon_\alpha)$. Now, fixation occurs after time in $T = T_{\varepsilon_\alpha \alpha}^{M^2} + \frac{\log(\alpha)}{\alpha}(1 + o(1)) = \frac{\log(\alpha)}{\alpha}(3 - \gamma + o(1))$ by (4.16).

2. A.(ii) Case $\mu \in \Theta(1)$, $d = 2$: Arguing exactly as in Case 2.A.(i), but now with $p = 1$, we obtain for any $\varepsilon_\alpha \downarrow 0$ with $\varepsilon_\alpha > 1/(\log \alpha)$ that M^1 hits $\varepsilon_\alpha \alpha$ by time $T_{\varepsilon_\alpha \alpha}^{M^1} \in \frac{\log \alpha}{\alpha}(1 + o(1))$ whp. In addition, by Lemma 4.8, M^1 has increased to $(1 - \varepsilon_\alpha)2\alpha\rho_1$ (maybe after modifying ε_α) by time $T_{(1 - \varepsilon_\alpha)2\alpha\rho_1}^{M^1} \in \frac{\log \alpha}{\alpha}(1 + o(1))$.

For bounding the time $T_1^{M^2}$ stochastically from below, fix $\varepsilon > 0$ and let \hat{M}^2 be as M^2 but with $\gamma = \varepsilon/2$. Since $T_1^{\hat{M}^2} \leq T_1^{M^2}$, we find that by Lemma 4.4

$$\mathbb{P}\left(\frac{\alpha}{\log \alpha} T_1^{M^2} - 1 < -\varepsilon\right) \leq \mathbb{P}\left(\frac{\alpha}{\log \alpha} T_1^{\hat{M}^2} - (1 - \varepsilon/2) < -\varepsilon/2\right) \xrightarrow{\alpha \rightarrow \infty} 0. \tag{4.19}$$

For bounding $T_1^{M^2}$ from above, consider migrants only after time $T_{(1 - \varepsilon_\alpha)2\alpha\rho_1}^{M^1} \in \frac{\log \alpha}{\alpha}(1 + o(1))$. Due to Corollary 4.3, whp the process M^1 will not drop below $2\alpha\rho_1(1 - 2\varepsilon_\alpha)$ for the

The fixation time of a strongly beneficial allele

t	$M_{t \log \alpha / \alpha}^\iota$	$M_{t \log \alpha / \alpha}^i, i \in D_1$	$M_{t \log \alpha / \alpha}^i, i \in D_2$	$M_{t \log \alpha / \alpha}^i, i \in D_3$	\dots
0	1	0	0	0	\dots
$1 - \gamma$	$\Theta(\alpha^{1-\gamma})$	1	0	0	\dots
$2(1 - \gamma)$	$\Theta(\alpha^{1 \wedge (2(1-\gamma))})$	$\Theta(\alpha^{1-\gamma})$	1	0	\dots
$3(1 - \gamma)$	$\Theta(\alpha^{1 \wedge (3(1-\gamma))})$	$\Theta(\alpha^{1 \wedge (2(1-\gamma))})$	$\Theta(\alpha^{1-\gamma})$	1	\dots

Table 1: The table gives (approximate) times and orders of magnitude in the case 2.B.(i) ($\gamma \in (0, 1)$, $d \geq 2$); see text for more explanation.

entire period remaining to fixation. The expected number of migrants between times $T_{(1-\varepsilon_\alpha)2\alpha\rho_1}^{M^1}$ and $T_{(1-\varepsilon_\alpha)2\alpha\rho_1}^{M^1} + \frac{1}{\log \log \alpha} \frac{\log \alpha}{\alpha}$ is at least $\mu a(1, 2)2\alpha\rho_1(1 - 2\varepsilon_\alpha) \frac{1}{\log \log \alpha} \frac{\log \alpha}{\alpha} \xrightarrow{\alpha \rightarrow \infty} \infty$ and hence we have $M_t^2 \geq 1$ for some $t \in \frac{\log \alpha}{\alpha}(1 + o(1))$ whp. Together with (4.19) this says that $T_1^{M^2} \in \frac{\log \alpha}{\alpha}(1 + o(1))$ whp. We can now apply Lemma 4.6 (with $\gamma := 0$ and $p = 1$) to infer that the process M^2 reaches $\varepsilon_\alpha \alpha$ in $\frac{\log \alpha}{\alpha}(2 + o(1))$ whp. From Lemma 4.8, we hence find some $t \in 2 \frac{\log \alpha}{\alpha}(1 + o(1))$ for which $M_t^1 \geq 2\alpha\rho_1(1 - \varepsilon_\alpha)$, $M_t^2 = 2\alpha\rho_2(1 - 2\varepsilon_\alpha)$. Then by (4.16), fixation occurs at time in $\frac{\log \alpha}{\alpha}(3 + o(1))$.

2. B.(i) Case $\mu \in \Theta(\alpha^\gamma)$ for $\gamma \in (0, 1)$, $d \geq 2$: Set $[d] := \{1, \dots, d\}$ and, for $s = 0, 1, \dots, \Delta_\iota$, let $B_s(\iota)$ be the set of vertices in $[d]$ which can be reached from ι by at most s steps (cf. Definition 2.5). We partition $[d] = \bigcup_{s=0}^{\Delta_\iota} D_s$ into $D_0 := \{\iota\}$ and $D_s := B_s(\iota) \setminus B_{s-1}(\iota)$, $s = 1, \dots, \Delta_\iota$. Arguing similarly as in part 2.A (i), now based on Corollary 4.5, we obtain the analogue of (4.17), simultaneously for all $i \in D_1$. In the language of the epidemic process $\mathcal{I}^{\iota, \gamma}$ this means that all colonies $i \in D_1$ are infected at times

$$T_1^{M^i} \in \frac{\log(\alpha)}{\alpha}(1 + o(1)) \text{ whp};$$

see also Table 1 for orientation.

Let us concentrate now on a colony $m \in D_2$ and set $D_1^{(m)} := \{j \in D_1 : a(j, m) > 0\}$. From the second assertion of Corollary 4.5 we obtain that there exists a sequence $\varepsilon_\alpha \downarrow 0$, such that for all colonies $i \in D_1^{(m)}$,

$$T_{\varepsilon_\alpha \alpha^{1-\gamma}}^{M^i} \in \frac{\log(\alpha)}{\alpha}(2(1 - \gamma) + o(1)) \text{ whp.}$$

Hence, all $i \in D_1^{(m)}$ will infect m by this time whp. Equation (4.7), translated to the pairs (M^ι, M^i) for $i \in D_1$ in a similar way as done in part 2.A.(i) for the pair (M^1, M^2) , implies that migration from the founder colony ι does not speed up (on the $\frac{\log(\alpha)}{\alpha}$ -timescale) the processes M^i till they reach $\varepsilon_\alpha \alpha^\gamma$ for an appropriate sequence ε_α ; in fact, during this period the rate of growth of M^i is that of a branching process with Malthusian parameter α . Lemma 4.6 carries this assertion further: Since $M^\iota \leq 2\alpha\rho_\iota(1 + 2\varepsilon_\alpha)$, migration from colony ι to colonies in D_1 is bounded by $c\alpha^{1+\gamma}$ for an appropriate constant c . In addition, the sequence ε_α fulfills the conditions of the sequence c_α in Lemma 4.6. Consequently, the process M^i continues to grow like a branching process with rate α by Lemma 4.6 and for $m \in D_2$ the assertions of Lemma 4.4 are fulfilled with $\sum_{j \in D_1^{(m)}} M^j$ playing the role of \mathcal{V} and M^m playing the role of \mathcal{W} , see also Table 1. It follows that $T_1^{M^m} \in \frac{\log(\alpha)}{\alpha}(2(1 - \gamma) + o(1))$ whp.

Repeating these arguments one finds that all colonies are, whp, infected by a time in $\frac{\log(\alpha)}{\alpha}(\Delta_\iota(1 - \gamma) + o(1))$, with Δ_ι as in Definition 2.5. Finally, arguing as in part 2.A.(i), it takes an additional time in $\frac{\log(\alpha)}{\alpha}(2 + o(1))$ until fixation occurs. This sums up to a total time in $\frac{\log(\alpha)}{\alpha}((2 + S_{\mathcal{I}^{\iota, \gamma}}) + o(1))$ whp, with $S_{\mathcal{I}^{\iota, \gamma}} = (1 - \gamma)\Delta_\iota$ according to Definition 2.5.

2. B.(ii) Case $\mu \in \Theta(1)$, $d \geq 2$: We will use the same notation as in Case 2.B.(i). Let $\varepsilon_\alpha > 0$. The argument from Case 2.A.(ii) works for all colonies $i \in D_1$ which are distance 1 apart from colony ι . Hence, whp, at some time in $\frac{\log \alpha}{\alpha}(2 + o(1))$, there is $M^i \in [2\alpha\rho_i(1 - 2\varepsilon_\alpha), 2\alpha\rho_i(1 + 2\varepsilon_\alpha)]$ for $i = \iota$ and $i \in D_1$. Similarly, each colony $m \in D_2$ has $M^m = 1$ (and in this sense is *infected*) within an additional time interval of length $o(\frac{\log(\alpha)}{\alpha})$, and then M^m increases to $2\alpha\rho_m(1 - 2\varepsilon_\alpha)$ after a duration in $\frac{\log \alpha}{\alpha}(1 + o(1))$. This procedure is iterated, and all colonies are infected by a time in $\frac{\log(\alpha)}{\alpha}(\Delta_\iota + 1 + o(1))$ whp. Then, from (4.16), fixation occurs at time in $\frac{\log(\alpha)}{\alpha}(\Delta_\iota + 2 + o(1))$ whp, giving the result.

3. A. Case $\mu = \frac{1}{\log \alpha}$, $d = 2$: The main step in this case is to show that

$$\frac{\alpha}{\log \alpha} T_1^{M^2} \xrightarrow{\alpha \rightarrow \infty} 1 + X, \text{ where } X \sim \exp(2\rho_1 a(1, 2)). \quad (4.20)$$

By the same arguments as in Case 2.A.(ii), for any sequence $\varepsilon_\alpha \downarrow 0$ with $\varepsilon_\alpha > 1/(\log \alpha)$, we have $T_{(1-\varepsilon_\alpha)\alpha}^{M^1} = \frac{\log \alpha}{\alpha}(1 + o(1))$ whp. In addition, $M^2 = 0$ before $T_{\varepsilon_\alpha\alpha}^{M^1}$ whp, as we can estimate the number of migrants from colony 1 to colony 2 by $\varepsilon_\alpha\alpha\mu a(1, 2)\frac{\log \alpha}{\alpha}(1 + o(1)) \xrightarrow{\alpha \rightarrow \infty} 0$. Here, the expected number of migrants from colony 1 to colony 2 during $[T_{\varepsilon_\alpha\alpha}^{M^1}, T_{(1-\varepsilon_\alpha)\alpha}^{M^1}]$ is bounded from above by $2\alpha\rho_1(1 - \varepsilon_\alpha)\mu a(1, 2)o(\frac{\log \alpha}{\alpha}) \xrightarrow{\alpha \rightarrow \infty} 0$ since $T_{(1-\varepsilon_\alpha)\alpha}^{M^1} - T_{\varepsilon_\alpha\alpha}^{M^1} = o(\frac{\log \alpha}{\alpha})$ by Lemma 4.8 with a possibly slower decreasing sequence ε_α . Hence, we have $M^2 = 0$ before $T_{(1-\varepsilon_\alpha)\alpha}^{M^1}$ whp as well. By Corollary 4.3, we have that $M_t^1 \in [2\alpha\rho_1(1 - 2\varepsilon_\alpha), 2\alpha\rho_1(1 + 2\varepsilon_\alpha)]$ after $T_{(1-\varepsilon_\alpha)\alpha}^{M^1}$ until fixation. Hence, for all $x > 0$,

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathbb{P}\left(\frac{\alpha}{\log \alpha} T_1^{M^2} - 1 > x\right) &= \lim_{\alpha \rightarrow \infty} \mathbb{E}\left[\exp\left(-\int_{T_{(1-\varepsilon_\alpha)\alpha}^{M^1}}^{\frac{\log \alpha}{\alpha}(1+x)} \mu a(1, 2) M_t^1 dt\right)\right] \\ &= \lim_{\alpha \rightarrow \infty} \exp\left(-\int_{\frac{\log \alpha}{\alpha}}^{\frac{\log \alpha}{\alpha}(1+x)} \frac{2\alpha\rho_1 a(1, 2)}{\log \alpha} dt\right) \\ &= e^{-2\rho_1 a(1, 2)x}, \end{aligned}$$

which gives (4.20). Analogously to the other cases we find $c > 0$ and a sequence $\varepsilon_\alpha \downarrow 0$ with $\varepsilon_\alpha > 1/(\log \alpha)$, such that if $(M^1, M^2) = (k, \ell)$, M^2 is a birth-death process with birth rate $b_\ell = \alpha\ell + \mu a(1, 2)k \leq \alpha\ell + c\alpha/\log \alpha$ and death rate $d_\ell = \frac{1}{\rho_2} \binom{\ell}{2} + \mu a(2, 1)\ell \leq \varepsilon_\alpha\alpha\ell$ for $\ell \leq \varepsilon_\alpha\alpha$. So, we can apply Lemma 4.6 (for $\gamma = 0$) in order to see that $T_{\varepsilon_\alpha\alpha}^{M^2}$ occurs after duration in $\frac{\log \alpha}{\alpha}(1 + o(1))$. Then, using Lemma 4.8, we see that $T_{2\alpha\rho_2(1-2\varepsilon_\alpha)}^{M^2} \in \frac{\log \alpha}{\alpha}(2 + X + o(1))$ for some X distributed as above. Then, using (4.16), fixation occurs at time in $\frac{\log \alpha}{\alpha}(3 + X + o(1))$, as claimed.

3. B. Case $\mu = \frac{1}{\log \alpha}$, $d \geq 2$: By the same arguments as in Case 3.A at a time $t \in \frac{\log \alpha}{\alpha}(1 + o(1))$, colony ι in the process \mathcal{J}^ι from Definition 2.5 switches from being infected to being infectious. From here on, each colony $i \in D_1$ can be *infected* by a migrant from colony ι at rate $2\rho_1 a(\iota, i)\alpha/(\log \alpha)$, i.e. at rate $2\rho_1 a(\iota, i)$ on the $\frac{\log \alpha}{\alpha}$ -timescale. After i is infected, M^i increases until there are of the order α particles, which happens after time of duration $\frac{\log \alpha}{\alpha}(1 + o(1))$. Then, the colony becomes *infectious*, meaning that other colonies can be infected from that colony. More precisely, if colony i is infectious and colony j satisfies $a(i, j) > 0$, then, as long as $M^j = 0$, a migrant from M^i arrives in colony j after an exponential time with rate $2\rho_j a(i, j)$ on the $\frac{\log \alpha}{\alpha}$ -timescale. Continuing in this way, the waiting time until all colonies are infectious is $\frac{\log(\alpha)}{\alpha}(S_{\mathcal{J}^\iota} + o(1))$ in the

approximating process \mathcal{J}^t . At this time, each colony i has $M^i \geq 2\alpha\rho_i(1 - \varepsilon_\alpha)$, $i = 1, \dots, d$. As in the other cases we conclude from (4.16) that after an additional time of duration in $\frac{\log \alpha}{\alpha}(1 + o(1))$, fixation has occurred.

Acknowledgments

We thank Wolfgang Stephan for posing the question initiating this work, Tom Kurtz for help and stimulating discussions related to the proof of Lemma 4.1, Jeff Jensen for valuable comments on the manuscript and two anonymous referees for their careful and critical reading which helped us to improve the presentation and to correct errors from previous versions of the manuscript. This research was supported by the DFG through the research unit 1078 and the priority program 1590, and in particular through grant GR 876/16-1 to AG, Pf-672/3-1 and Pf-672/6-1 to PP, and Wa-967/4-1 to AW.

References

- [1] Aldous, D. (1985). Exchangeability and related topics. In P. Hennequin (Ed.), *École d'Été de Probabilités de Saint-Flour XIII-1983*, Volume 1117 of *Lecture Notes in Mathematics*, Berlin, pp. 1–198. Springer. MR-0883646
- [2] Athreya, K. and P. Ney (1972). *Branching Processes*. Springer. MR-0373040
- [3] Athreya, S. and J. Swart (2005). Branching-coalescing particle systems. *Prob. Theory Relat. Fields* 131, 376–414. MR-2123250
- [4] Dawson, D. (1993). Measure-valued Markov processes. In P. Hennequin (Ed.), *École d'Été de Probabilités de Saint-Flour XXI-1991*, Volume 1541 of *Lecture Notes in Mathematics*, Berlin, pp. 1–260. Springer. MR-1242575
- [5] A. Depperschmidt, A. Greven, and P. Pfaffelhuber. Tree-valued Fleming–Viot dynamics with mutation and selection. *Ann. Appl. Probab.*, 22(6):2560–2615, 2012. MR-3024977
- [6] Etheridge, A., P. Pfaffelhuber, and A. Wakolbinger (2006). An approximate sampling formula under genetic hitchhiking. *Ann. Appl. Probab.* 16, 685–729. MR-2244430
- [7] Ethier, S. and T. Kurtz (1986). *Markov Processes. Characterization and Convergence*. John Wiley, New York. MR-0838085
- [8] Fearnhead, P. (2002). The common ancestor at a non-neutral locus. *J. Appl. Probab.* 39, 38–54. MR-1895142
- [9] Feinberg, M. (1979). *Lectures on chemical reaction networks. Notes of lectures given at the Mathematics Research Centre*. University of Wisconsin.
- [10] Harris, T. (1963). *The Theory of Branching Processes*. Springer. MR-0163361
- [11] Hartfield, M. (2012). A framework for estimation the fixation time of an advantageous allele in stepping-stone models. *J. Evol. Biol.* 25, 1751–1764.
- [12] Kaplan, N. L., R. R. Hudson, and C. H. Langley (1989). The ‘Hitchhiking effect’ revisited. *Genetics* 123, 887–899.
- [13] Kim, Y. and T. Maruki (2011). Hitchhiking effect of a beneficial mutation spreading in a subdivided population. *Genetics* 189, 213–226.
- [14] Krone, S. and C. Neuhauser (1997). Ancestral processes with selection. *Theo. Pop. Biol.* 51, 210–237.
- [15] Mano, S. (2009). Duality, ancestral and diffusion processes in models with selection. *Theo. Pop. Biol.* 75, 164–175.
- [16] Maynard Smith, J. and J. Haigh (1974). The hitch-hiking effect of a favorable gene. *Genetic Research* 23, 23–35.
- [17] Nagylaki, T. (1982). Geographical invariance in population genetics. *J. Theo. Biol.* 99(1), 159–172. MR-0685073

The fixation time of a strongly beneficial allele

- [18] Neuhauser, C. and S. Krone (1997). The genealogy of samples in models with selection. *Genetics* 154, 519–534.
- [19] Nielsen, R. (2005). Molecular Signatures of Natural Selection. *Annu. Rev. Genet.* 39, 197–218.
- [20] Norris, J. R. (1998). *Markov Chains*. Cambridge University Press. MR-1600720
- [21] Pfaffelhuber, P. and C. Pokalyuk (2013). The ancestral selection graph under strong directional selection. *Theo. Pop. Biol.* 87, 25–33.
- [22] Sabeti, P., S. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. Mikkelsen, D. Altshuler, and E. Lander (2006). Positive natural selection in the human lineage. *Science* 312, 1614–1620.
- [23] Schweinsberg, J. and R. Durrett (2005). Random partitions approximating the coalescence of lineages during a selective sweep. *Ann. Appl. Probab.* 15, 1591–1651. MR-2152239
- [24] Shiga, T. and K. Uchiyama (1986). Stationary states and their stability of the stepping stone model involving mutation and selection. *Prob. Theo. Rel. Fields* 73, 87–116. MR-0849066
- [25] Slatkin, M. (1981). Fixation probabilities and fixation times in a subdivided population. *Evolution* 35, 477–488.
- [26] Slatkin, M. (1976). The rate of spread of an advantageous allele in a subdivided population. *Population Genetics and Ecology*, 767–780.
- [27] Stephan, W., T. H. E. Wiehe, and M. W. Lenz (1992). The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. *Theo. Pop. Biol.* 41, 237–254.
- [28] Thornton, K., J. Jensen, C. Becquet, and P. Andolfatto (2007). Progress and prospects in mapping recent selection in the genome. *Heredity* 98, 340–348.
- [29] Wakeley, J. and O. Sargsyan (2009). The conditional ancestral selection graph with strong balancing selection. *Theo. Pop. Biol.* 75, 355–364.
- [30] Whitlock, M. C. (2003). Fixation probability and time in subdivided populations. *Genetics* 164(2), 767–779.