An Eco-Evolutionary approach of Adaptation and Recombination in a large population of varying size

Charline Smadi

AN ECO-EVOLUTIONARY APPROACH OF ADAPTATION AND RECOMBINATION IN A LARGE POPULATION OF VARYING SIZE

CHARLINE SMADI

ABSTRACT. We identify the genetic signature of a selective sweep in a population described by a birth-and-death process with density dependent competition. We study the limit behaviour for large K, where K scales the population size. We focus on two loci: one under selection and one neutral. We distinguish a soft sweep occurring after an environmental change, from a hard sweep occurring after a mutation, and express the neutral proportion variation as a function of the ecological parameters, recombination probability r_K, and K. We show that for a hard sweep, two recombination regimes appear according to the order of $r_K \log K$.

1. INTRODUCTION

There are at least two different ways of adaptation for a population: selection can either act on a new mutation (hard selective sweep), either on preexisting alleles that become advantageous after an environmental change (soft selective sweep). New mutations are sources of diversity, and hard selective sweep was until recently the only considered way of adaptation. Soft selective sweep allows a faster adaptation to novel environments, and its importance is growing in empirical and theoretical studies (Prezeworski, Coop and Wall [27], Barrett and Schluter [2]). These distinct selective sweeps entail different genetic signatures in the vicinity of the novely fixed allele, and the multiplication of genetic data available allows to detect these signatures in current populations as described by Peter, Huerta-Sanchez and Nielsen [25]. To do this in an effective way, it is necessary to identify accurately the signatures left by these two modes of adaptation.

In this work, we consider a sexual haploid population of varying size, modeled by a birth and death process with density dependent competition. The ability to survive and reproduce of each individual depends on its own genotype and on the population state. More precisely, each individual is characterized by some ecological parameters: birth rate, intrinsic death rate and competition kernel describing the competition with other individuals depending on their genotype. The differential reproductive success of individuals generated by their interactions entail progressive variations in the number of individuals carrying a given genotype. This process, called natural selection, is a key mechanism of evolution. Such eco-evolutionary approach has been introduced by Metz and coauthors in [24] and made rigorous in the seminal paper of Fournier and Méléard [18]. Then it has been developed by Champagnat and coauthors [5, 6, 7, 8, 9], and Méléard and Tran [23] for the haploid asexual case and by Collet, Méléard and Metz [10] and Coron [11, 12] for the diploid sexual case. The recent work of Billiard and coauthors [4] studies the dynamics of a two-locus model in an haploid asexual population. Following these works, we introduce a parameter K called carrying capacity which scales the population size, and study the limit behavior for large K. But unlike them, we focus on two loci in a sexual haploid population and take into

UNIVERSITÉ PARIS-EST, CERMICS (ENPC), F-77455 MARNE LA VALLÉE, FRANCE AND CMAP UMR 7641, ÉCOLE POLYTECHNIQUE CNRS, ROUTE DE SACLAY, 91128 PALAISEAU CEDEX, FRANCE

E-mail address: charline.smadi@polytechnique.edu.

Date: March 17, 2014.

2010 *Mathematics Subject Classification.* 60J75; 60F99; 60K35; 92D15.

Key words and phrases. birth and death process; coalescent; eco-evolution; genetic hitch-hiking; selective sweep.

account recombinations: one locus is under selection and has two possible alleles A and a and the second one is neutral with allele b_1 or b_2. When two individuals give birth, either a recombination occurs with probability r_K and the newborn inherits one allele from each parent, or he is the clone of one parent.

We first focus on soft selective sweep occurring after a change in the environment (new pathogen, environmental catastrophe, occupation of a new ecological niche,...). We assume that before the change the alleles A and a were neutral and represented both a positive fraction of the population, and that in the new environment the allele a becomes favorable and goes to fixation. We can divide the selective sweep in two periods: a first one where the population process is well approximated by the solution of a deterministic dynamical system, and a second one where A-individuals are near extinction, the deterministic approximation fails and the fluctuations of the A-population size become predominant. We give the asymptotic value of the final neutral allele proportions as a function of the ecological parameters, recombination probability r_K and solutions of a two-dimensional competitive Lotka-Volterra system.

We then focus on hard selective sweep. We assume that a mutant a appears in a monomorphic A-population at ecological equilibrium. As stated by Champagnat in [5], the selective sweep is divided in three periods: during the first one, the resident population size stays near its equilibrium value, and the mutant population size grows until it reaches a non-negligible fraction of the total population size. The two other periods are the ones described for the soft selective sweep. Moreover, the time needed for the mutant a to fix in the population is of order $\log K$. We prove that the distribution of neutral alleles at the end of the sweep has different shapes according to the order of the recombination probability per reproductive event r_K with respect to $1/\log K$. More precisely, we find two recombination regimes: a strong one were $r_K \log K$ is large, and a weak one were $r_K \log K$ is bounded. In both recombination regimes, we give the asymptotic value of the final neutral allele proportions as a function of the ecological parameters and recombination probability r_K. In the strong recombination regime, the frequent exchanges of neutral alleles between the A and a-populations yield an homogeneous neutral repartition in the two populations and the latter is not modified by the sweep. In the weak recombination regime, the frequency of the neutral allele carried by the first mutant increases because it is linked to the positively selected allele. This phenomenon, called genetic hitch-hiking by Maynard Smith and Haigh [29], has been studied by many authors. Maynard Smith and Haigh [29] and Stephan and coauthors [30] use deterministic models for the change in the frequency of the selected allele. Kaplan and coauthors [20] and Barton [3] present more precise models taking into account the randomness of the first and third periods of the mutant invasion. Durrett and Schweinsberg [14, 28], Etheridge and coauthors [16], Pfaffelhuber and Studeny [26], and Leocard [21] describe the population process by a structured coalescent and finely study genealogies of neutral alleles during the sweep. Eriksson and coauthors [15] describe a deterministic approximation for the growth of the favored allele frequency during a sweep, which leads to more accurate approximation than previous models for large values of the recombination probability. Unlike our model, in all these works, the population size is constant and the individuals' "selective value" does not depend on the population state, but only on the individuals' genotype.

The structure of the paper is the following. In Section 2 we describe the model, review some results of Champagnat in [5] about the two-dimensional population process when we do not consider the neutral locus, and present the main results. In Section 3 we state a semimartingale decomposition of neutral proportions, key tool in the different proofs. Section 4 is devoted to the proof for the soft sweep. It relies on a comparison of the population process with a four dimensional dynamical system. In Section 5 we describe a coupling of the population process with two birth and death processes widely use in Sections 6 and 7, respectively devoted to the proofs for

the strong and the weak recombination regimes of hard sweep. The proof for the weak regime requires a fine study of the genealogies in a structured coalescent process during the first phase of the selective sweep. We use here some ideas developed in [28]. Finally in Appendix we state technical results.

This work stems from the papers of Champagnat [5] and Schweinsberg and Durrett [28]. In the sequel, c is used to denote a positive finite constant. Its value can change from line to line but it is always independent of the integer K and the positive real number ε. The set $\mathbb{N} := \{1, 2, ...\}$ denotes the set of positive integers.

2. Model and main results

We introduce the sets $\mathscr{A} = \{A, a\}$, $\mathscr{B} = \{b_1, b_2\}$, and $\mathscr{E} = \{A, a\} \times \{b_1, b_2\}$ to describe the genetic background of individuals. The state of the population will be given by the four dimensional Markov process $N^{(z,K)} = (N^{(z,K)}_{\alpha\beta}(t), (\alpha, \beta) \in \mathscr{E})_{t\geq 0}$ where $N^{(z,K)}_{\alpha\beta}(t)$ denotes the number of individuals with alleles (α, β) at time t when the carrying capacity is $K \in \mathbb{N}$ and the initial state is $\lfloor zK \rfloor$ with $z = (z_{\alpha\beta}, (\alpha, \beta) \in \mathscr{E}) \in \mathbb{R}^{\mathscr{E}}_+$. We recall that b_1 and b_2 are neutral, thus ecological parameters only depend on the allele, A or a, carried by the individuals at their first locus. There are the following:

- For $\alpha \in \mathscr{A}$, f_α and D_α denote the birth rate and the intrinsic death rate of an individual carrying allele α.
- For $(\alpha_1, \alpha_2) \in \mathscr{A}^2$, C_{α_1,α_2} represents the competitive pressure felt by an individual carrying allele α_1 from an individual carrying allele α_2.
- $K \in \mathbb{N}$ is a parameter rescaling the competition between individuals. It can be interpreted as a scale of resources or area available, and is related to the concept of carrying capacity, which is the maximum population size that the environment can sustain indefinitely. In the sequel K will be large.
- r_K is the recombination probability per reproductive event. When two individuals with respective genotypes (α_1, β_1) and (α_2, β_2) in $\mathscr{E}$ give birth, the newborn individual, either is a clone of one parent and carries alleles (α_1, β_1) or (α_2, β_2) with probability $(1 - r_K)/2$, or has a mixed genotype (α_1, β_2) or (α_2, β_1) with probability $r_K/2$.

We will use, for every $n = (n_{\alpha\beta}, (\alpha, \beta) \in \mathscr{E}) \in \mathbb{Z}^{\mathscr{E}}_+$, and $(\alpha, \beta) \in \mathscr{E}$, the notations

$$n_\alpha = n_{\alpha b_1} + n_{\alpha b_2}, \quad n_\beta = n_{A\beta} + n_{a\beta}, \quad \text{and} \quad |n| = n_A + n_a = n_{b_1} + n_{b_2}.$$

Let us now give the transition rates of $N^{(z,K)}$ when $N^{(z,K)}(t) = n \in \mathbb{Z}^{\mathscr{E}}_+$. An individual can die either from a natural death or from competition, whose strength depends on the carrying capacity K. Thus death rate of individuals $(\alpha, \beta) \in \mathscr{E}$ is given by:

$$d^K_{\alpha\beta}(n) = \left[D_\alpha + C_{\alpha,A} n_A/K + C_{\alpha,a} n_a/K\right] n_{\alpha\beta}. \tag{2.1}$$

An individual carrying allele $\alpha \in \mathscr{A}$ produces gametes with rate f_α, thus the relative frequencies of gametes available for reproduction are $p_{\alpha\beta}(n) = f_\alpha n_{\alpha\beta}/(f_A n_A + f_a n_a)$, $(\alpha, \beta) \in \mathscr{E}$. When an individual gives birth, he chooses his mate uniformly among the gametes available. Then the probability of giving birth to an individual of a given genotype depends on the parents (the couple (ab_2, ab_1) is not able to generate an individual Ab_1). We detail the computation of $b^K_{Ab_1}(n)$:

$$\begin{aligned} b^K_{Ab_1}(n) &= f_A n_{Ab_1}[p_{Ab_1} + p_{Ab_2}/2 + p_{ab_1}/2 + (1 - r_K)p_{ab_2}/2] + f_A n_{Ab_2}[p_{Ab_1}/2 + r_K p_{ab_1}/2] \\ &\quad + f_a n_{ab_1}[p_{Ab_1}/2 + r_K p_{Ab_2}/2] + f_a n_{ab_2}(1 - r_K)p_{Ab_1}/2 \\ &= f_A n_{Ab_1} + r_K f_A f_a (n_{ab_1} n_{Ab_2} - n_{Ab_1} n_{ab_2})/(f_A n_A + f_a n_a). \end{aligned}$$

If we denote by $\bar{\alpha}$ (resp. $\bar{\beta}$) the complement of α in $\mathscr{A}$ (resp. β in $\mathscr{B}$), we obtain in the same way:

$$b^K_{\alpha\beta}(n) = f_\alpha n_{\alpha\beta} + r_K f_a f_A \frac{n_{\bar{\alpha}\beta} n_{\alpha\bar{\beta}} - n_{\alpha\beta} n_{\bar{\alpha}\bar{\beta}}}{f_A n_A + f_a n_a}, \quad (\alpha, \beta) \in \mathscr{E}. \tag{2.2}$$

The definitions of death and birth rates in (2.1) and (2.2) ensure that the number of jumps is finite on every finite interval, and the population process is well defined.

When we focus on the dynamics of traits under selection A and a, we get the process $(N_A^{(z,K)}, N_a^{(z,K)})$. It has been studied by Champagnat in [5] and its death and birth rates, which are direct consequences of (2.1) and (2.2), satisfy:

$$d_\alpha^K(n) = \sum_{\beta\in\mathscr{B}} d_{\alpha\beta}^K(n) = \Big[D_\alpha + C_{\alpha,A}\frac{n_A}{K} + C_{\alpha,a}\frac{n_a}{K}\Big] n_\alpha, \quad b_\alpha^K(n) = \sum_{\beta\in\mathscr{B}} b_{\alpha\beta}^K(n) = f_\alpha n_\alpha, \quad \alpha\in\mathscr{A}. \tag{2.3}$$

Champagnat has proved that under some conditions the rescaled population process $(N_A^{(z,K)}/K, N_a^{(z,K)}/K)$ is well approximated by the following dynamical system,

$$\dot{n}_\alpha^{(z)} = (f_\alpha - D_\alpha - C_{\alpha,A}n_A^{(z)} - C_{\alpha,a}n_a^{(z)})n_\alpha^{(z)}, \quad n_\alpha^{(z)}(0) = z_\alpha, \quad \alpha\in\mathscr{A}. \tag{2.4}$$

More precisely Theorem 3 (b) in [5] states that for every compact subset $B \subset (\mathbb{R}_+^{A\times\mathscr{B}})^* \times (\mathbb{R}_+^{a\times\mathscr{B}})^*$ and finite real number T, we have for any $\delta > 0$,

$$\lim_{K\to\infty} \sup_{z\in B} \mathbb{P}\Big(\sup_{0\le t\le T, \alpha\in\mathscr{A}} |N_\alpha^{(z,K)}(t)/K - n_\alpha^{(z)}(t)| \ge \delta \Big) = 0. \tag{2.5}$$

Moreover, if we assume

$$f_A > D_A, \quad f_a > D_a, \quad \text{and} \quad f_a - D_a > (f_A - D_A).\sup\Big\{C_{a,A}/C_{A,A}, C_{a,a}/C_{A,a}\Big\}, \tag{2.6}$$

then the dynamical system (2.4) has a unique attracting equilibrium $(0, \bar{n}_a)$ for initial condition z satisfying $z_a > 0$, and an unstable steady state $(\bar{n}_A, 0)$ where

$$\bar{n}_\alpha = \frac{f_\alpha - D_\alpha}{C_{\alpha,\alpha}} > 0, \quad \alpha\in\mathscr{A}. \tag{2.7}$$

Hence, Assumption (2.6) avoids the coexistence of alleles A and a, and $\bar{n}_\alpha$ is the equilibrium density of a monomorphic α-population per unit of carrying capacity. This implies that when K is large, the size of a monomorphic α-population stays near $\bar{n}_\alpha K$ for a long time (Theorem 3 (c) in [5]). Moreover, if we introduce the invasion fitness $S_{\alpha\bar{\alpha}}$ of a mutant α in a population $\bar{\alpha}$,

$$S_{\alpha\bar{\alpha}} = f_\alpha - D_\alpha - C_{\alpha,\bar{\alpha}}\bar{n}_{\bar{\alpha}}, \quad \alpha\in\mathscr{A}, \tag{2.8}$$

it corresponds to the per capita growth rate of a mutant α when it appears in a population $\bar{\alpha}$ at its equilibrium density $\bar{n}_{\bar{\alpha}}$. Assumption (2.6) is equivalent to

Assumption 1. *Ecological parameters satisfy*

$$\bar{n}_A > 0, \quad \bar{n}_a > 0, \quad \textit{and} \quad S_{Aa} < 0 < S_{aA}.$$

Under Assumption 1, with positive probability, the A-population becomes extinct and the a-population size reaches a vicinity of its equilibrium value $\bar{n}_a K$.

Let us now present the main results of this paper. We introduce the extinction time of the A-population, and the fixation event of the a-population. For $(z,K)\in\mathbb{R}_+^{\mathscr{E}}\times\mathbb{N}$:

$$T_{\text{ext}}^{(z,K)} := \inf\Big\{t\ge 0, N_A^{(z,K)}(t) = 0\Big\}, \quad \text{and} \quad \text{Fix}^{(z,K)} := \Big\{T_{\text{ext}}^{(z,K)} < \infty, N_a^{(z,K)}(T_{\text{ext}}^{(z,K)}) > 0\Big\}. \tag{2.9}$$

We are interested in the neutral allele proportions. We thus define for $t\ge 0$,

$$P_{\alpha,\beta}^{(z,K)}(t) = \frac{N_{\alpha\beta}^{(z,K)}(t)}{N_\alpha^{(z,K)}(t)}, \quad (\alpha,\beta)\in\mathscr{E}, K\in\mathbb{N}, z\in\mathbb{R}_+^{\mathscr{E}}, \tag{2.10}$$

the proportion of alleles β in the α-population at time t, with the convention $0/0 = 0$. More precisely, we are interested in these proportions at the end of the sweep, that is at time $T_{\text{ext}}^{(z,K)}$ when the last A-individual dies. We then introduce the neutral proportion at this time:

$$\mathscr{P}_{a,b_1}^{(z,K)} = P_{a,b_1}^{(z,K)}(T_{\text{ext}}^{(z,K)}). \tag{2.11}$$

We first focus on soft selective sweep. We assume that the alleles A and a were neutral and coexisted in a population with large carrying capacity K. At time 0, an environmental change makes the allele a favorable (in the sense of Assumption (1)). Before stating the result, let us introduce the function F, defined for every $(z,r,t) \in (\mathbb{R}_+^{\mathscr{E}})^* \times [0,1] \times \mathbb{R}_+$ by

$$F(z,r,t) = \int_0^t \frac{r f_A f_a n_A^{(z)}(s)}{f_A n_A^{(z)}(s) + f_a n_a^{(z)}(s)} \exp\Big(- r f_A f_a \int_0^s \frac{n_A^{(z)}(u) + n_a^{(z)}(u)}{f_A n_A^{(z)}(u) + f_a n_a^{(z)}(u)} du \Big) ds, \tag{2.12}$$

where $(n_A^{(z)}, n_a^{(z)})$ is the solution of the dynamical system (2.4). We notice that $F : t \in \mathbb{R}^+ \mapsto F(z,r,t)$ is non-negative and non-decreasing. Moreover, if we introduce the function $h : (z,r,t) \in (\mathbb{R}_+^{\mathscr{E}})^* \times [0,1] \times \mathbb{R}_+ \mapsto r f_A f_a \int_0^t n_A^{(z)}(s)/(f_A n_A^{(z)}(s) + f_a n_a^{(z)}(s)) ds$ non-decreasing in time, then

$$0 \le F(z,r,t) \le \int_0^t \partial_s h(z,r,s) e^{-h(z,r,s)} ds = e^{-h(z,r,0)} - e^{-h(z,r,t)}.$$

Thus $F(z,r,t)$ has a limit in $[0,1]$ when t goes to infinity and we can define

$$F(z,r) := \lim_{t\to\infty} F(z,r,t) \in [0,1]. \tag{2.13}$$

In the case of soft sweep, the selected allele gets to fixation with high probability. More precisely, Champagnat proved the following asymptotic result in [5]: under Assumption 1,

$$\lim_{K\to\infty} \mathbb{P}(\text{Fix}^{(z,K)}) = 1, \quad \forall z \in \mathbb{R}_+^{A\times\mathscr{B}} \times (\mathbb{R}_+^{a\times\mathscr{B}})^*. \tag{2.14}$$

We consider the soft selective sweep with recombination probability r_K satisfying:

Assumption 2.

$$\lim_{K\to\infty} r_K = r \in [0,1].$$

Then recalling (2.11) we get the following result whose proof is deferred in Section 4:

Theorem 1. *Let z be in $\mathbb{R}_+^{A\times\mathscr{B}} \times (\mathbb{R}_+^{a\times\mathscr{B}})^*$ and Assumptions 1 and 2 hold. Then on the fixation event $Fix^{(z,K)}$, the proportion of alleles b_1 when the A-population becomes extinct (time $T_{ext}^{(z,K)}$) converges in probability:*

$$\lim_{K\to\infty} \mathbb{P}\Big(\mathbb{1}_{Fix^{(z,K)}} \Big| \mathscr{P}_{a,b_1}^{(z,K)} - \Big[\frac{z_{Ab_1}}{z_A} F(z,r) + \frac{z_{ab_1}}{z_a}(1 - F(z,r)) \Big] \Big| > \varepsilon \Big) = 0, \quad \forall \varepsilon > 0.$$

The neutral proportion at the end of the soft sweep is thus a weighted mean of initial proportions in populations A and a. In particular, soft sweep is responsible for a diminution of the number of neutral alleles with very low or very high proportions in the population, as remarked in [27]. We notice that the weight $F(z,r)$ does not depend on the initial neutral proportions. It only depends on r and on the dynamical system (2.4) with initial condition $(n_A(0), n_a(0)) = (z_A, z_a)$.

Now we focus on hard selective sweep: a mutant a appears in a large population and gets to fixation. We assume that the mutant appears when the A-population is at ecological equilibrium, and carries the neutral allele b_1. In other words, recalling Definition (2.7), we assume:

Assumption 3. *There exists $z_{Ab_1} \in]0, \bar{n}_A[$ such that $N^{(z^{(K)},K)}(0) = \lfloor z^{(K)} K \rfloor$ with*

$$z^{(K)} = (z_{Ab_1}, \bar{n}_A - z_{Ab_1}, K^{-1}, 0).$$

In this case, the selected allele gets to fixation with positive probability. More precisely, Champagnat proved the following asymptotic result in [5]: under Assumptions 1 and 3,

$$\lim_{K\to\infty} \mathbb{P}\Big(\mathrm{Fix}^{(z^{(K)},K)}\Big) = \frac{S_{aA}}{f_a}. \tag{2.15}$$

In the case of strong selective sweep we will distinguish two different recombination regimes:

Assumption 4. *Strong recombination*

$$\lim_{K\to\infty} r_K \log K = \infty.$$

Assumption 5. *Weak recombination*

$$\limsup_{K\to\infty} r_K \log K < \infty.$$

Recall (2.11). Then we have the following results whose proofs are deferred in Sections 6 and 7:

Theorem 2. *Suppose that Assumptions 1 and 3 hold. Then on the fixation event $Fix^{(z,K)}$ and under Assumption 4 or 5, the proportion of alleles b_1 when the A-population becomes extinct (time $T_{ext}^{(z,K)}$) converges in probability. More precisely, if Assumption 4 holds,*

$$\lim_{K\to\infty} \mathbb{P}\Big(\mathbb{1}_{Fix^{(z,K)}} \Big| \mathscr{P}_{a,b_1}^{(z^{(K)},K)} - \frac{z_{Ab_1}}{z_A}\Big| > \varepsilon\Big) = 0, \quad \forall \varepsilon > 0,$$

and if Assumption 5 holds,

$$\lim_{K\to\infty} \mathbb{P}\Big(\mathbb{1}_{Fix^{(z,K)}} \Big| \mathscr{P}_{a,b_1}^{(z^{(K)},K)} - \Big[\frac{z_{Ab_1}}{z_A} + \frac{z_{Ab_2}}{z_A}\exp\Big(-\frac{f_a r_K \log K}{S_{aA}}\Big)\Big]\Big| > \varepsilon\Big) = 0, \quad \forall \varepsilon > 0.$$

As stated in [5], the selective sweep has a duration of order $\log K$. Thus, when $r_K \log K$ is large, a lot of recombinations occur during the sweep, and the neutral alleles are constantly exchanged by the populations A and a. Hence in the strong recombination case, the sweep does not modifiy the proportion of neutral alleles. On the contrary, when r_K is of order $1/\log K$ the number of recombinations undergone by a given lineage does not go to infinity, and the frequency of the neutral allele b_1 carried by the first mutant a increases. This phenomenon is called genetic hitchhiking [29]: the selective sweep leads to a diminution of diversity around the selected allele.

Remark 1. *The limits in the two regimes are consistent in the sense that*

$$\lim_{r_K \log K\to\infty} \frac{z_{Ab_2}}{z_A}\exp\Big(-\frac{f_a r_K \log K}{S_{aA}}\Big) = 0.$$

Moreover, let us notice that we can easily extend the results of Theorems 1 and 2 to a finite number of possible alleles b_1, b_2, ..., b_i on the neutral locus.

3. A SEMI-MARTINGALE DECOMPOSITION

The expression of birth rate in (2.2) shows that the effect of recombination depends on the recombination probability r_K but also on the population state via the term $n_{\tilde{\alpha}\beta}n_{\alpha\tilde{\beta}} - n_{\alpha\beta}n_{\tilde{\alpha}\tilde{\beta}}$. This quantity is linked with the linkage disequilibrium of the population, which is the occurrence of some allele combinations more or less often than would be expected from a random formation of haplotypes (see [13] Section 3.3 for an introduction to this notion or [22] for a study of its structure around a sweep). Proposition 1 states a semi-martingale representation of the neutral allele proportions.

Proposition 1. *Let (α, z, K) be in $\mathscr{A} \times (\mathbb{R}_+^{\mathscr{E}})^* \times \mathbb{N}$. The process $(P_{\alpha,b_1}^{(z,K)}(t), t \geq 0)$ defined in (2.10) is a semi-martingale and we have the following decomposition:*

$$P_{\alpha,b_1}^{(z,K)}(t) = P_{\alpha,b_1}^{(z,K)}(0) + M_\alpha^{(z,K)}(t) + r_K f_A f_a \int_0^t \frac{N_{\tilde{\alpha}b_1}^{(z,K)} N_{\alpha b_2}^{(z,K)} - N_{\alpha b_1}^{(z,K)} N_{\tilde{\alpha}b_2}^{(z,K)}}{(N_\alpha^{(z,K)}+1)(f_A N_A^{(z,K)} + f_a N_a^{(z,K)})}, \tag{3.1}$$

where the process $(M_\alpha^{(z,K)}(t), t \geq 0)$ *is a martingale bounded on every interval* $[0,t]$ *whose quadratic variation is given by* (3.7).

To enlight the presentation in remarks and proofs we shall mostly write N instead $N^{(z,K)}$.

Remark 2. *The process* $N_{ab_2}N_{Ab_1} - N_{ab_1}N_{Ab_2}$ *will play a major role in the dynamics of neutral proportions. Indeed it is a measure of the neutral proportion disequilibrium between the* A *and* a*-populations as it satisfies:*

$$N_A N_a (P_{A,b_1} - P_{a,b_1}) = N_{ab_2}N_{Ab_1} - N_{ab_1}N_{Ab_2}. \tag{3.2}$$

Proof of Proposition 1. In the vein of Fournier and Méléard [18] we represent the population process in terms of Poisson measure. Let $Q(ds,d\theta)$ be a Poisson random measure on $\mathbb{R}_+^2$ with intensity $dsd\theta$, and $(e_{\alpha\beta}, (\alpha,\beta) \in \mathscr{E})$ the canonical basis of $\mathbb{R}^{\mathscr{E}}$. According to (2.3) a jump occurs at rate $\sum_{(\alpha,\beta)\in\mathscr{E}}(b^K_{\alpha\beta}(N) + d^K_{\alpha\beta}(N)) = f_a N_a + d_a^K(N) + f_A N_A + d_A^K(N)$. We decompose on possible jumps that may occur: births and deaths for a-individuals and births and deaths for A-individuals. Itô's formula with jumps (see [19] p. 66) yields for every function h measurable and bounded on $\mathbb{R}_+^{\mathscr{E}}$:

$$\begin{aligned} h(N(t)) \;=\; & h(N(0)) + \int_0^t \int_{R_+} \Big\{ \sum_{\alpha\in\mathscr{A}} \Big(h(N(s^-) + e_{\alpha b_1}) \mathbb{1}_{0<\theta-\mathbb{1}_{\alpha=A}(f_aN_a(s^-)+d_a^K(N(s^-))\leq b^K_{ab_1}(N(s^-))} \\ & + h(N(s^-) + e_{\alpha b_2}) \mathbb{1}_{b^K_{ab_1}(N(s^-))<\theta-\mathbb{1}_{\alpha=A}(f_aN_a(s^-)+d_a^K(N(s^-))\leq f_\alpha N_\alpha(s^-)} \\ & + h(N(s^-) - e_{\alpha b_1}) \mathbb{1}_{0<\theta-f_\alpha N_\alpha(s^-)-\mathbb{1}_{\alpha=A}(f_aN_a(s^-)+d_a^K(N(s^-))\leq d^K_{ab_1}(N(s^-))} \\ & + h(N(s^-) - e_{\alpha b_2}) \mathbb{1}_{d^K_{ab_1}(N(s^-))<\theta-f_\alpha N_\alpha(s^-)-\mathbb{1}_{\alpha=A}(f_aN_a(s^-)+d_a^K(N(s^-))\leq d_\alpha^K(N(s^-))} \Big) \\ & - h(N(s^-)) \mathbb{1}_{\theta\leq f_aN_a(s^-)+d_a^K(N(s^-))+f_AN_A(s^-)+d_A^K(N(s^-))} \Big\} Q(ds,d\theta). \end{aligned} \tag{3.3}$$

Let us introduce the functions μ_K^α defined for $\alpha \in \mathscr{A}$ and (s,θ) in $\mathbb{R}_+ \times \mathbb{R}_+$ by,

$$\begin{aligned} \mu_K^\alpha(N,s,\theta) \;=\; & \frac{N_{\alpha b_2}(s)}{(N_\alpha(s)+1)N_\alpha(s)} \mathbb{1}_{0<\theta-\mathbb{1}_{\alpha=A}(f_aN_a(s)+d_a^K(N(s))\leq b^K_{ab_1}(N(s))} \\ & - \frac{N_{\alpha b_1}(s)}{(N_\alpha(s)+1)N_\alpha(s)} \mathbb{1}_{b^K_{ab_1}(N(s))<\theta-\mathbb{1}_{\alpha=A}(f_aN_a(s)+d_a^K(N(s))\leq f_\alpha N_\alpha(s)} \\ & - \frac{N_{\alpha b_2}(s)}{(N_\alpha(s)-1)N_\alpha(s)} \mathbb{1}_{0<\theta-f_\alpha N_\alpha(s)-\mathbb{1}_{\alpha=A}(f_aN_a(s)+d_a^K(N(s))\leq d^K_{ab_1}(N(s))} \\ & + \frac{N_{\alpha b_1}(s)}{(N_\alpha(s)-1)N_\alpha(s)} \mathbb{1}_{d^K_{ab_1}(N(s))<\theta-f_\alpha N_\alpha(s)-\mathbb{1}_{\alpha=A}(f_aN_a(s)+d_a^K(N(s))\leq d_\alpha^K(N(s))}, \end{aligned} \tag{3.4}$$

with the convention $0/0 = 0$. Then we can represent the neutral allele proportions P_{α,b_1} as,

$$P_{\alpha,b_1}(t) = P_{\alpha,b_1}(0) + \int_0^t \int_{\mathbb{R}^+} \mu_K^\alpha(N,s^-,\theta) Q(ds,d\theta), \quad t \geq 0. \tag{3.5}$$

A direct calculation gives

$$\int_0^\infty \mu_K^\alpha(N,s,\theta)d\theta = r_K f_A f_a \frac{N_{\tilde\alpha b_1}(s)N_{\alpha b_2}(s) - N_{\alpha b_1}(s)N_{\tilde\alpha b_2}(s)}{(N_\alpha(s)+1)(f_AN_A(s)+f_aN_a(s))}.$$

Thus if we introduce the compensated Poisson measure $\tilde Q(ds,d\theta) := Q(ds,d\theta) - dsd\theta$, then

$$M_\alpha(t) := \int_0^t \int_{\mathbb{R}^+} \mu_K^\alpha(N,s^-,\theta)\tilde Q(ds,d\theta) = P_{\alpha,b_1}(t) - P_{\alpha,b_1}(0) - r_K f_A f_a \int_0^t \frac{N_{\tilde\alpha b_1}N_{\alpha b_2} - N_{\alpha b_1}N_{\tilde\alpha b_2}}{(N_\alpha+1)(f_AN_A+f_aN_a)}$$

is a local martingale. By construction the process P_{α,b_1} has values in $[0,1]$ and as $r_K \leq 1$,

$$\sup_{s\leq t} \Big| r_K f_A f_a \int_0^s \frac{N_{\tilde\alpha b_1}N_{\alpha b_2} - N_{\alpha b_1}N_{\tilde\alpha b_2}}{(N_\alpha+1)(f_AN_A+f_aN_a)} \Big| \leq r_K f_\alpha t \leq f_\alpha t, \quad t \geq 0. \tag{3.6}$$

Thus M_α is a square integrable pure jump martingale bounded on every finite interval with quadratic variation

$$\begin{aligned}\langle M_\alpha\rangle_t &= \int_0^t\int_{\mathbb{R}^+}\Big(\mu_K^\alpha(N,s,\theta)\Big)^2 ds d\theta \\ &= \int_0^t\Big\{P_{\alpha,b_1}(1-P_{\alpha,b_1})\Big[\Big(D_\alpha+\frac{C_{\alpha,\alpha}}{K}N_\alpha+\frac{C_{\alpha,\bar\alpha}}{K}N_{\bar\alpha}\Big)\frac{\mathbb{1}_{N_\alpha\geq 2}N_\alpha}{(N_\alpha-1)^2} \\ &\qquad +\frac{f_\alpha N_\alpha}{(N_\alpha+1)^2}\Big]+\frac{r_K f_A f_a(N_{\bar\alpha b_1}N_{\alpha b_2}-N_{\alpha b_1}N_{\bar\alpha b_2})(1-2P_{\alpha,b_1})}{(N_\alpha+1)^2(f_A N_A+f_a N_a)}\Big\}. \qquad (3.7)\end{aligned}$$

This ends up the proof of Proposition 1. □

Remark 3. *Let us mention two properties of the functions μ_K^α, defined in* (3.4)*, which will be useful in the sequel. Firstly by definition we have for all (s,θ) in $\mathbb{R}_+\times\mathbb{R}_+$,*

$$\mu_K^A(N,s,\theta)\mu_K^a(N,s,\theta)=0. \qquad (3.8)$$

Secondly, the convention $0/0=0$ yields the following equality for $\alpha\in\mathscr{A}$ and all (s,θ) in $\mathbb{R}_+\times\mathbb{R}_+$,

$$\mu_K^\alpha(N,s,\theta)\mathbb{1}_{N_\alpha(s)\geq 1}=\mu_K^\alpha(N,s,\theta). \qquad (3.9)$$

For sake of simplicity we will use more often the second notation, but in Section 6 the first notation will also be useful.

Lemma 3.1 states properties of the quadratic variation widely used in the forthcoming proofs. We introduce a compact interval containing the equilibrium size of the A-population,

$$I_\varepsilon^K:=\Big[K\Big(\bar n_A-2\varepsilon\frac{C_{A,a}}{C_{A,A}}\Big),K\Big(\bar n_A+2\varepsilon\frac{C_{A,a}}{C_{A,A}}\Big)\Big], \qquad (3.10)$$

and the stopping times T_ε^K and $\tilde T_\varepsilon^K$, which denote respectively the hitting time of $\lfloor\varepsilon K\rfloor$ by the mutant population and the exit time of I_ε^K by the resident population,

$$T_\varepsilon^K:=\inf\Big\{t\geq 0,N_a^K(t)=\lfloor\varepsilon K\rfloor\Big\},\quad \tilde T_\varepsilon^K:=\inf\Big\{t\geq 0,N_A^K(t)\notin I_\varepsilon^K\Big\}. \qquad (3.11)$$

Lemma 3.1. *For $v<\infty$, there exists a finite $C(v)$ such that $(N_A^{(z,K)}(t),N_a^{(z,K)}(t))\in[0,vK]^2$ implies*

$$\frac{d}{dt}\langle M_\alpha^{(z,K)}\rangle_t=\int_{\mathbb{R}^+}\Big(\mu_K^\alpha(N^{(z,K)},t,\theta)\Big)^2 d\theta\leq C(v)\frac{\mathbb{1}_{N_\alpha(t)\geq 1}}{N_\alpha(t)},\quad \alpha\in\mathscr{A}. \qquad (3.12)$$

Under Assumptions 1 and 3, there exist $k_0\in\mathbb{N}$, $\varepsilon_0>0$ and a pure jump martingale $\tilde M$ such that for $\varepsilon\leq\varepsilon_0$ and $t\geq 0$,

$$e^{\frac{S_{aA}}{2(k_0+1)}t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K}\int_{\mathbb{R}^+}\Big[\mu_K^a(N^{(z^{(K)},K)},t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K,\theta)\Big]^2 d\theta\leq(k_0+1)C\Big(\bar n_A+2\varepsilon\frac{C_{A,a}}{C_{A,A}}\Big)\tilde M_{t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K}, \qquad (3.13)$$

and

$$\mathbb{E}\Big[\tilde M_{t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K}\Big]\leq\frac{1}{k_0+1}. \qquad (3.14)$$

Proof. Equation (3.12) is a direct consequence of (3.7). To prove (3.13) and (3.14), let us first notice that according to Assumption 1, there exists $k_0\in\mathbb{N}$ such that for ε small enough and $k\in\mathbb{Z}_+$,

$$\frac{f_a(k_0+k-1)-(D_a+C_{a,A}\bar n_A+\varepsilon(C_{a,a}+2C_{A,a}C_{a,A}/C_{A,A}))(k_0+k+1)}{k_0+k-1}\geq\frac{S_{aA}}{2}.$$

This implies in particular that for every $t<T_\varepsilon^K\wedge\tilde T_\varepsilon^K$,

$$\frac{f_aN_a(t)(N_a(t)+k_0-1)-d_a(N(t))(N_a(t)+k_0+1)}{(N_a(t)+k_0-1)(N_a(t)+k_0+1)}\geq\frac{S_{aA}N_a(t)}{2(N_a(t)+k_0+1)}\geq\frac{S_{aA}\mathbb{1}_{N_a(t)\geq 1}}{2(k_0+1)}, \qquad (3.15)$$

where the death rate d_a has been defined in (2.3). For sake of simplicity let us introduce the process X defined as follows:

$$X(t) = \frac{\mathbb{1}_{N_a(t)\geq 1}}{N_a(t)+k_0}\exp\Big(\frac{S_{aA}t}{2(k_0+1)}\Big), \quad \forall\, t\geq 0.$$

Applying Itô's formula with jumps we get for every $t\geq 0$:

$$X(t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K) = \tilde M(t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K)+ \int_0^{t\wedge T_\varepsilon^K\wedge\tilde T_\varepsilon^K}\Big(\frac{S_{aA}}{2(k_0+1)} - \frac{f_aN_a(s)(N_a(s)+k_0-1)-d_a^K(N(s))(N_a(s)+k_0+1)}{(N_a(s)+k_0-1)(N_a(s)+k_0+1)}\Big)X(s)ds, \tag{3.16}$$

where the martingale $\tilde M$ has the following expression:

$$\tilde M(t) = \frac{1}{k_0+1}+\int_0^t\int_{\mathbb{R}_+}\tilde Q(ds,d\theta)\mathbb{1}_{N_a(s^-)\geq 1}\exp\Big(\frac{S_{aA}s}{2(k_0+1)}\Big) \Big[\frac{\mathbb{1}_{\theta\leq f_aN_a(s^-)}}{N_a(s^-)+k_0+1}+\frac{\mathbb{1}_{f_aN_a(s^-)<\theta\leq f_aN_a(s^-)+d_a(N(s^-))}}{N_a(s^-)+k_0-1}-\frac{\mathbb{1}_{\theta\leq f_aN_a(s^-)+d_a(N(s^-))}}{N_a(s^-)+k_0}\Big]. \tag{3.17}$$

Thanks to (3.15) the integral in (3.16) is nonpositive. Moreover, according to (3.12), for $t\leq T_\varepsilon^K\wedge\tilde T_\varepsilon^K$,

$$\int_{\mathbb{R}^+}\Big(\mu_K^a(N^{(z^{(K)},K)},t,\theta)\Big)^2d\theta \leq C\Big(\bar n_A+2\varepsilon\frac{C_{A,a}}{C_{A,A}}\Big)\frac{\mathbb{1}_{N_\alpha(t)\geq 1}}{N_\alpha(t)} \leq (k_0+1)C\Big(\bar n_A+2\varepsilon\frac{C_{A,a}}{C_{A,A}}\Big)X(t)\exp\Big(-\frac{S_{aA}t}{2(k_0+1)}\Big) \tag{3.18}$$

which ends the proof. □

4. Proof of Theorem 1

In this section we suppose that Assumptions 1 and 2 hold. For $\varepsilon\leq C_{a,a}/C_{a,A}\wedge 2|S_{Aa}|/C_{A,a}$ and z in $(\mathbb{R}_+^{A\times\mathscr{B}})^*\times(\mathbb{R}_+^{a\times\mathscr{B}})^*$ we introduce a deterministic time $t_\varepsilon(z)$ after which the solution $(n_A^{(z)},n_a^{(z)})$ of the dynamical system (2.4) is close to the stable equilibrium $(0,\bar n_a)$:

$$t_\varepsilon(z) := \inf\big\{s\geq 0,\forall\, t\geq s,(n_A^{(z)}(t),n_a^{(z)}(t))\in[0,\varepsilon^2/2]\times[\bar n_a-\varepsilon/2,\infty)\big\}. \tag{4.1}$$

Once $(n_A^{(z)},n_a^{(z)})$ has reached the set $[0,\varepsilon^2/2]\times[\bar n_a-\varepsilon/2,\infty)$ it no more escapes from it. Moreover, according to Assumption 1 on stable equilibrium, $t_\varepsilon(z)$ is finite.

First we compare the population process with a four dimensional dynamical system on the time interval $[0,t_\varepsilon(z)]$. Then we study this dynamical system and get an approximation of the neutral proportions at time $t_\varepsilon(z)$. Finally, we state that during the A-population extinction period, this proportion stays nearly constant.

4.1. Comparison with a four dimensional dynamical system.

We denote by $n^{(z)}=(n_{\alpha\beta}^{(z)},(\alpha,\beta)\in\mathscr{E})$ the solution of

$$\dot n_{\alpha\beta}^{(z)} = \Big(f_\alpha-(D_\alpha+C_{\alpha,A}n_A^{(z)}+C_{\alpha,a}n_a^{(z)})\Big)n_{\alpha\beta}^{(z)}+rf_Af_a\frac{n_{\tilde\alpha\beta}^{(z)}n_{\alpha\tilde\beta}^{(z)}-n_{\alpha\beta}^{(z)}n_{\tilde\alpha\tilde\beta}^{(z)}}{f_An_A^{(z)}+f_an_a^{(z)}}, \quad (\alpha,\beta)\in\mathscr{E}, \tag{4.2}$$

with initial condition $n^{(z)}(0)=z\in\mathbb{R}_+^{\mathscr{E}}$. Then we have the following comparison result:

Lemma 4.1. *Let z be in $\mathbb{R}_+^{\mathscr{E}}$ and ε be in $\mathbb{R}_+^*$. Then*

$$\lim_{K\to\infty}\sup_{s\leq t_\varepsilon(z)}\|N^{(z,K)}(s)/K-n^{(z)}(s)\| = 0 \quad a.s. \tag{4.3}$$

where $\|.\|$ denotes the L^1-Norm on $\mathbb{R}^{\mathscr{E}}$.

Proof. The proof relies on a slight modification of Theorem 2.1 p. 456 in Ethier and Kurtz [17]. According to (2.1) and (2.2), the rescaled birth and death rates

$$\tilde{b}^K_{\alpha\beta}(n) = \frac{1}{K} b^K_{\alpha\beta}(Kn) = f_\alpha n_{\alpha\beta} + r_K f_a f_A \frac{n_{\bar{\alpha}\beta} n_{\alpha\bar{\beta}} - n_{\alpha\beta} n_{\bar{\alpha}\bar{\beta}}}{f_A n_A + f_a n_a}, \quad (\alpha,\beta) \in \mathscr{E}, n \in N^{\mathscr{E}},$$

and

$$\tilde{d}_{\alpha\beta}(n) = \frac{1}{K} d^K_{\alpha\beta}(Kn) = \left[D_\alpha + C_{\alpha,A} n_A + C_{\alpha,a} n_a \right] n_{\alpha\beta}, \quad (\alpha,\beta) \in \mathscr{E}, n \in N^{\mathscr{E}},$$

are Lipschitz and bounded on every compact subset of $\mathbb{N}^{\mathscr{E}}$. The only difference with [17] is that $\tilde{b}^K_{\alpha\beta}$ depends on K via the term r_K. Applying Itô's formula with jumps we get:

$$\frac{N^{(z,K)}(t)}{K} = \frac{\lfloor zK \rfloor}{K} + Mart^{(z,K)}(t) + \int_0^t \sum_{(\alpha,\beta)\in\mathscr{E}} e_{\alpha\beta} \Big(\tilde{b}^K_{\alpha\beta}\Big(\frac{N^{(z,K)}(s)}{K} \Big) - \tilde{d}_{\alpha\beta}\Big(\frac{N^{(z,K)}(s)}{K} \Big) \Big) ds,$$

where $Mart^{(z,K)}$ is a martingale, and we recall that $(e_{\alpha\beta}, (\alpha,\beta) \in \mathscr{E})$ is the canonical basis of $\mathbb{R}^{\mathscr{E}}_+$. If we denote by $\tilde{b}^\infty_{\alpha\beta}$ the function

$$\tilde{b}^\infty_{\alpha\beta}(n) = f_\alpha n_{\alpha\beta} + r f_a f_A \frac{n_{\bar{\alpha}\beta} n_{\alpha\bar{\beta}} - n_{\alpha\beta} n_{\bar{\alpha}\bar{\beta}}}{f_A n_A + f_a n_a}, \quad (\alpha,\beta) \in \mathscr{E}, n \in N^{\mathscr{E}},$$

we get

$$n^{(z)}(t) = z + \int_0^t \sum_{(\alpha,\beta)\in\mathscr{E}} e_{\alpha\beta} \Big(\tilde{b}^\infty_{\alpha\beta}\Big(n^{(z)}(s) \Big) - \tilde{d}_{\alpha\beta}\Big(n^{(z)}(s) \Big) \Big) ds.$$

Hence we have for every $t \le t_\varepsilon(z)$,

$$\Big| \frac{N^{(z,K)}(t)}{K} - n^{(z)}(t) \Big| \le \Big| \frac{\lfloor zK \rfloor}{K} - z \Big| + \Big| Mart^{(z,K)}(t) \Big| + \int_0^t \sum_{(\alpha,\beta)\in\mathscr{E}} \Big| \tilde{b}^\infty_{\alpha\beta}\Big(n^{(z)}(s) \Big) - \tilde{b}^K_{\alpha\beta}\Big(n^{(z)}(s) \Big) \Big| ds$$
$$+ \int_0^t \sum_{(\alpha,\beta)\in\mathscr{E}} \Big| \Big(\tilde{b}^K_{\alpha\beta} - \tilde{d}_{\alpha\beta} \Big)\Big(\frac{N^{(z,K)}(s)}{K} \Big) - \Big(\tilde{b}^K_{\alpha\beta} - \tilde{d}_{\alpha\beta} \Big)\Big(n^{(z)}(s) \Big) \Big| ds,$$

and there exists a finite constant M such that

$$\Big| \frac{N^{(z,K)}(t)}{K} - n^{(z)}(t) \Big| \le \frac{1}{K} + \Big| Mart^{(z,K)}(t) \Big| + |r - r_K| M t_\varepsilon(z) + M \int_0^t \Big| \frac{N^{(z,K)}(s)}{K} - n^{(z)}(s) \Big| ds.$$

But following Ethier and Kurtz, we get

$$\lim_{K\to\infty} \sup_{s \le t_\varepsilon(z)} |Mart^{(z,K)}| = 0, \quad \text{a.s.},$$

and we end the proof by using Assumption 2 and Gronwall's Lemma. □

Once we know that the rescaled population process is close to the solution of the dynamical system (4.2), we can study this latter.

Lemma 4.2. *Let z be in $\mathbb{R}^{\mathscr{E}}_+$ such that $z_A > 0$ and $z_a > 0$. Then $n^{(z)}_a(t)$ and $n^{(z)}_{ab_1}(t)$ have a finite limit when t goes to infinity, and there exists a positive constant ε_0 such that for every $\varepsilon \le \varepsilon_0$,*

$$\Big| \frac{n^{(z)}_{ab_1}(\infty)}{n^{(z)}_a(\infty)} - \frac{n^{(z)}_{ab_1}(t_\varepsilon(z))}{n^{(z)}_a(t_\varepsilon(z))} \Big| \le \frac{2 f_a \varepsilon^2}{\bar{n}_A |S_{aA}|}.$$

Proof. Assumption 1 ensures that $n^{(z)}_a(t)$ goes to $\bar{n}_a$ at infinity. If we define the functions

$$p^{(z)}_{\alpha,b_1} = n^{(z)}_{\alpha b_1} / n^{(z)}_\alpha, \quad \alpha \in \mathscr{A}, \quad \text{and} \quad g^{(z)} = p^{(z)}_{A,b_1} - p^{(z)}_{a,b_1},$$

we easily check that $\phi : (n^{(z)}_{Ab_1}, n^{(z)}_{Ab_2}, n^{(z)}_{ab_1}, n^{(z)}_{ab_2}) \mapsto (n^{(z)}_A, n^{(z)}_a, g^{(z)}, p^{(z)}_{a,b_1})$ defines a change of variables from $(\mathbb{R}_+^*)^{\mathscr{E}}$ to $\mathbb{R}_+^{2*} \times]-1,1[\times]0,1[$, and (4.2) is equivalent to:

$$\left\{\begin{array}{l} \dot{n}^{(z)}_\alpha = (f_\alpha - (D_\alpha + C_{\alpha,A} n^{(z)}_A + C_{\alpha,a} n^{(z)}_a)) n^{(z)}_\alpha, \quad \alpha \in \mathscr{A} \\ \dot{g}^{(z)} = -g^{(z)}\Big(r f_A f_a (n^{(z)}_A + n^{(z)}_a)/(f_A n^{(z)}_A + f_a n^{(z)}_a)\Big) \\ \dot{p}^{(z)}_{a,b_1} = g^{(z)}\Big(r f_A f_a n^{(z)}_A/(f_A n^{(z)}_A + f_a n^{(z)}_a)\Big), \end{array}\right. \tag{4.4}$$

with initial condition $(n^{(z)}_A(0), n^{(z)}_a(0), g^{(z)}(0), p^{(z)}_{a,b_1}(0)) = (z_A, z_a, z_{Ab_1}/z_A - z_{ab_1}/z_a, z_{ab_1}/z_a)$. Moreover, a direct integration yields

$$p^{(z)}_{a,b_1}(t) = p^{(z)}_{a,b_1}(0) - (p^{(z)}_{a,b_1}(0) - p^{(z)}_{A,b_1}(0)) F(z,r,t), \tag{4.5}$$

where F has been defined in (2.12). According to (2.13), $F(z,r,t)$ has a finite limit when t goes to infinity. Hence $p^{(z)}_{a,b_1}$ also admits a limit at infinity. Let $\varepsilon \le |S_{Aa}|/C_{A,a} \wedge C_{a,a}/C_{A,a} \wedge \bar{n}_a/2$, and $t_\varepsilon(z)$ defined in (4.1). Then for $t \ge t_\varepsilon(z)$,

$$\dot{n}^{(z)}_A(t) \le (f_A - D_A - C_{Aa}(\bar{n}_a - \varepsilon/2)) n^{(z)}_A(t) \le S_{Aa} n^{(z)}_A(t)/2 < 0.$$

Recalling that $r \le 1$ and $|g(t)| \le 1$ for all $t \ge 0$ we get:

$$\Big| p^{(z)}_{a,b_1}(\infty) - p^{(z)}_{a,b_1}(t_\varepsilon(z)) \Big| \le \int_{t_\varepsilon(z)}^\infty \frac{f_A f_a n^{(z)}_A}{f_A n^{(z)}_A + f_a n^{(z)}_a} \le \frac{f_A \varepsilon^2}{\bar{n}_a} \int_0^\infty e^{S_{Aa} s/2} ds \le \frac{2 f_a \varepsilon^2}{\bar{n}_a |S_{aA}|}, \tag{4.6}$$

which ends up the proof. □

4.2. A-**population extinction.** The deterministic approximation (4.2) fails when the A-population size becomes too small. We shall compare N_A with birth and death processes to study the last period of mutant invasion. We show that during this period, the number of A individuals is so small that it has no influence on the neutral proportion in the a-population, which stays nearly constant. Before stating the result, we recall Definition (2.9), introduce the compact set Θ:

$$\Theta := \big\{ z \in \mathbb{R}_+^{A\times\mathscr{B}} \times (\mathbb{R}_+^{a\times\mathscr{B}})^*, z_A \le \varepsilon^2 \quad \text{and} \quad |z_a - \bar{n}_a| \le \varepsilon \big\}, \tag{4.7}$$

the constant $M'' = 3 + (f_a + C_{a,A})/C_{a,a}$, and the stopping time:

$$S^K_\varepsilon(z) := \inf\Big\{ t \ge 0, N^{(z,K)}_A(t) > \varepsilon K \text{ or } |N^{(z,K)}_a(t) - \bar{n}_a K| > M'' \varepsilon K \Big\}. \tag{4.8}$$

Lemma 4.3. *Let z be in Θ. Then under Assumption 1, there exist two positive finite constants c and ε_0 such that for $\varepsilon \le \varepsilon_0$,*

$$\limsup_{K\to\infty} \mathbb{P}\Big(\sup_{t \le T^{(z,K)}_{ext}} \Big| P^{(z,K)}_{a,b_1}(t) - P^{(z,K)}_{a,b_1}(0) \Big| > \varepsilon \Big) \le c\varepsilon.$$

Proof. Let z be in Θ and Z^1 be a birth and death process with birth rate f_A, death rate $D_A + (\bar{n}_a - M''\varepsilon) C_{A,a}$, and initial state $\lceil \varepsilon^2 K \rceil$. Then on $[0, S^K_\varepsilon(z)[$, N_A and Z^1 have the same birth rate, and Z^1 has a smaller death rate than N_A. Thus according to Theorem 2 in [5], we can construct the processes N and Z^1 on the same probability space such that:

$$N_A(t) \le Z^1_t, \quad \forall\, t < S^K_\varepsilon(z). \tag{4.9}$$

Moreover, if we denote by T^1_0 the extinction time of Z^1, $T^1_0 := \inf\{t \ge 0, Z^1_t = 0\}$, and recall that

$$f_A - D_A - (\bar{n}_a - M''\varepsilon) C_{A,a} = S_{Aa} + M'' C_{A,a} \varepsilon < S_{Aa}/2 < 0, \quad \forall \varepsilon < |S_{Aa}|/(2M'' C_{Aa}), \tag{4.10}$$

we get according to (A.10) that for $z \le \varepsilon^2$ and $L(\varepsilon, K) = 2\log K/|S_{Aa} + M''\varepsilon C_{A,a}|$,

$$\mathbb{P}_{\lceil zK \rceil}\Big(T^1_0 \le L(\varepsilon, K) \Big) \ge \exp\Big(\lceil \varepsilon^2 K \rceil \Big[\log(K^2 - 1) - \log(K^2 - f_A (D_A + (\bar{n}_a - M''\varepsilon) C_{A,a})^{-1}) \Big] \Big).$$

Thus:

$$\lim_{K\to\infty} \mathbb{P}_{\lceil zK \rceil}\Big(T^1_0 < L(\varepsilon, K) \Big) = 1. \tag{4.11}$$

Moreover, Equation (A.4) ensures the existence of a finite c such that for ε small enough,

$$\mathbb{P}\Big(L(\varepsilon,K) < S_\varepsilon^K(z)\Big) \geq 1 - c\varepsilon. \tag{4.12}$$

Equations (4.11) and (4.12) imply

$$\liminf_{K\to\infty} \mathbb{P}\Big(T_0^1 < L(\varepsilon,K) < S_\varepsilon^K(z)\Big) \geq 1 - c\varepsilon \tag{4.13}$$

for a finite c. According to Coupling (4.9) he have the inclusion $\{T_0^1 < L(\varepsilon,K) < S_\varepsilon^K(z)\} \subset \{T_{ext}^K < L(\varepsilon,K) < S_\varepsilon^K(z)\}$. Adding (4.13) we finally get:

$$\liminf_{K\to\infty} \mathbb{P}(T_{ext}^K < L(\varepsilon,K) < S_\varepsilon^K(z)) \geq 1 - c\varepsilon. \tag{4.14}$$

Recall the martingale decomposition of P_{a,b_1} in (3.1). To bound the difference $|P_{a,b_1}(t) - P_{a,b_1}(0)|$ we bound independently the martingale $M_a(t)$ and the integral $|P_{a,b_1}(t) - P_{a,b_1}(0) - M_a(t)|$. On one hand Doob's Maximal Inequality and Equation (3.12) imply:

$$\mathbb{P}\Big(\sup_{t\leq L(\varepsilon,K)\wedge S_\varepsilon^K} |M_a(t)| > \frac{\varepsilon}{2}\Big) \leq \frac{4}{\varepsilon^2}\mathbb{E}\Big[\langle M_a\rangle_{L(\varepsilon,K)\wedge S_\varepsilon^K(z)}\Big] \leq \frac{4C(\bar{n}_a + M''\varepsilon)L(\varepsilon,K)}{\varepsilon^2 K(\bar{n}_a - M''\varepsilon)}. \tag{4.15}$$

On the other hand the inequality $|N_{Ab_1}N_{ab_2} - N_{ab_1}N_{Ab_2}| \leq N_A N_a$ yields for $t \geq 0$

$$\Big|\int_0^{t\wedge S_\varepsilon^K(z)} \frac{r_K f_A f_a (N_{Ab_1}N_{ab_2} - N_{ab_1}N_{Ab_2})}{(N_a+1)(f_A N_A + f_a N_a)}\Big| \leq \int_0^{t\wedge S_\varepsilon^K(z)} \frac{f_A N_A}{(\bar{n}_a - \varepsilon M'')K}.$$

Hence decomposition (3.1), Markov's Inequality, and Equations (4.9), (A.8) and (4.10) yield

$$\mathbb{P}\Big(|(P_{a,b_1} - M_a)(t\wedge S_\varepsilon^K(z)) - P_{a,b_1}(0)| > \frac{\varepsilon}{2}\Big) \leq \frac{2f_A\varepsilon^2}{\varepsilon(\bar{n}_a - \varepsilon M'')}\int_0^t e^{S_{Aa}s/2}ds \leq \frac{4f_A\varepsilon}{(\bar{n}_a - \varepsilon M'')|S_{Aa}|}. \tag{4.16}$$

Taking the limit of (4.15) when K goes to infinity and adding (4.16) end the proof. □

4.3. **End of the proof of Theorem 1.** Recall Definitions (2.9) and (4.1). We have:

$$\Big|P_{a,b_1}^{(z,K)}(T_{\text{ext}}^{(z,K)}) - p_{a,b_1}^{(z)}(\infty)\Big| \leq$$
$$\Big|P_{a,b_1}^{(z,K)}(T_{\text{ext}}^{(z,K)}) - P_{a,b_1}^{(z,K)}(t_\varepsilon(z))\Big| + \Big|P_{a,b_1}^{(z,K)}(t_\varepsilon(z)) - p_{a,b_1}^{(z)}(t_\varepsilon(z))\Big| + \Big|p_{a,b_1}^{(z)}(t_\varepsilon(z)) - p_{a,b_1}^{(z)}(\infty)\Big|.$$

To bound the two last terms we use respectively Lemmas 4.1 and 4.2. For the first term of right hand side, (2.5) ensures that with high probability, $N^{(z,K)}(t_\varepsilon(z)) \in \Theta$ and $t_\varepsilon(z) < T_{\text{ext}}^{(z,K)}$. Lemma 4.3, Equation (2.14) and Markov's Inequality allow us to conclude that for ε small enough

$$\limsup_{K\to\infty} \mathbb{P}(\mathbb{1}_{\text{Fix}^{(z,K)}}|P_{a,b_1}^{(z,K)}(T_{\text{ext}}^{(z,K)}) - p_{a,b_1}^{(z)}(\infty)| > 3\varepsilon) \leq c\varepsilon,$$

for a finite c. It is equivalent to the convergence in probability, which concludes the proof.

5. A COUPLING WITH TWO BIRTH AND DEATH PROCESSES

In Sections 6 and 7 we suppose that Assumptions 1 and 3 hold and we denote by N^K the process $N^{(z^{(K)},K)}$. As it will appear in the proof of Theorem 2 the first period of mutant invasion, which ends at time T_ε^K when the mutant population size hits $\lfloor \varepsilon K \rfloor$, is the most important for the neutral proportion dynamics. Indeed, the neutral proportion in the a-population has already reached its final value at time T_ε^K. Let us describe a coupling of the process N_a^K with two birth and death processes which will be a key argument to control the growing of the population a during the first period. To this aim we recall Definition (3.11) and define for $\varepsilon < S_{aA}/(2C_{a,A}C_{A,a}/C_{A,A} + C_{a,a})$,

$$s_-(\varepsilon) := \frac{S_{aA}}{f_a} - \varepsilon\frac{2C_{a,A}C_{A,a} + C_{a,a}C_{A,A}}{f_a C_{A,A}}, \quad \text{and} \quad s_+(\varepsilon) := \frac{S_{aA}}{f_a} + 2\varepsilon\frac{C_{a,A}C_{A,a}}{f_a C_{A,A}}. \tag{5.1}$$

Definitions (2.3) and (2.8) ensure that for $t < T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K$,

$$f_a(1-s_+(\varepsilon)) \le \frac{d_a^K(N^K(t))}{N_a^K(t)} = f_a - S_{aA} + \frac{C_{a,A}}{K}(N_A^K(t) - \bar{n}_A K) + \frac{C_{a,a}}{K} N_a^K(t) \le f_a(1-s_-(\varepsilon)), \tag{5.2}$$

and following Theorem 2 in [5], we can construct on the same probability space the processes Z_ε^-, N^K and Z_ε^+ such that almost surely:

$$Z_\varepsilon^-(t) \le N_a^K(t) \le Z_\varepsilon^+(t), \quad \text{for all } t < \tilde{T}_\varepsilon^K \wedge T_\varepsilon^K, \tag{5.3}$$

where for $* \in \{-,+\}$, Z_ε^* is a birth and death process with initial state 1, and individual birth and death rates f_a and $f_a(1-s_*(\varepsilon))$. We want to prove convergences on the fixation event Fix^K, defined in (2.9). Inequality (A.6) allows us to restrict our attention to the conditional probability measure:

$$\hat{\mathbb{P}}(.) = \mathbb{P}(.|T_\varepsilon^K \le \tilde{T}_\varepsilon^K). \tag{5.4}$$

To study expectations ($\hat{\mathbb{E}}$) and variances ($\hat{\mathrm{Var}}$) associated with this probability measure, we express the event $\{T_\varepsilon^K \le \tilde{T}_\varepsilon^K\}$ in a form easier to handle. More precisely, if we introduce the events:

$$\mathscr{L}_\varepsilon^K := \left\{ f_a(1-s_+(\varepsilon)) \le \frac{d_a^K(N^K(t))}{N_a^K(t)} \le f_a(1-s_-(\varepsilon)), \forall\, t \le T_\varepsilon^K \right\}, \quad \mathscr{H}_\varepsilon^K := \{N_A^K(t) \in I_\varepsilon^K, \forall\, t \le T_\varepsilon^K\}, \tag{5.5}$$

then we can check that

$$\{T_\varepsilon^K \le \tilde{T}_\varepsilon^K\} = \{T_\varepsilon^K < \infty, \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K\}.$$

The term $T_\varepsilon^K < \infty$ is due to the almost sure finiteness of $\tilde{T}_\varepsilon^K$. Indeed the extinction time of a birth and death process with competition is almost surely finite.

6. Proof of Theorem 2 in the strong recombination regime

We distinguish the three periods of the selective sweep: (i) rare mutants and resident population size near its equilibrium value, (ii) quasi-deterministic period governed by the dynamical system (2.4), and (iii) A-population extinction. First we prove that at time T_ε^K proportions of b_1 alleles in the populations A and a are close to z_{Ab_1}/z_A. Once the neutral proportions are the same in the two populations, they do not evolve anymore until the end of the sweep.

Lemma 6.1. *There exist two positive finite constants c and ε_0 such that for $\varepsilon \le \varepsilon_0$:*

$$\limsup_{K\to\infty} \mathbb{E}\left[\mathbb{1}_{T_\varepsilon^K \le \tilde{T}_\varepsilon^K} \left\{ \left| P_{A,b_1}^K(T_\varepsilon^K) - \frac{z_{Ab_1}}{z_A} \right| + \left| P_{A,b_1}^K(T_\varepsilon^K) - P_{a,b_1}^K(T_\varepsilon^K) \right| \right\} \right] \le c\varepsilon.$$

Proof. First we bound the difference between the neutral proportions in the two populations, $|P_{a,b_1}(t) - P_{A,b_1}(t)|$, then we bound $|P_{A,b_1}(t) - z_{Ab_1}/z_A|$. For sake of simplicity we introduce:

$$G(t) := P_{A,b_1}(t) - P_{a,b_1}(t) = \frac{N_{ab_2}(t)N_{Ab_1}(t) - N_{ab_1}(t)N_{Ab_2}(t)}{N_A(t)N_a(t)}, \quad \forall\, t \ge 0, \tag{6.1}$$

$$Y(t) = \mathbb{1}_{\{N_A(t)\ge 1, N_a(t)\ge 1\}} G^2(t) e^{r_K(f_A \wedge f_a)t}, \quad \forall\, t \ge 0.$$

Recalling (3.8) and applying Itô's formula with jumps we get

$$\begin{aligned} Y(t \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K) = Y(0) + \hat{M}_{t\wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K} + r_K \int_0^t \mathbb{1}_{s<T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K} \Big(f_A \wedge f_a - H(s) \Big) Y(s) ds \\ + \int_0^t \mathbb{1}_{s<T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K} e^{r_K(f_A\wedge f_a)s} ds \int_{\mathbb{R}_+} \left[\Big(\mu_A^K(N,s,\theta)\Big)^2 + \Big(\mu_a^K(N,s,\theta)\Big)^2 \right] \mathbb{1}_{\{N_A(s)\ge 1, N_a(s)\ge 1\}} d\theta, \end{aligned}$$

where $\hat{M}$ is a martingale with zero mean, and H is defined by

$$H(t) = \frac{2 f_a f_A N_A(t) N_a(t)}{f_A N_A(t) + f_a N_a(t)} \left[\frac{1}{N_A(t)+1} + \frac{1}{N_a(t)+1} \right], \quad t \ge 0. \tag{6.2}$$

In particular we can check that for all $s \geq 0$ we have $Y(s)H(s) \geq (f_A \wedge f_a)Y(s)$. Equations (3.12) and (3.13) give bounds for the integrals $\int (\mu_\alpha^K)^2$, $\alpha \in \mathscr{A}$, and adding (3.14) we obtain:

$$\begin{aligned}\mathbb{E}[Y(t \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K)] &\leq 1 + \frac{C(\bar{n}_A + 2\varepsilon C_{A,a}/C_{A,A})}{r_K(f_A \wedge f_a)(\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A})K} e^{r_K(f_A \wedge f_a)t} \\ &\quad + \int_0^t (k_0+1)C\Big(\bar{n}_A + 2\varepsilon \frac{C_{A,a}}{C_{A,A}}\Big)\mathbb{E}\Big[\tilde{M}_{s \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K}\Big] e^{(r_K(f_A \wedge f_a) - \frac{S_{aA}}{2(k_0+1)})s} ds \\ &\leq c\Big(1 + \frac{1}{Kr_K} e^{r_K(f_A \wedge f_a)t} + e^{(r_K(f_A \wedge f_a) - \frac{S_{aA}}{2(k_0+1)})t}\Big), \qquad (6.3)\end{aligned}$$

where c is a finite constant which can be chosen independently of ε and K if ε is small enough and K large enough. Combining Equation (3.1), Cauchy-Schwarz Inequality, and Equations (3.12) and (6.3) we get for every $t \geq 0$,

$$\begin{aligned}&\mathbb{E}\Big[\Big|P_{A,b_1}(t \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K) - \frac{\lfloor z_{Ab_1} K \rfloor}{\lfloor z_A K \rfloor}\Big|\Big] \\ &\quad \leq \mathbb{E}\Big[|M_A(t \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K)|\Big] + \frac{r_K f_a \varepsilon}{\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A}} \int_0^t \mathbb{E}\Big[\mathbb{1}_{s < T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K} |G(s)| \mathbb{1}_{\{N_A(s) \geq 1, N_a(s) \geq 1\}}\Big] ds \\ &\quad \leq \mathbb{E}^{1/2}\Big[\langle M_A \rangle_{t \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K}\Big] + c r_K \varepsilon \int_0^t \mathbb{E}^{1/2}\Big[Y(s \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K)\Big] e^{-r_K(f_A \wedge f_a)s/2} ds \\ &\quad \leq c\Big(\sqrt{t/K} + \varepsilon \int_0^t \Big(r_K e^{-r_K(f_A \wedge f_a)s} + \frac{1}{K} + e^{-S_{aA}s/2(k_0+1)}\Big)^{1/2} ds\Big),\end{aligned}$$

where c is finite. A simple integration then yields the existence of a finite c such that:

$$\mathbb{E}\Big[\Big|P_{A,b_1}(t \wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K) - \frac{\lfloor z_{Ab_1} K \rfloor}{\lfloor z_A K \rfloor}\Big|\Big] \leq c\Big(\sqrt{t/K} + \varepsilon\Big(1 + \frac{t}{\sqrt{K}}\Big)\Big). \qquad (6.4)$$

But according to Coupling (5.3) and limit (A.11) we have the asymptotic behavior

$$\lim_{K \to \infty} \mathbb{P}(T_\varepsilon^K > 2S_{aA} \log K | T_\varepsilon^K \leq \tilde{T}_\varepsilon^K) = 0, \qquad (6.5)$$

Hence applying (6.4) at time $t = 2S_{aA} \log K$ and using (A.3) and (6.5), we bound the first term in theexpectation. The second bound is obtained in the same way. □

The following Lemma states that during the second period, the neutral proportion stays constant in the a-population.

Lemma 6.2. *There exist two positive finite constants c and ε_0 such that for $\varepsilon \leq \varepsilon_0$:*

$$\limsup_{K \to \infty} \mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K \leq \tilde{T}_\varepsilon^K} \Big| P_{a,b_1}(T_\varepsilon^K + t_\varepsilon(\tfrac{N(T_\varepsilon^K)}{K})) - \tfrac{z_{Ab_1}}{z_A}\Big|\Big] \leq c\varepsilon.$$

Proof. Let us introduce, for $z \in \mathbb{R}_+^{\mathscr{E}}$ and $\varepsilon > 0$ the set Γ and the time t_ε defined as follows:

$$\Gamma := \Big\{z \in \mathbb{R}_+^{\mathscr{E}}, z_A \in \Big[\bar{n}_A - 2\varepsilon \frac{C_{A,a}}{C_{A,A}}, \bar{n}_A + 2\varepsilon \frac{C_{A,a}}{C_{A,A}}\Big], z_a \in \Big[\frac{\varepsilon}{2}, 3\frac{\varepsilon}{2}\Big]\Big\}, \quad t_\varepsilon := \sup\{t_\varepsilon(z), z \in \Gamma\}, \qquad (6.6)$$

where $t_\varepsilon(z)$ has been defined in (4.1). According to Assumption 1, $t_\varepsilon < \infty$, and

$$I(\Gamma, \varepsilon) := \inf_{z \in \Gamma} \inf_{t \leq t_\varepsilon} \{n_A^{(z)}(t), n_a^{(z)}(t)\} > 0,$$

and we can introduce the stopping time

$$L_\varepsilon^K(z) = \inf\big\{t \geq 0, (N_A^{(z,K)}(t), N_a^{(z,K)}(t)) \notin [I(\Gamma, \varepsilon)K/2, (\bar{n}_A + \bar{n}_a)K]^2\big\}. \qquad (6.7)$$

Finally, we denote by $(\mathscr{F}_t^K, t \geq 0)$ the canonical filtration of N^K. Notice that on the event $\{T_\varepsilon^K \leq \tilde{T}_\varepsilon^K\}$, $N(T_\varepsilon^K)/K \in \Gamma$, thus $t_\varepsilon(N(T_\varepsilon^K)/K) \leq t_\varepsilon$. The semi-martingale decomposition (3.1) and the

definition of G in (6.1) then twice the Strong Markov property and the Cauchy-Schwarz Inequality yield:

$$\begin{aligned}(6.8)\quad &\mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}\Big|P_{a,b_1}\Big(T_\varepsilon^K+t_\varepsilon(\tfrac{N(T_\varepsilon^K)}{K})\wedge L_\varepsilon^K(\tfrac{N(T_\varepsilon^K)}{K})\Big)-P_{a,b_1}(T_\varepsilon^K)\Big|\Big]\\ &\le\mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}\mathbb{E}\Big[\Big|M_a\Big(T_\varepsilon^K+t_\varepsilon(\tfrac{N(T_\varepsilon^K)}{K})\wedge L_\varepsilon^K(\tfrac{N(T_\varepsilon^K)}{K})-M_a(T_\varepsilon^K)\Big|+f_a\int_{T_\varepsilon^K}^{T_\varepsilon^K+t_\varepsilon\wedge L_\varepsilon^K(\frac{N(T_\varepsilon^K)}{K})}|G|\Big|\mathscr{F}_{T_\varepsilon^K}\Big]\Big]\\ &\le\mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}\Big\{\mathbb{E}^{1/2}\Big[\langle M_a\rangle_{T_\varepsilon^K+t_\varepsilon\wedge L_\varepsilon^K(\frac{N(T_\varepsilon^K)}{K})}-\langle M_a\rangle_{T_\varepsilon^K}\Big|\mathscr{F}_{T_\varepsilon^K}\Big]+f_a\sqrt{t_\varepsilon}\,\mathbb{E}^{1/2}\Big[\int_{T_\varepsilon^K}^{T_\varepsilon^K+t_\varepsilon\wedge L_\varepsilon^K(\frac{N(T_\varepsilon^K)}{K})}G^2\Big|\mathscr{F}_{T_\varepsilon^K}\Big]\Big\}\Big].\end{aligned}$$

To bound the first term of the right hand side we use Strong Markov property, Equation (3.12) and the definition of L_ε^K in (6.7). We get

$$(6.9)\qquad \mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}\mathbb{E}^{1/2}\Big[\langle M_a\rangle_{T_\varepsilon^K+t_\varepsilon\wedge L_\varepsilon^K(\frac{N(T_\varepsilon^K)}{K})}-\langle M_a\rangle_{T_\varepsilon^K}\Big|\mathscr{F}_{T_\varepsilon^K}\Big]\Big]\le\frac{2t_\varepsilon C(\bar n_A+\bar n_a)}{I(\Gamma,\varepsilon)K}.$$

For the second term, Strong Markov property and Itô's formula with jumps yield

$$\begin{aligned}&\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}\mathbb{E}^{1/2}\Big[\int_{T_\varepsilon^K}^{T_\varepsilon^K+t_\varepsilon\wedge L_\varepsilon^K(\frac{N(T_\varepsilon^K)}{K})}G^2\Big|\mathscr{F}_{T_\varepsilon^K}\Big]\\ &\qquad\le\sqrt{t_\varepsilon}\sup_{z\in\Gamma,t\le t_\varepsilon}\mathbb{E}\Big[|G^2(t\wedge L_\varepsilon^K(z))-G^2(0)|\Big|N(0)=\lfloor zK\rfloor\Big]+\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}|G(T_\varepsilon^K)|\\ &\qquad\qquad\le 2\sqrt{t_\varepsilon}\sum_{\alpha\in\mathscr{A}}\int_0^{t_\varepsilon}\mathbb{E}\Big[\mathbb{1}_{s<L_\varepsilon^K(z)}\frac{d}{ds}\langle M_\alpha\rangle_s\Big]ds+\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}|G(T_\varepsilon^K)|.\end{aligned}$$

Equations (3.12) and Lemma 6.1 finally lead to

$$(6.10)\qquad \mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K\le\tilde T_\varepsilon^K}\mathbb{E}^{1/2}\Big[\int_{T_\varepsilon^K}^{T_\varepsilon^K+t_\varepsilon\wedge L_\varepsilon^K(\frac{N(T_\varepsilon^K)}{K})}G^2\Big|\mathscr{F}_{T_\varepsilon^K}\Big]\Big]\le\sqrt{t_\varepsilon}\Big(\frac{8t_\varepsilon C(\bar n_A+\bar n_a)}{I(\Gamma,\varepsilon)K}\Big)^{1/2}+o_K(1)$$

where $o_K(1)$ denotes a function of K going to 0 at infinity. Moreover (2.5) ensures that

$$\mathbb{P}\Big(T_\varepsilon^K\le\tilde T_\varepsilon^K,L_\varepsilon^K(\tfrac{N(T_\varepsilon^K)}{K})\le t_\varepsilon(\tfrac{N(T_\varepsilon^K)}{K})\Big)\le\mathbb{P}\Big(\tfrac{N(T_\varepsilon^K)}{K}\in\Theta,L_\varepsilon^K(\tfrac{N(T_\varepsilon^K)}{K})\le t_\varepsilon(\tfrac{N(T_\varepsilon^K)}{K})\Big)\underset{K\to\infty}{\to}0,$$

where Θ has been defined in (4.7). Adding Equations (6.8), (6.9), (6.10) and Lemma 6.1, we finally end the proof of Lemma 6.2. □

Proof of Theorem 2 in the strong recombination regime. Let us focus on the A-population extinction period. We have thanks to Strong Markov property:

$$\begin{aligned}(6.11)\quad &\mathbb{P}\Big(\mathbb{1}_{N(T_\varepsilon^K+t_\varepsilon(N(T_\varepsilon^K)/K))\in\Theta}\Big|P_{a,b_1}(T_{\text{ext}})-P_{a,b_1}(T_\varepsilon^K+t_\varepsilon(\tfrac{N(T_\varepsilon^K)}{K}))\Big|>\sqrt{\varepsilon}\Big)\\ &\qquad\le\sup_{z\in\Theta}\mathbb{P}\Big(|P_{a,b_1}(T_{\text{ext}})-P_{a,b_1}(0)|>\sqrt{\varepsilon}\Big|N(0)=\lfloor zK\rfloor\Big).\end{aligned}$$

But Equation (2.5) yields $\mathbb{P}(N(T_\varepsilon^K+t_\varepsilon(N(T_\varepsilon^K)/K))/K\in\Theta|N(T_\varepsilon^K)/K\in\Gamma)\to_{K\to\infty}1$, and $\{T_\varepsilon^K\le\tilde T_\varepsilon^K\}\subset\{N(T_\varepsilon^K)/K\in\Gamma\}$. Adding Equation (A.6) and Lemma 6.2, Triangle inequality allows us to conclude that for ε small enough

$$\limsup_{K\to\infty}\ \mathbb{P}\Big(\Big|P_{a,b_1}^K(T_{\text{ext}}^K)-\frac{z_{Ab_1}}{z_A}\Big|>\sqrt{\varepsilon}\Big|\text{Fix}^K\Big)\le c\varepsilon.$$

As $\mathbb{P}(\text{Fix}^K)\to_{K\to\infty}S_{aA}/f_a>0$, it is equivalent to the claim of Theorem 2 in the strong regime. □

7. Proof of Theorem 2 in the weak recombination regime

In this section we suppose that Assumptions 3 and 5 hold. We first focus on the neutral proportion in the a population at time T_ε^K. The idea is to consider the neutral alleles at time T_ε^K and follow their ancestral lines back until the beginning of the sweep, to know whether they are descended from the first mutant or not. Two kinds of event can happen to a neutral lineage: coalescences and m-recombinations (see Section 7.1); we show that we can neglect the coalescences and the occurrence of several m-recombinations for a lineage during the first period. Therefore, our approximation of the genealogy is the following: two neutral lineages are independent, and each of them undergoes one recombination with an A-individual during the first period with probability $1-\exp(-r_K f_a \log K/S_{aA})$. If it has undergone a recombination with an A-individual, it can be an allele b_1 or b_2. Otherwise it is descended from the first mutant and is an allele b_1. To get this approximation we follow the line presented by Schweinsberg and Durrett in [28]. In this paper, the authors describe the population dynamics by a variation of Moran model with two loci and recombinations. In their model, the population size is constant and each individual has a constant selective advantage, 0 or s. In our model the size is varying and the individual's ability to survive depends on the population state. After the study of the first period we check that the second and third periods have little influence on the neutral proportion in the a-population.

7.1. **Coalescence and m-recombination times.** Let us introduce the jump times of the stopped Markov process $(N^K(t), t \le T_\varepsilon^K)$, $0=:\tau_0^K<\tau_1^K<\ldots<\tau_{J^K}^K:=T_\varepsilon^K$, where J^K denotes the jump number of N^K between 0 and T_ε^K, and the time of the m-th jump is:

$$\tau_m^K=\inf\left\{t>\tau_{m-1}^K, N^K(t)\neq N^K(\tau_{m-1}^K)\right\},\ 1\le m\le J^K.$$

Let us sample two individuals uniformly at random at time T_ε^K and denote by β_p and β_q their neutral alleles. We want to follow their genealogy backward in time and know at each time between 0 and T_ε^K the types (A or a) of the individuals carrying β_p and β_q.

We say that β_p and β_q coalesce at time τ_m^K if they are carried by two different individuals at time τ_m^K and by the same individual at time τ_{m-1}^K. In other words the individual carrying the allele β_p (or β_q) at time τ_m^K is a newborn and has inherited his neutral allele from the individual carrying allele β_q (or β_p) at time τ_{m-1}^K. The jump number at the coalescence time is denoted by

$$TC^K(\beta_p,\beta_q):=\begin{cases}\sup\{m\le J^K, \beta_p \text{ and } \beta_q \text{ coalesce at time } \tau_m^K\}, & \text{if } \beta_p \text{ and } \beta_q \text{ coalesce}\\ -\infty, & \text{otherwise.}\end{cases}$$

We say that β_p m-recombines at time τ_m^K if the individual carrying the allele β_p at time τ_m^K is a newborn, carries the allele $\alpha\in\mathscr{A}$, and has inherited his allele β_p from an individual carrying allele $\bar{\alpha}$. In other words, a m-recombination is a recombination which modifies the selected allele connected to the neutral allele. The jump numbers of the first and second (backward in time) m-recombinations are denoted by:

$$TR_1^K(\beta_p):=\begin{cases}\sup\{m\le J^K, \beta_p \text{ m-recombines at time } \tau_m^K\}, & \text{if there is at least one m-recombination}\\ -\infty, & \text{otherwise,}\end{cases}$$

$$TR_2^K(\beta_p):=\begin{cases}\sup\{m<TR_1^K(\beta_p), \beta_p \text{ m-recombines at time } \tau_m^K\}, & \text{if there are at least two}\\ & \text{m-recombinations}\\ -\infty, & \text{otherwise.}\end{cases}$$

Let us now focus on the probability for a coalescence to occur conditionally on the state of the process (N_A,N_a) at two successive jump times. We denote by $p_{\alpha_1\alpha_2}^{c_K}(n)$ the probability that the genealogies of two random neutral alleles associated respectively with alleles α_1 and $\alpha_2\in\mathscr{A}$ at time τ_m^K coalesce at this time conditionally on $(N_A^K(\tau_{m-1}^K),N_a^K(\tau_{m-1}^K))=n\in\mathbb{N}^2$ and on the birth of an individual carrying allele $\alpha_1\in\mathscr{A}$ at time τ_m^K. Then we have the following result:

Lemma 7.1. *For every $n=(n_A,n_a)\in\mathbb{N}^2$ and $\alpha\in\mathscr{A}$, we have:*

$$p^{c_K}_{\alpha\alpha}(n)=\frac{2}{n_\alpha(n_\alpha+1)}\Big(1-\frac{r_K f_{\bar\alpha} n_{\bar\alpha}}{f_A n_A+f_a n_a}\Big)\quad\text{and}\quad p^{c_K}_{\alpha\bar\alpha}(n)=\frac{r_K f_{\bar\alpha}}{(n_\alpha+1)(f_A n_A+f_a n_a)}. \tag{7.1}$$

Proof. We only state the expression of $p^{c_K}_{\alpha\alpha}(n)$, as the calculations are similar for $p^{c_K}_{\alpha\bar\alpha}(n)$. If there is a m-recombination, we cannot have the coalescence of two neutral alleles associated with allele α at time τ^K_m. With probability $1-r_K f_{\bar\alpha} n_{\bar\alpha}/(f_A n_A+f_a n_a)$ there is no m-recombination and the parent giving its neutral allele carries the allele α. When there is no m-recombination, two individuals among those, who carry allele α also carry a neutral allele which was in the same individual at time τ^K_{m-1}. We have a probability $2/n_\alpha(n_\alpha+1)$ to pick this couple of individuals among the $(n_\alpha+1)$ α-individuals. □

Remark 4. *A m-recombination for a neutral allele associated with an α allele is a coalescence with an $\bar\alpha$ individual. Thus if we denote by $p^{r_K}_\alpha(n)$ the probability that an α-individual, chosen at random at time τ^K_m, is the newborn and underwent a m-recombination at his birth, conditionally on $(N^K_A(\tau^K_{m-1}),N^K_a(\tau^K_{m-1}))=n\in\mathbb{N}^2$ and on the birth of an individual α at time τ^K_m we get*

$$p^{r_K}_\alpha(n)=n_{\bar\alpha}p^{c_K}_{\alpha\bar\alpha}(n)=\frac{n_{\bar\alpha}r_K f_{\bar\alpha}}{(n_\alpha+1)(f_A n_A+f_a n_a)}. \tag{7.2}$$

Moreover, if we recall the definition of I^K_ε in (3.10)*, we notice that there exists a finite constant c such that for $k<\lfloor\varepsilon K\rfloor$,*

$$(1-c\varepsilon)\frac{r_K}{k+1}\le\inf_{n_A\in I^K_\varepsilon}p^{r_K}_a(n_A,k)\le\sup_{n_A\in I^K_\varepsilon}p^{r_K}_a(n_A,k)\le\frac{r_K}{k+1}. \tag{7.3}$$

7.2. **Jumps of mutant population during the first period.** We want to count the number of coalescences and m-recombinations in the lineages of the two randomly chosen neutral alleles β_p and β_q. By definition, these events can only occur at a birth time. Thus we need to study the upcrossing number of the process N^K_a before T^K_ε (Lemma 7.2). It allows us to prove that the probability that a lineage is affected by two m-recombinations or that two lineages coalesce, and then (backward in time) are affected by a m-recombination are negligible (Lemma 7.3). Then we obtain an approximation of the probability that a lineage is affected by a m-recombination (Lemma 7.4), and finally we check that two lineages are approximately independent (Equation (7.20)). The last step consists in controlling the neutral proportion in the population A (Lemma 7.5). Indeed it will give us the probability that a neutral allele which has undergone a m-recombination is a b_1 or a b_2.

Let us denote by ζ^K_k the jump number of last visit to k before the hitting of $\lfloor\varepsilon K\rfloor$,

$$\zeta^K_k:=\sup\{m\le J^K,N^K_a(\tau^K_m)=k\},\quad 1\le k\le\lfloor\varepsilon K\rfloor. \tag{7.4}$$

This allows us to introduce for $0<j\le k<\lfloor\varepsilon K\rfloor$ the number of upcrossings from k to $k+1$ for the process N^K_a before and after the last visit to j:

$$U^{(K,1)}_{j,k}:=\#\{m\in\{0,...,\zeta^K_j-1\},(N^K_a(\tau^K_m),N^K_a(\tau^K_{m+1}))=(k,k+1)\}, \tag{7.5}$$

$$U^{(K,2)}_{j,k}:=\#\{m\in\{\zeta^K_j,...,J^K-1\},(N^K_a(\tau^K_m),N^K_a(\tau^K_{m+1}))=(k,k+1)\}. \tag{7.6}$$

We also introduce the number of jumps of the A-population size when there are k a-individuals and the total number of upcrossings from k to $k+1$ before T^K_ε:

$$H^K_k:=\#\{m<J^K,N^K_a(\tau^K_m)=N^K_a(\tau^K_{m+1})=k\}, \tag{7.7}$$

$$U^K_k:=U^{(K,1)}_{j,k}+U^{(K,2)}_{j,k}=\#\{m<J^K,(N^K_a(\tau^K_m),N^K_a(\tau^K_{m+1}))=(k,k+1)\}. \tag{7.8}$$

The next Lemma states moment properties of these jump numbers. Recall Definition (5.1). Then

Lemma 7.2. *There exist two positive and finite constants ε_0 and c such that for $\varepsilon \le \varepsilon_0$, K large enough and $1 \le j \le k < \lfloor \varepsilon K \rfloor$, we have*

$$\hat{\mathbb{E}}[H_j^K] \le \frac{12 f_A \bar{n}_A K}{s_-^4(\varepsilon) f_a j}, \quad \lambda_\varepsilon := \frac{(1-s_-(\varepsilon))^3}{(1-s_+(\varepsilon))^2} \in (0,1), \quad \hat{\mathbb{E}}[(U_{j,k}^{(K,1)})^2] \le \frac{4\lambda_\varepsilon^{k-j}}{s_-^7(\varepsilon)(1-s_+(\varepsilon))}, \tag{7.9}$$

$$\hat{\mathbb{E}}[(U_j^K)^2] \le \frac{2}{s_-^2(\varepsilon)}, \quad \left|\hat{\mathrm{Cov}}(U_{j,k}^{(K,2)}, U_j^K)\right| \le c(\varepsilon + (1-s_-(\varepsilon))^{k-j}), \tag{7.10}$$

and

$$r_K \Big| \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{\hat{\mathbb{E}}[U_k^K]}{k+1} - \frac{f_a \log K}{S_{aA}} \Big| \le c\varepsilon. \tag{7.11}$$

This Lemma is widely used in Sections 7.3 and 7.4. Indeed, we shall decompose on the possible states of the population when a birth occurs, and apply Equations (7.1) and (7.2) to express the probability of coalescences and m-recombinations at each birth event. The proof of Lemma 7.2 is quite technical and is postponed to Appendix B.

7.3. **Negligible events.** The next Lemma bounds the probability that two m-recombinations occur in a neutral lineage and the probability that a couple of neutral lineages coalesce and then m-recombine when we consider the genealogy backward in time.

Lemma 7.3. *There exist two positive finite constants c and ε_0 such that for $K \in \mathbb{N}$ and $\varepsilon \le \varepsilon_0$,*

$$\hat{\mathbb{P}}\Big(TR_2^K(\beta_p) \ne -\infty\Big) \le \frac{c}{\log K}, \quad \textit{and} \quad \hat{\mathbb{P}}\Big(0 \le TR_1^K(\beta_p) \le TC^K(\beta_p, \beta_q)\Big) \le \frac{c}{\log K}.$$

Proof. By definition, the neutral allele β_p is associated with an allele a at time T_ε^K. If there are at least two m-recombinations it implies that there exists a time between 0 and T_ε^K at which β_p has undergone a m-recombination when it was associated with an allele A. We shall work conditionally on the stopped process $((N_A(\tau_m^K), N_a(\tau_m^K)), m \le J^K)$ and decompose according to the a-population size when this m-recombination occurs. We get the inclusion:

$$\{TR_2^K(\beta_p) \ne -\infty\} \subset \bigcup_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \bigcup_{m=1}^{J^K} \Big\{ TR_2^K(\beta_p) = m, N_a(\tau_{m-1}^K) = N_a(\tau_m^K) = k \Big\}.$$

We recall the definition of I_ε^K in (3.10). Thanks to Equations (7.2) and (7.9), we get:

$$\hat{\mathbb{P}}(TR_2^K(\beta_p) \ne -\infty) \le \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \sup_{n_A \in I_\varepsilon^K} p_A^{r_K}(n_A, k) \hat{\mathbb{E}}[H_k^K] \le \frac{12 r_K \bar{n}_A \varepsilon}{s_-^4(\varepsilon)(\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A})^2}.$$

The Assumption 5 on weak recombination completes the proof of the first inequality in Lemma 7.3. The proof of the second one is divided in two steps, presented after introducing the notations

$$(\alpha\beta_p)_m := \{\text{the neutral allele } \beta_p \text{ is associated with the allele } \alpha \text{ at time } \tau_m^K\}, \quad \alpha \in \mathscr{A}, \quad m \le J^K,$$

$$(\alpha_1\beta_p, \alpha_2\beta_q)_m := (\alpha_1\beta_p)_m \cap (\alpha_2\beta_q)_m, \quad (\alpha_1, \alpha_2) \in \mathscr{A}^2, \quad m \le J^K.$$

First step: We show that the probability that β_p is associated with an allele A at the coalescence time is negligible. We first recall the inclusion,

$$\{TC^K(\beta_p, \beta_q) \ne -\infty, (A\beta_p)_{TC^K(\beta_p,\beta_q)}\} \subset \bigcup_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \bigcup_{m=1}^{J^K} \Big\{ TC^K(\beta_p, \beta_q) = m, N_a(\tau_{m-1}^K) = k, (A\beta_p)_m \Big\},$$

and decompose on the possible selected alleles associated with β_q and on the type of the newborn at coalescence time. Using Lemma 7.1, Equations (7.9) and (7.10), and $r_K \le 1$, we get

$$\begin{aligned}(7.12)\quad &\hat{\mathbb{P}}(TC^K(\beta_p,\beta_q) \neq -\infty, (A\beta_p)_{TC^K(\beta_p,\beta_q)}) \\ &\le \sum_{k=1}^{\lfloor \varepsilon K\rfloor -1}\Big[\sup_{n_A\in I_\varepsilon^K} p_{AA}^{c_K}(n_A,k) + \sup_{n_A\in I_\varepsilon^K} p_{Aa}^{c_K}(n_A,k)\Big]\hat{\mathbb{E}}[H_k^K] + \sup_{n_A\in I_\varepsilon^K} p_{aA}^{c_K}(n_A,k)\hat{\mathbb{E}}[U_k^K] \le \frac{c}{K}\sum_{k=1}^{\lfloor \varepsilon K\rfloor -1}\frac{1}{k},\end{aligned}$$

for a finite c, which is of order $\log K/K$.

Second step: Then, we focus on the case where β_p and β_q are associated with an allele a at coalescence time. The inclusion

$$\{N_a(\tau^K_{TC^K(\beta_p,\beta_q)-1}) = k, (a\beta_p, a\beta_q)_{TC^K(\beta_p,\beta_q)}\} \subset \bigcup_{m=1}^{J^K}\Big\{TC^K(\beta_p,\beta_q) = m, N_a(\tau^K_{m-1}) = k, (a\beta_p, a\beta_q)_m\Big\},$$

and Equations (7.1) and (7.10) yield for every $k \in \{1,...,\lfloor \varepsilon K\rfloor -1\}$:

$$\hat{\mathbb{P}}(N_a(\tau^K_{TC^K(\beta_p,\beta_q)-1}) = k, (a\beta_p, a\beta_q)_{TC^K(\beta_p,\beta_q)}) \le \sup_{n_A\in I_\varepsilon^K} p_{aa}^{c_K}(n_A,k)\hat{\mathbb{E}}[U_k^K] \le \frac{4}{s_-^2(\varepsilon)k(k+1)}.$$

If β_p and β_q coalesce then undergo their first m-recombination when we look backward in time, and if the a-population has the size k at the coalescence time, it implies that the m-recombination occurs before the ζ_k^K-th jump when we look forward in time. For $k,l < \lfloor \varepsilon K\rfloor$,

$$\begin{aligned}&\hat{\mathbb{P}}\Big(N_a\Big(\tau^K_{TR_1^K(\beta_p)}\Big) = l, 0 \le TR_1^K(\beta_p) \le TC^K(\beta_p,\beta_q)\Big| N_a\Big(\tau^K_{TC^K(\beta_p,\beta_q)-1}\Big) = k, (a\beta_p, a\beta_q)_{TC^K(\beta_p,\beta_q)}\Big) \\ &\qquad \le \sup_{n_A\in I_\varepsilon^K} p_a^{r_K}(n_A,l)\Big(\mathbb{1}_{k>l}\hat{\mathbb{E}}[U_l^K] + \mathbb{1}_{k\le l}\hat{\mathbb{E}}[U_{k,l}^{(K,1)}]\Big) \le \frac{2r_K}{(l+1)s_-^2(\varepsilon)}\Big(\mathbb{1}_{k>l} + \frac{2\mathbb{1}_{k\le l}\lambda_\varepsilon^{l-k}}{s_-^5(\varepsilon)(1-s_+(\varepsilon))}\Big),\end{aligned}$$

where the last inequality is a consequence of (7.3), (7.9) and (7.10). The two last equations finally yield the existence of a finite c such that for every $K \in \mathbb{N}$:

$$\hat{\mathbb{P}}(0 \le TR_1^K(\beta_p) \le TC^K(\beta_p,\beta_q), (a\beta_p, a\beta_q)_{TC^K(\beta_p,\beta_q)}) \le cr_K\sum_{k,l=1}^{\lfloor \varepsilon K\rfloor}\frac{\mathbb{1}_{k>l} + \mathbb{1}_{k\le l}\lambda_\varepsilon^{l-k}}{k(k+1)(l+1)} \le cr_K,$$

which ends up the proof of Lemma 7.3 with Assumption 5. □

7.4. **Probability to be descended from the first mutant.** We want to estimate the probability for the neutral lineage of β_p to undergo no m-recombination.

Lemma 7.4. *There exist two positive finite constants c and ε_0 such that for $\varepsilon \le \varepsilon_0$:*

$$\limsup_{K\to\infty}\Big|\hat{\mathbb{P}}(TR_1^K(\beta_p) = -\infty) - \exp\Big(-\frac{f_a r_K \log K}{S_{aA}}\Big)\Big| \le c\varepsilon.$$

Proof. We introduce ρ_m^K, the conditional probability that the neutral lineage of β_p m-recombines at time τ_m^K, given $(N_A(\tau_n^K), N_a(\tau_n^K), n \le J^K)$ and given that it has not m-recombined during the time interval $]\tau_m^K, T_\varepsilon^K]$. The last condition implies that β_p is associated with an allele a at time τ_m^K.

$$\rho_m^K := \mathbb{1}_{\{N_a^K(\tau_m^K) - N_a^K(\tau_{m-1}^K) = 1\}} p_a^{r_K}(N_A^K(\tau_{m-1}^K), N_a^K(\tau_{m-1}^K)).$$

We also introduce η^K, the sum of these conditional probabilities for $1 \le m \le J^K$:

$$\eta^K := \sum_{m=1}^{J^K}\rho_m^K.$$

We want to give a rigourous meaning to the sequence of equivalents:

$$\hat{\mathbb{P}}\Big(TR_1^K(\beta_p) = -\infty\Big|(N_A(\tau_m^K), N_a(\tau_m^K))_{m\le J^K}\Big) = \prod_{m=1}^{J^K}(1-\rho_m^K) \sim \prod_{m=1}^{J^K}\exp(-\rho_m^K) \sim \exp(-\mathbb{E}[\eta^K]),$$

when K goes to infinity. Jensen and Triangle inequalities and the Mean Value Theorem imply

$$\hat{\mathbb{E}}\left|\hat{\mathbb{P}}\Big(TR_1^K(\beta_p)=-\infty\Big|(N_A(\tau_m^K),N_a(\tau_m^K))_{m\le J^K}\Big)-e^{-\frac{far_K\log K}{S_{aA}}}\right|\le \tag{7.13}$$
$$\hat{\mathbb{E}}\left|\hat{\mathbb{P}}\Big(TR_1^K(\beta_p)=-\infty\Big|(N_A(\tau_m^K),N_a(\tau_m^K))_{m\le J^K}\Big)-e^{-\eta^K}\right|+\left|e^{-\hat{\mathbb{E}}\eta^K}-e^{-\frac{far_K\log K}{S_{aA}}}\right|+\hat{\mathbb{E}}\left|\eta^K-\hat{\mathbb{E}}\eta^K\right|.$$

We aim to bound the right hand side of (7.13). The bounding of the first term follows the method developed in Lemma 3.6 in [28]. We refer to this proof, and get

$$\hat{\mathbb{E}}\left|\hat{\mathbb{P}}\Big(TR_1^K(\beta_p)=-\infty\Big|(N_A(\tau_m^K),N_a(\tau_m^K))_{m\le J^K}\Big)-e^{-\eta^K}\right|\le\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\sup_{n_A\in I_\varepsilon^K}\Big(p_a^{r_K}(n_A,k)\Big)^2\hat{\mathbb{E}}[U_k^K]\le\frac{\pi^2r_K^2}{3s_-^2(\varepsilon)}, \tag{7.14}$$

where I_ε^K has been defined in (3.10) and the last inequality follows from (7.3) and (7.10). To bound the second term, we need to estimate $\hat{\mathbb{E}}[\eta^K]$. Inequality (7.3) implies

$$(1-c\varepsilon)r_K\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{U_k^K}{k+1}\le\eta^K\le r_K\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{U_k^K}{k+1}. \tag{7.15}$$

Adding (7.11) we get that for ε small enough,

$$\limsup_{K\to\infty}\left|\exp(-\hat{\mathbb{E}}[\eta^K])-\exp\Big(-\frac{r_Kf_a\log K}{S_{aA}}\Big)\right|\le c\varepsilon. \tag{7.16}$$

The bounding of the last term of (7.13) requires a fine study of dependences between upcrossing numbers before and after the last visit to a given integer by the mutant population size. In particular, we widely use Equation (7.10). We observe that $\hat{\mathbb{E}}[|\eta^K-\hat{\mathbb{E}}\eta^K|]\le(\hat{\mathrm{Var}}\,\eta^K)^{1/2}$, but the variance of η^K is quite involved to study and according to Assumption 5 and Equations (7.15) and (7.10),

$$\begin{aligned}\left|\hat{\mathrm{Var}}\,\eta^K-\hat{\mathrm{Var}}\left(r_K\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{U_k^K}{k+1}\right)\right| &\le c\varepsilon\hat{\mathbb{E}}\Big[\Big(r_K\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{U_k^K}{k+1}\Big)^2\Big]\\ &\le c\varepsilon r_K^2\sum_{k,l=1}^{\lfloor\varepsilon K\rfloor-1}\frac{\hat{\mathbb{E}}[(U_k^K)^2]+\hat{\mathbb{E}}[(U_l^K)^2]}{(k+1)(l+1)}\le c\varepsilon,\end{aligned} \tag{7.17}$$

for a finite c. Let $k\le l<\lfloor\varepsilon K\rfloor$, and recall that by definition, $U_l^K=U_{k,l}^{(K,1)}+U_{k,l}^{(K,2)}$. Then we have

$$\left|\hat{\mathrm{Cov}}(U_k^K,U_l^K)\right|\le\Big(\hat{\mathbb{E}}[(U_k^K)^2]\hat{\mathbb{E}}[(U_{k,l}^{(K,1)})^2]\Big)^{1/2}+\left|\hat{\mathrm{Cov}}(U_k^K,U_{k,l}^{(K,2)})\right|. \tag{7.18}$$

Applying Inequalities (7.9) and (7.10) and noticing that $(1-s_-(\varepsilon))<\lambda_\varepsilon^{1/2}<1$ (see proof of Lemma 7.2 in Appendix B) lead to

$$\left|\hat{\mathrm{Cov}}(U_k^K,U_l^K)\right|\le c(\lambda_\varepsilon^{(l-k)/2}+\varepsilon+(1-s_-(\varepsilon))^{l-k})\le c(\lambda_\varepsilon^{(l-k)/2}+\varepsilon)$$

for a finite c. We finally get:

$$\begin{aligned}\hat{\mathrm{Var}}\Big(r_K\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{U_k^K}{k+1}\Big) &\le 2r_K^2\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{1}{k+1}\sum_{l=k}^{\lfloor\varepsilon K\rfloor-1}\frac{\hat{\mathrm{Cov}}(U_k^K,U_l^K)}{l+1}\\ &\le cr_K^2\sum_{k=1}^{\lfloor\varepsilon K\rfloor-1}\frac{1}{k+1}\sum_{l=k}^{\lfloor\varepsilon K\rfloor-1}\frac{\lambda_\varepsilon^{(l-k)/2}+\varepsilon}{l+1}\le cr_K^2\varepsilon\log^2K,\end{aligned} \tag{7.19}$$

where we used (7.18) for the second inequality. Applying Jensen's Inequality to the left hand side of (7.13) and adding Equations (7.14), (7.16), (7.17) and (7.19) complete the proof of Lemma 7.4. □

We finally focus on the dependence between genealogies of β_p and β_q. Following [28] pp. 1622 to 1624 in the case $J=1$, we can prove that for d in $\{0,1,2\}$,

$$\limsup_{K\to\infty}\left|\hat{\mathbb{P}}(\mathbb{1}_{TR_1^K(\beta_p)\ge0}+\mathbb{1}_{TR_1^K(\beta_p)\ge0}=d)-\binom{2}{d}\Big(1-e^{\frac{-far_K\log K}{S_{aA}}}\Big)^d\Big(e^{\frac{-far_K\log K}{S_{aA}}}\Big)^{2-d}\right|\le c\varepsilon. \tag{7.20}$$

In particular we use here the weak dependence between two neutral lineages stated in Lemma 7.3 and the probability to descend from the first mutant for β_p and β_q obtained in Lemma 7.4.

7.5. **Neutral proportion at time T_ε^K.** According to Lemma 7.3 and Equation (7.20), it is enough to distinguish two cases for the randomly chosen neutral allele β_p: either its lineage has undergone one m-recombination, or no m-recombination. In the second case, β_p is a b_1. In the first one, the probability that β_p is a b_1 depends on the neutral proportion in the A population at the coalescence time. We now state that this proportion stays nearly constant during the first period.

Lemma 7.5. *There exist two positive finite constants c and ε_0 such that for $\varepsilon \le \varepsilon_0$,*

$$\limsup_{K\to\infty} \mathbb{P}\Big(\sup_{t\le T_\varepsilon^K} \Big| P^K_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} \Big| > \sqrt{\varepsilon}, T_\varepsilon^K < \infty \Big) \le c\varepsilon.$$

Lemma 7.5 whose proof is postponed to Appendix B, allows us to state the following lemma.

Lemma 7.6. *There exist two positive finite constants c and ε_0 such that for $\varepsilon \le \varepsilon_0$,*

$$\limsup_{K\to\infty} \hat{\mathbb{P}}\Big(\Big| P^K_{a,b_2}(T_\varepsilon^K) - \Big(1 - \exp\Big(-\frac{f_a r_K \log K}{S_{aA}}\Big)\Big) \frac{z_{Ab_2}}{z_A} \Big| > \varepsilon^{1/3} \Big) \le c\varepsilon^{1/6}.$$

Proof. $(\beta_i, i \le \lfloor \varepsilon K \rfloor)$ denotes the neutral alleles carried by the a-individuals at time T_ε^K and

$$A_2^K(i) := \{\beta_i \text{ has undergone exactly one m-recombination and is an allele } b_2\}.$$

If β_i is a b_2, either its genealogy has undergone one m-recombination with an individual Ab_2, either it has undergone more than two m-recombinations. Thus

$$0 \le N^K_{ab_2}(T_\varepsilon^K) - \sum_{i=1}^{\lfloor \varepsilon K \rfloor} \mathbb{1}_{A_2^K(i)} \le \sum_{i=1}^{\lfloor \varepsilon K \rfloor} \mathbb{1}_{\{TR_2^K(\beta_i) \ne -\infty\}}.$$

Moreover, the probability of $A_2^K(i)$ depends on the neutral proportions in the A-population when β_i m-recombines. For $i \le \lfloor \varepsilon K \rfloor$,

$$\Big| \hat{\mathbb{P}}\Big(A_2^K(i) \Big| TR_1^K(\beta_i) \ge 0, TR_2^K(\beta_i) = -\infty, \sup_{t\le T_\varepsilon^K} \Big| P^K_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} \Big| \le \sqrt{\varepsilon} \Big) - \Big(1 - \frac{z_{Ab_1}}{z_A}\Big) \Big| \le \sqrt{\varepsilon}. \tag{7.21}$$

Lemma 7.5 and Equation (A.5) ensure that $\limsup_{K\to\infty} \hat{\mathbb{P}}(\sup_{t\le T_\varepsilon^K} |P^K_{A,b_1}(t) - z_{Ab_1}/z_A| > \sqrt{\varepsilon}) \le c\varepsilon$, and Lemmas 7.3 and 7.4 that $|\hat{\mathbb{P}}(TR_1^K(\beta_i) \ge 0, TR_2^K(\beta_i) = -\infty) - (1 - e^{-f_a r_K \log K / S_{aA}})| \le c\varepsilon$. It yields:

$$\Big| \hat{\mathbb{P}}\Big(TR_1^K(\beta_i) \ge 0, TR_2^K(\beta_i) = -\infty, \sup_{t\le T_\varepsilon^K} \Big| P^K_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} \Big| \le \sqrt{\varepsilon} \Big) - \Big(1 - \exp\Big(-\frac{f_a r_K \log K}{S_{aA}}\Big)\Big) \Big| \le c\sqrt{\varepsilon}$$

for a finite c. Adding (7.21) we get:

$$\limsup_{K\to\infty} \Big| \hat{\mathbb{E}}[P^K_{a,b_2}(T_\varepsilon^K)] - (1 - \exp(-f_a r_K \log K / S_{aA}))(1 - z_{Ab_1}/z_A) \Big| \le c\sqrt{\varepsilon}. \tag{7.22}$$

In the same way, using the weak dependence between lineages stated in (7.20), we prove that $\limsup_{K\to\infty} |\hat{\mathbb{E}}[(P^K_{a,b_2})^2](T_\varepsilon^K) - (1 - e^{-f_a r_K \log K / S_{aA}})^2 (1 - z_{Ab_1}/z_A)^2| \le c\sqrt{\varepsilon}$. This implies, adding (7.22) that $\limsup_{K\to\infty} \hat{\mathrm{Var}}(P^K_{a,b_2}(T_\varepsilon^K)) \le c\sqrt{\varepsilon}$. We end the proof by using Chebyshev's Inequality. □

7.6. **Second and third periods.** Now we prove that during the second period the neutral proportion in the a population remains nearly constant. This is due to the short duration of this period, which does not go to infinity with the carrying capacity K.

Lemma 7.7. *There exist two positive finite constants c and ε_0 such that for $\varepsilon \le \varepsilon_0$,*

$$\limsup_{K\to\infty} \hat{\mathbb{P}}\Big(\Big| P^K_{a,b_1}(T^K_{ext}) - P^K_{a,b_1}(T_\varepsilon^K) \Big| > \varepsilon^{1/3} \Big) \le c\varepsilon^{1/3}. \tag{7.23}$$

Proof. Let us introduce the stopping time V_ε^K:

$$V_\varepsilon^K := \inf\Big\{t \geq 0, \sup_{t\leq t_\varepsilon, \alpha\in\mathscr{A}} |N_\alpha(T_\varepsilon^K + t)/K - n_\alpha{}^{(N(T_\varepsilon^K)/K)}(t)| > \varepsilon^3\Big\}.$$

Recall the definition of t_ε in (6.6) and that $(\mathscr{F}_t^K, t \geq 0)$ denotes the canonical filtration of N^K. Strong Markov property, Doob's Maximal Inequality and Equation (3.12) yield:

$$\begin{aligned}
&\mathbb{P}\Big(T_\varepsilon^K \leq \tilde{T}_\varepsilon^K, \sup_{t\leq t_\varepsilon} |M_a^K(T_\varepsilon^K + t\wedge V_\varepsilon^K) - M_a^K(T_\varepsilon^K)| > \sqrt{\varepsilon}\Big) \\
&\qquad = \mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K \leq \tilde{T}_\varepsilon^K} \mathbb{P}\Big(\sup_{t\leq t_\varepsilon} |M_a^K(T_\varepsilon^K + t\wedge V_\varepsilon^K) - M_a^K(T_\varepsilon^K)| > \sqrt{\varepsilon} | \mathscr{F}_{T_\varepsilon^K}\Big)\Big] \\
&\qquad\qquad \leq \frac{1}{\varepsilon}\mathbb{E}\Big[\mathbb{1}_{T_\varepsilon^K \leq \tilde{T}_\varepsilon^K}\Big(\langle M_a\rangle_{T_\varepsilon^K + t\wedge V_\varepsilon^K} - \langle M_a\rangle_{T_\varepsilon^K}\Big)\Big] \leq \frac{c(\varepsilon)\, t_\varepsilon}{\varepsilon K},
\end{aligned}$$

where $c(\varepsilon)$ is finite. But according to Equation (2.5) with $\delta = \varepsilon^3$, $\limsup_{K\to\infty} \mathbb{P}(V_\varepsilon^K < t_\varepsilon | T_\varepsilon^K \leq \tilde{T}_\varepsilon^K) = 0$. Moreover, Equations (3.1) and (3.6) imply for every $t \geq 0$

$$\sup_{t\leq t_\varepsilon} |P_{a,b_1}^K(T_\varepsilon^K + t) - P_{a,b_1}^K(T_\varepsilon^K)| \leq \sup_{t\leq t_\varepsilon} |M_a^K(T_\varepsilon^K + t) - M_a^K(T_\varepsilon^K)| + r_K t_\varepsilon f_a.$$

As r_K goes to 0 under Assumption 5, we finally get:

$$\limsup_{K\to\infty} \mathbb{P}\Big(\sup_{t\leq t_\varepsilon} |P_{a,b_1}^K(T_\varepsilon^K + t) - P_{a,b_1}^K(T_\varepsilon^K)| > \sqrt{\varepsilon}, T_\varepsilon^K \leq \tilde{T}_\varepsilon^K\Big) = 0. \tag{7.24}$$

Adding Lemma 4.3 ends the proof of Lemma 7.7 □

7.7. **End of the proof of Theorem 2 in the weak recombination regime.** Thanks to Lemmas 7.6 and 7.7 we get that for ε small enough,

$$\limsup_{K\to\infty} \hat{\mathbb{P}}\Big(\Big|P_{a,b_2}^K(T_{\text{ext}}^K) - (1 - \exp(-f_a r_K \log K / S_{aA}))z_{Ab_2}/z_A\Big| > 2\varepsilon^{1/3}\Big) \leq c\varepsilon^{1/6}.$$

Moreover, (A.6) ensures that $\liminf_{K\to\infty} \mathbb{P}(T_\varepsilon^K \leq \tilde{T}_\varepsilon^K | \text{Fix}^K) \geq 1 - c\varepsilon$, which implies

$$\limsup_{K\to\infty} \mathbb{P}\Big(\mathbb{1}_{\text{Fix}^K}\Big|P_{a,b_2}^K(T_{\text{ext}}^K) - (1 - \exp(-f_a r_K \log K / S_{aA}))z_{Ab_2}/z_A\Big| > 2\varepsilon^{1/3}\Big) \leq c\varepsilon^{1/6}.$$

This is equivalent to the convergence in probability and ends the proof of Theorem 2.

Appendix A. Technical results

We first present some results stated in [5]. We recall Definitions (2.8), (2.9), (4.7), (4.8), (3.11) and (6.6) and that the notation $.^K$ refers to processes that satisfy Assumption 3. Proposition 2 is a direct consequence of Equations (42), (71), (72) and (74) in [5]:

Proposition 2. *There exist two posivite finite constants M_1 and ε_0 such that for every $\varepsilon \leq \varepsilon_0$*

$$\lim_{K\to\infty} \mathbb{P}\Big(\Big|N_a^K(T_{ext}^K) - K\bar{n}_a\Big| > \varepsilon K \Big| Fix^K\Big) = 0, \quad \textit{and} \quad \limsup_{K\to\infty} \Big|\mathbb{P}(T_\varepsilon^K < \infty) - \frac{S_{aA}}{f_a}\Big| \leq M_1\varepsilon. \tag{A.1}$$

Moreover there exists $M_2 > 0$ such that for every $\varepsilon \leq \varepsilon_0$, the probability of the event

$$F_\varepsilon^K = \Big\{T_\varepsilon^K \leq \tilde{T}_\varepsilon^K, N_A^K(T_\varepsilon^K + t_\varepsilon) < \frac{\varepsilon^2 K}{2}, |N_a^K(T_\varepsilon^K + t_\varepsilon) - \bar{n}_a K| < \frac{\varepsilon K}{2}\Big\} \tag{A.2}$$

satisfies

$$\liminf_{K\to\infty} \mathbb{P}(T_\varepsilon^K \leq \tilde{T}_\varepsilon^K) \geq \liminf_{K\to\infty} \mathbb{P}(F_\varepsilon^K) \geq \frac{S_{aA}}{f_a} - M_2\varepsilon, \tag{A.3}$$

and if $z \in \Theta$, then there exist $V > 0$ and $c < \infty$ such that:

$$\liminf_{K\to\infty} \mathbb{P}(S_\varepsilon^K(z) > e^{VK}) \geq 1 - c\varepsilon. \tag{A.4}$$

Thanks to these results we can state the following Lemma, which allows us to focus on the event $\{T_\varepsilon^K \le \tilde{T}_\varepsilon^K\}$ rather than on Fix^K in Section 7.

Lemma A.1. *There exist two posivite finite constants c and ε_0 such that for every $\varepsilon \le \varepsilon_0$*

$$\limsup_{K\to\infty} \mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K > \tilde{T}_\varepsilon^K) \le c\varepsilon, \tag{A.5}$$

and

$$\limsup_{K\to\infty} \left[\mathbb{P}(\{T_\varepsilon^K \le \tilde{T}_\varepsilon^K\} \setminus Fix^K) + \mathbb{P}(Fix^K \setminus \{T_\varepsilon^K \le \tilde{T}_\varepsilon^K\})\right] \le c\varepsilon. \tag{A.6}$$

Proof. From Equation (A.1), we deduce that for $\varepsilon < \bar{n}_a/2$

$$\lim_{K\to\infty} \mathbb{P}(T_\varepsilon^K = \infty | \mathrm{Fix}^K) \le \lim_{K\to\infty} \mathbb{P}\left(\left|N_a^K(T_{\mathrm{ext}}^K) - K\bar{n}_a\right| > \varepsilon K \middle| \mathrm{Fix}^K\right) = 0. \tag{A.7}$$

Adding (2.15) we get that $\lim_{K\to\infty}\mathbb{P}(T_\varepsilon^K = \infty, \mathrm{Fix}^K) = 0$. The equality

$$\mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K > \tilde{T}_\varepsilon^K) = \mathbb{P}(T_\varepsilon^K < \infty) + \mathbb{P}(T_\varepsilon^K > \tilde{T}_\varepsilon^K) - \mathbb{P}(\{T_\varepsilon^K = \infty, \mathrm{Fix}^K\}^C)$$

with Equations (A.1) and (A.3) ends the proof of (A.5). From Equation (2.14) we deduce that $\lim_{K\to\infty}\mathbb{P}(\mathrm{Fix}^K | T_\varepsilon^K \le \tilde{T}_\varepsilon^K) = 1$. We end the proof of (A.6) thanks to Equations (A.7) and (A.5). □

We also recall some results on birth and death processes which proofs can be found in Lemma 3.1 in [28] and in [1] p 109 and 112.

Proposition 3. *Let $Z = (Z_t)_{t\ge 0}$ be a birth a death process with birth and death rates b and d. For $i \in \mathbb{Z}^+$, $T_i = \inf\{t \ge 0, Z_t = i\}$ and $\mathbb{P}_i$ (resp. $\mathbb{E}_i$) is the law (resp. expectation) of Z when $Z_0 = i$. Then*

- *For $i \in \mathbb{N}$ and $t \ge 0$,*

$$\mathbb{E}_i[Z_t] = ie^{(b-d)t}. \tag{A.8}$$

- *For $(i,j,k) \in \mathbb{Z}_+^3$ such that $j \in (i,k)$,*

$$\mathbb{P}_j(T_k < T_i) = \frac{1-(d/b)^{j-i}}{1-(d/b)^{k-i}}. \tag{A.9}$$

- *If $0 < d < b$, for every $i \in \mathbb{Z}_+$ and $t \ge 0$,*

$$\mathbb{P}_i(T_0 \le t) = \left(\frac{d(1-e^{(d-b)t})}{b-de^{(d-b)t}}\right)^i. \tag{A.10}$$

- *If $0 < d < b$, on the non-extinction event of Z, which has a probability $1-(d/b)^{Z_0}$, the following convergence holds:*

$$T_N/\log N \underset{N\to\infty}{\to} (1-d/b)^{-1}, a.s. \tag{A.11}$$

For $0 < s < 1$, if $\tilde{Z}^{(s)}$ denotes a random walk with jumps ± 1 where up jumps occur with probability $1/(2-s)$ and down jumps with probability $(1-s)/(2-s)$, we denote by $\mathbb{P}_i^{(s)}$ the law of $\tilde{Z}^{(s)}$ when the initial state is $i \in \mathbb{N}$ and introduce for every $a \in \mathbb{R}_+$ the stopping time

$$\tau_a := \inf\{n \in \mathbb{Z}_+, \tilde{Z}_n^{(s)} = \lfloor a \rfloor\}. \tag{A.12}$$

We also introduce for ε small enough and $0 \le j,k < \lfloor \varepsilon K \rfloor$, the quantities

$$q_{j,k}^{(s_1,s_2)} := \frac{\mathbb{P}_{k+1}^{(s_1)}(\tau_{\varepsilon K} < \tau_k)}{\mathbb{P}_{k+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_j)} = \frac{s_1}{1-(1-s_1)^{\lfloor \varepsilon K\rfloor - k}} \frac{1-(1-s_2)^{\lfloor \varepsilon K\rfloor - j}}{1-(1-s_2)^{k+1-j}}, \quad 0 < s_1, s_2 < 1, \tag{A.13}$$

whose expressions are direct consequences of (A.9). Let us now state two technical results. The first one is necessary to get Lemma 7.5. The second one helps us to control upcrossing numbers of the process N_a^K before reaching the size $\lfloor \varepsilon K \rfloor$.

Lemma A.2. • *Let $\mathscr{D}, \mathscr{C}, \mathscr{F}, \mathscr{G}$ and $\mathscr{H}$ be measurable events such that $\mathscr{D} \subset \mathscr{C} \subset \mathscr{F}$. Then we have*

$$\mathbb{P}(\mathscr{H} \cap (\mathscr{C} \cup \mathscr{G})) - \mathbb{P}(\mathscr{H} \cap \mathscr{D}) \leq \mathbb{P}(\mathscr{F}) - \mathbb{P}(\mathscr{D}) + \mathbb{P}(\mathscr{G} \setminus \mathscr{C}).$$

• *For $a \in]0, 1/2[$, $(s_1, s_2) \in [a, 1-a]^2$, and $0 \leq j \leq k < l < \lfloor \varepsilon K \rfloor$,*

$$q_{0,k}^{(s_1 \wedge s_2, s_1 \vee s_2)} \geq s_1 \wedge s_2 \quad \text{and} \quad \left| \frac{1}{q_{k,l}^{(s_1,s_2)}} - \frac{1}{q_{j,l}^{(s_2,s_1)}} \right| \leq \frac{2(1+1/s_2)}{ea^2 |\log(1-a)|} |s_2 - s_1| + \frac{(1-s_2)^{l+1-k}}{s_2^3}. \tag{A.14}$$

Proof. The proof of the first result is left to the reader. The first part of (A.14) is a direct consequence of Definiton (A.13). Let a be in $]0, 1/2[$ and consider functions $f_{\alpha,\beta} : x \mapsto (1-x^\alpha)/(1-x^\beta)$, $(\alpha, \beta) \in \mathbb{N}^2$, $x \in [a, 1-a]$. Then for $x \in [a, 1-a]$,

$$\| f'_{\alpha,\beta} \|_\infty \leq 2(ea^2 |\log(1-a)|)^{-1}. \tag{A.15}$$

From Equation (A.9), we get for $0 < s < 1$ and $0 \leq j \leq k < \lfloor \varepsilon K \rfloor$,

$$\begin{aligned} \left| \mathbb{P}_{l+1}^{(s)}(\tau_{\lfloor \varepsilon K \rfloor} < \tau_k) - \mathbb{P}_{l+1}^{(s)}(\tau_{\lfloor \varepsilon K \rfloor} < \tau_j) \right| &= \frac{(1-(1-s)^{k-j})((1-s)^{l+1-k} - (1-s)^{\lfloor \varepsilon K \rfloor - k})}{(1-(1-s)^{\lfloor \varepsilon K \rfloor - k})(1-(1-s)^{\lfloor \varepsilon K \rfloor - j})} \\ &\leq (1-s)^{l+1-k} s^{-2}. \end{aligned} \tag{A.16}$$

Triangle Inequality leads to:

$$\begin{aligned} \left| \frac{1}{q_{k,l}^{(s_1,s_2)}} - \frac{1}{q_{j,l}^{(s_2,s_1)}} \right| &= \left| \frac{\mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_k)}{\mathbb{P}_{l+1}^{(s_1)}(\tau_{\varepsilon K} < \tau_l)} - \frac{\mathbb{P}_{l+1}^{(s_1)}(\tau_{\varepsilon K} < \tau_j)}{\mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_l)} \right| \\ &\leq \left| \frac{1}{\mathbb{P}_{l+1}^{(s_1)}(\tau_{\varepsilon K} < \tau_l)} - \frac{1}{\mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_l)} \right| \mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_k) \\ &\quad + \frac{1}{\mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_l)} \left| \mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_k) - \mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_j) \right| \\ &\quad + \frac{1}{\mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_l)} \left| \mathbb{P}_{l+1}^{(s_1)}(\tau_{\varepsilon K} < \tau_j) - \mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_j) \right|. \end{aligned}$$

Noticing that $\mathbb{P}_{l+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_l) \geq \mathbb{P}_{l+1}^{(s_2)}(\tau_\infty < \tau_l) = \mathbb{P}_1^{(s_2)}(\tau_\infty < \tau_0) = s_2$, and using (A.16) and the Mean Value Theorem with (A.15), we get the second part of (A.14). □

Appendix B. Proofs of Lemmas 7.2 and 7.5

Proof of Equation (7.10). In the whole proof, the integer n_A denotes the state of N_A at some time smaller than $\tilde{T}_\varepsilon^K$ and thus belongs to I_ε^K which has been defined in (3.10). $\mathbb{P}_{(n_A,n_a)}$ (resp. $\hat{\mathbb{P}}_{(n_A,n_a)}$) denotes the probability $\mathbb{P}$ (resp. $\hat{\mathbb{P}}$) when $(N_A(0), N_a(0)) = (n_A, n_a) \in \mathbb{Z}_+^2$. We introduce for $u \in \mathbb{R}_+$ the hitting time of $\lfloor u \rfloor$ by the process N_a:

$$\sigma_u^K := \inf\{t \geq 0, N_a^K(t) = \lfloor u \rfloor\}. \tag{B.1}$$

Let (i, j, k) be in $\mathbb{Z}_+^3$ with $j < k < \lfloor \varepsilon K \rfloor$. Between jumps ζ_j^K and J^K the process N_a necessarily jumps from k to $k+1$. Then, either it reaches $\lfloor \varepsilon K \rfloor$ before returning to k, either it again jumps from k to $k+1$ and so on. Thus we approximate the probability that there is only one jump from k to $k+1$ by comparing $U_{j,k}^{(K,2)}$ with geometrically distributed random variables. As we do not know the value of N_A when N_a hits $k+1$ for the first time, we take the maximum over all the possible values in I_ε^K. Recall Definition (5.5). We get:

$$\begin{aligned} \hat{\mathbb{P}}(U_{j,k}^{(K,2)} = 1 | U_j^K = i) &\leq \sup_{n_A \in I_\varepsilon^K} \hat{\mathbb{P}}_{(n_A,k+1)}(T_\varepsilon^K < \sigma_k^K | T_\varepsilon^K < \sigma_j^K, U_j^K = i) \\ &= \sup_{n_A \in I_\varepsilon^K} \mathbb{P}_{(n_A,k+1)}(T_\varepsilon^K < \sigma_k^K | T_\varepsilon^K < \sigma_j^K, U_j^K = i, \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K). \end{aligned}$$

The value of U_j^K is correlated with the value of N_A when N_a hits $k+1$. But here we take the maximum over all the possible values of N_A. Hence the value of U_j^K has no influence on the last probability and we can ignore it. We thus obtain the inequality:

$$\hat{\mathbb{P}}(U_{j,k}^{(K,2)} = 1 | U_j^K = i) \leq \sup_{n_A \in I_\varepsilon^K} \frac{\mathbb{P}_{(n_A,k+1)}(T_\varepsilon^K < \sigma_k^K | \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K)}{\mathbb{P}_{(n_A,k+1)}(T_\varepsilon^K < \sigma_j^K | \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K)}.$$

Definition (5.5) allows us to compare these conditional probabilities with the probabilities of the same events under $\mathbb{P}^{(s_-(\varepsilon))}$ and $\mathbb{P}^{(s_+(\varepsilon))}$, namely $\sup_{n_A \in I_\varepsilon^K} \mathbb{P}_{(n_A,k+1)}(T_\varepsilon^K < \sigma_k^K | \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K) \leq \mathbb{P}_{k+1}^{(s_+(\varepsilon))}(\tau_{\varepsilon K} < \tau_k)$, and $\inf_{n_A \in I_\varepsilon^K} \mathbb{P}_{(n_A,k+1)}(T_\varepsilon^K < \sigma_j^K | \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K) \geq \mathbb{P}_{k+1}^{(s_-(\varepsilon))}(\tau_{\varepsilon K} < \tau_j)$. Recall (A.13). Then

$$\hat{\mathbb{P}}(U_{j,k}^{(K,2)} = 1 | U_j^K = i) \leq \frac{\mathbb{P}_{k+1}^{(s_+(\varepsilon))}(\tau_{\varepsilon K} < \tau_k)}{\mathbb{P}_{k+1}^{(s_-(\varepsilon))}(\tau_{\varepsilon K} < \tau_j)} = q_{j,k}^{(s_+(\varepsilon), s_-(\varepsilon))}.$$

In an analogous way we show that $\hat{\mathbb{P}}(U_{j,k}^{(K,2)} = 1 | U_j^K = i) \geq q_{j,k}^{(s_-(\varepsilon), s_+(\varepsilon))}$. We deduce that we can construct two geometrically distributed random variables G_1 and G_2, possibly on an enlarged space, with respective parameters $q_{j,k}^{(s_+(\varepsilon), s_-(\varepsilon))} \wedge 1$ and $q_{j,k}^{(s_-(\varepsilon), s_+(\varepsilon))}$ such that on the event $\{U_j^K = i\}$,

$$G_1 \leq U_{j,k}^{(K,2)} \leq G_2. \tag{B.2}$$

For the same reasons we obtain $q_{j,k}^{(s_-(\varepsilon), s_+(\varepsilon))} \leq \hat{\mathbb{P}}(U_{j,k}^{(K,2)} = 1) \leq q_{j,k}^{(s_+(\varepsilon), s_-(\varepsilon))} \wedge 1$, and again we can construct two random variables $G_1' \stackrel{d}{=} G_1$ and $G_2' \stackrel{d}{=} G_2$ such that

$$G_1' \leq U_{j,k}^{(K,2)} \leq G_2'. \tag{B.3}$$

Recall that $U_{0,k}^{(K,2)} = U_k^K$. Hence taking $j = 0$ and adding the first part of Equation (A.14) give the first inequality of (7.10). According to Definition (5.1), for ε small enough, $|s_+(\varepsilon) - s_-(\varepsilon)| \leq c\varepsilon$ for a finite c. Hence Equations (B.2), (B.3) and (A.14) entail the existence of a finite c such that for ε small enough $|\hat{\mathbb{E}}[U_{j,k}^{(K,2)} | U_j^K = i] - \hat{\mathbb{E}}[U_{j,k}^{(K,2)}]| \leq c\varepsilon + (1 - s_-(\varepsilon))^{k+1-j} / s_-^3(\varepsilon)$. Thus according to the first part of Equation (7.10),

$$\begin{aligned} \left| \hat{\mathrm{Cov}}(U_{j,k}^{(K,2)}, U_j^K) \right| &\leq \sum_{i \in \mathbb{N}^*} i \hat{\mathbb{P}}(U_j^K = i) \left| \hat{\mathbb{E}}[U_{j,k}^{(K,2)} | U_j^K = i] - \hat{\mathbb{E}}[U_{j,k}^{(K,2)}] \right| \\ &\leq \frac{2}{s_-^2(\varepsilon)} \left(c\varepsilon + \frac{(1 - s_-(\varepsilon))^{k+1-j}}{s_-^3(\varepsilon)} \right), \end{aligned} \tag{B.4}$$

where we use that $U_j^K \leq (U_j^K)^2$. This ends up the proof of (7.10). □

Proof of Equation (7.9). Definitions (2.3) and (5.5) ensure that if $n_A \in I_\varepsilon^K$,

$$\begin{aligned} \hat{\mathbb{P}}_{(n_A,k)}(N_a(dt) = k+1) &= \frac{\mathbb{P}_{(n_A,k)}(T_\varepsilon^K < \infty | N_a(dt) = k+1, \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K)}{\mathbb{P}_{(n_A,k)}(T_\varepsilon^K < \infty | \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K)} \mathbb{P}_{(n_A,k)}(N_a(dt) = k+1 | \mathscr{L}_\varepsilon^K, \mathscr{H}_\varepsilon^K) \\ &\geq \frac{\mathbb{P}_{k+1}^{(s_-(\varepsilon))}(\tau_{\varepsilon K} < \tau_0)}{\mathbb{P}_k^{(s_+(\varepsilon))}(\tau_{\varepsilon K} < \tau_0)} f_a k (dt + o(dt)) \\ &= \frac{1 - (1 - s_-(\varepsilon))^{k+1}}{1 - (1 - s_-(\varepsilon))^{\lfloor \varepsilon K \rfloor}} \frac{1 - (1 - s_+(\varepsilon))^{\lfloor \varepsilon K \rfloor}}{1 - (1 - s_+(\varepsilon))^k} f_a k (dt + o(dt)) \\ &\geq s_-^2(\varepsilon) f_a k (dt + o(dt)), \end{aligned}$$

and

$$\begin{aligned}\hat{\mathbb{P}}_{(n_A,k)}(N_A(dt)\neq n_A) &\leq \frac{\mathbb{P}_k^{(s_+(\varepsilon))}(\tau_{\varepsilon K}<\tau_0)}{\mathbb{P}_k^{(s_-(\varepsilon))}(\tau_{\varepsilon K}<\tau_0)}\mathbb{P}_{(n_A,k)}(N_A(dt)\neq n_A|\mathscr{L}_\varepsilon^K,\mathscr{H}_\varepsilon^K)\\ &\leq (1+c\varepsilon)2f_A\bar{n}_AK(dt+o(dt)).\end{aligned}$$

for a finite c, where we use (A.9) and that $D_A+C_{A,A}\bar{n}_A=f_A$. Thus for ε small enough:

$$\hat{\mathbb{P}}(N_a(\tau_{m+1}^K)\neq N_a(\tau_m^K)|N_a(\tau_m^K)=k)\geq\frac{s_-^2(\varepsilon)f_ak}{3f_A\bar{n}_AK}.$$

If D_k^K denotes the downcrossing number from k to $k-1$ before T_ε^K, then under the probability $\hat{\mathbb{P}}$, we can bound $U_k^K+D_k^K+H_k^K$ by the sum of $U_k^K+D_k^K$ independent geometrically distributed random variables G_i^K with parameter $s_-^2(\varepsilon)f_ak/3f_A\bar{n}_AK$ and $H_k^K\leq\sum_{1\leq i\leq U_k^K+D_k^K}(G_i^K-1)$. Let us notice that if $k\geq 2$, $D_k^K=U_{k-1}^K-1$, and $D_1^K=0$. Using the first part of (7.10) twice we get

$$\hat{\mathbb{E}}[H_k^K]\leq\Big(\frac{4}{s_-^2(\varepsilon)}-1\Big)\Big(\frac{3f_A\bar{n}_AK}{s_-^2(\varepsilon)f_ak}-1\Big),$$

which ends up the proof of the first inequality in (7.9).
As the mutant population size is not Markovian we cannot use symmetry and Strong Markov property to control the dependence of jumps before and after the last visit to a given state as in [28]. Hence we describe the successive excursions of N_a^K above a given level to get the last inequality in (7.9). Let $\tilde{U}_{j,k}^{(i)}$ be the number of jumps from k to $k+1$ during the ith excursion above j. We first bound the expectation $\hat{\mathbb{E}}[(\tilde{U}_{j,k}^{(i)})^2]$. During an excursion above j, N_a hits $j+1$, but we do not know the value of N_A at this time. Thus we take the maximum value for the probability when n_A belongs to I_ε^K, and $\hat{\mathbb{P}}(\tilde{U}_{j,k}^{(i)}\geq 1)\leq\sup_{n_A\in I_\varepsilon^K}\hat{\mathbb{P}}_{(j+1,n_A)}(\sigma_{k+1}^K<\sigma_j^K|\sigma_j^K<T_\varepsilon^K)$. Then using Coupling (5.3) and Definitions (5.4) and (5.5) we obtain

$$\begin{aligned}\hat{\mathbb{P}}\Big(\tilde{U}_{j,k}^{(i)}\geq 1\Big) &\leq \sup_{n_A\in I_\varepsilon^K}\frac{\mathbb{P}_{(j+1,n_A)}(T_\varepsilon^K<\infty|\sigma_{k+1}^K<\sigma_j^K<T_\varepsilon^K,\mathscr{L}_\varepsilon^K,\mathscr{H}_\varepsilon^K)\mathbb{P}_{(j+1,n_A)}(\sigma_j^K<T_\varepsilon^K|\sigma_{k+1}^K<\sigma_j^K,\mathscr{L}_\varepsilon^K,\mathscr{H}_\varepsilon^K)}{\mathbb{P}_{(j+1,n_A)}(T_\varepsilon^K<\infty|\sigma_j^K<T_\varepsilon^K,\mathscr{L}_\varepsilon^K,\mathscr{H}_\varepsilon^K)\mathbb{P}_{(j+1,n_A)}(\sigma_j^K<T_\varepsilon^K|\mathscr{L}_\varepsilon^K,\mathscr{H}_\varepsilon^K)}\\ &\quad .\mathbb{P}_{(j+1,n_A)}(\sigma_{k+1}^K<\sigma_j^K|\mathscr{L}_\varepsilon^K,\mathscr{H}_\varepsilon^K)\\ &\leq \frac{\mathbb{P}_j^{(s_+(\varepsilon))}(\tau_{\varepsilon K}<\tau_0)\mathbb{P}_{k+1}^{(s_-(\varepsilon))}(\tau_j<\tau_{\varepsilon K})\mathbb{P}_{j+1}^{(s_+(\varepsilon))}(\tau_{k+1}<\tau_j)}{\mathbb{P}_j^{(s_-(\varepsilon))}(\tau_{\varepsilon K}<\tau_0)\mathbb{P}_{j+1}^{(s_+(\varepsilon))}(\tau_j<\tau_{\varepsilon K})}.\end{aligned}$$

Adding Equation (A.9) we finally get

$$\hat{\mathbb{P}}\Big(\tilde{U}_{j,k}^{(i)}\geq 1\Big)\leq\frac{(1-s_-(\varepsilon))^{k+1-j}}{s_-(\varepsilon)(1-s_+(\varepsilon))}. \tag{B.5}$$

Moreover if $\tilde{U}_{j,k}^{(i)}\geq 1$, N_a necessarily hits k after its first jump from k to $k+1$, and before its return to j. Using the same techniques as before we get:

$$\hat{\mathbb{P}}\Big(\tilde{U}_{j,k}^{(i)}=1|\tilde{U}_{j,k}^{(i)}\geq 1\Big)\geq\inf_{n_A\in I_\varepsilon^K}\hat{\mathbb{P}}_{(n_A,k)}(\sigma_j^K<\sigma_{k+1}^K|\sigma_j^K<T_\varepsilon^K)\geq\frac{\mathbb{P}_j^{(s_-(\varepsilon))}(\tau_{\varepsilon K}<\tau_0)\mathbb{P}_k^{(s_+(\varepsilon))}(\tau_j<\tau_{k+1})}{\mathbb{P}_j^{(s_+(\varepsilon))}(\tau_{\varepsilon K}<\tau_0)\mathbb{P}_k^{(s_-(\varepsilon))}(\tau_j<\tau_{\varepsilon K})},$$

which yields

$$\hat{\mathbb{P}}\Big(\tilde{U}_{j,k}^{(i)}=1|\tilde{U}_{j,k}^{(i)}\geq 1\Big)\geq s_-(\varepsilon)s_+(\varepsilon)\Big(\frac{1-s_+(\varepsilon)}{1-s_-(\varepsilon)}\Big)^{k-j}\geq s_-^2(\varepsilon)\Big(\frac{1-s_+(\varepsilon)}{1-s_-(\varepsilon)}\Big)^{k-j}. \tag{B.6}$$

Furthermore, given that $\tilde{U}^{(i)}_{j,k}$ is non-null, $\tilde{U}^{(i)}_{j,k}$ is smaller than a geometrically distributed random variable with parameter $q = s_-^2(\varepsilon)[(1-s_+(\varepsilon))/(1-s_-(\varepsilon))]^{k-j}$. In particular,

$$\mathbb{E}\Big[(\tilde{U}^{(i)}_{j,k})^2 | \tilde{U}^{(i)}_{j,k} \geq 1\Big] \leq \frac{2}{q^2} = \frac{2}{s_-^4(\varepsilon)}\Big(\frac{1-s_-(\varepsilon)}{1-s_+(\varepsilon)}\Big)^{2(k-j)}.$$

Adding Equations (B.5) and recalling that $|s_+(\varepsilon) - s_-(\varepsilon)| \leq c\varepsilon$ for c finite and ε small enough yield

$$\hat{\mathbb{E}}\Big[(\tilde{U}^{(i)}_{j,k})^2\Big] \leq \frac{2\lambda_\varepsilon^{k-j}}{s_-^5(\varepsilon)(1-s_+(\varepsilon))}, \quad \text{where} \quad \lambda_\varepsilon := \frac{(1-s_-(\varepsilon))^3}{(1-s_+(\varepsilon))^2} < 1.$$

Using that for $n \in \mathbb{N}$ and $(x_i, 1 \leq i \leq n) \in \mathbb{R}^n$, $(\sum_{1\leq i\leq n} x_i)^2 \leq n\sum_{1\leq i\leq n} x_i^2$ and that the number of excursions above j before T_ε^K is $U_j^K - 1$, we get

$$\hat{\mathbb{E}}\Big[(U^{(K,1)}_{j,k})^2\Big] \leq \hat{\mathbb{E}}\Big[U_j^K - 1\Big]\frac{2\lambda_\varepsilon^{k-j}}{s_-^5(\varepsilon)(1-s_+(\varepsilon))} \leq \frac{4\lambda_\varepsilon^{k-j}}{s_-^7(\varepsilon)(1-s_+(\varepsilon))},$$

where we use the first part of Equation (7.10). This ends the proof of Equation (7.9). □

Proof of Equation (7.11). Definition (A.13), Inequality (B.3) and Equation (A.9) yield:

$$r_K \sum_{k=1}^{\lfloor \varepsilon K\rfloor -1} \frac{\hat{\mathbb{E}}[U_k^K]}{k+1} \geq r_K \sum_{k=1}^{\lfloor \varepsilon K\rfloor -1} \Big[(k+1)q_{0,k}^{(s_+(\varepsilon),s_-(\varepsilon))}\Big]^{-1} = \frac{r_K(A-B)}{s_+(\varepsilon)(1-(1-s_-(\varepsilon))^{\lfloor \varepsilon K\rfloor})},$$

with

$$A := \sum_{k=1}^{\lfloor \varepsilon K\rfloor -1} \frac{1-(1-s_-(\varepsilon))^{k+1}}{k+1}, \quad \text{and} \quad B := (1-s_+(\varepsilon))^{\lfloor \varepsilon K\rfloor} \sum_{k=1}^{\lfloor \varepsilon K\rfloor -1} \frac{1-(1-s_-(\varepsilon))^{k+1}}{(1-s_+(\varepsilon))^k(k+1)}.$$

For large K, $A = \log(\varepsilon K) + O(1)$. Moreover, Lemma 3.5 in [28] ensures that for $u > 1$ there exists $D(u) < \infty$ such that $\sum_{k=1}^{\lfloor \varepsilon K\rfloor} u^k/(k+1) \leq D(u)u^{\lfloor \varepsilon K\rfloor}/\lfloor \varepsilon K\rfloor$. This implies that $B \leq c/\lfloor \varepsilon K\rfloor$ for a finite c. Finally, by definition, for ε small enough, $|S_{aA}/f_a - s_+(\varepsilon)| \leq c\varepsilon$ for a finite constant c. This yields

$$r_K \sum_{k=1}^{\lfloor \varepsilon K\rfloor -1} \frac{\hat{\mathbb{E}}[U_k^K]}{k+1} \geq (1-c\varepsilon)\frac{r_K f_a \log K}{S_{aA}}$$

for a finite c and concludes the proof for the lower bound. The upper bound is obtained in the same way. This concludes the proof of Lemma 7.2. □

Proof of Lemma 7.5. We use Coupling (5.3) to control the growing of the mutant population during the first period of invasion, and the semi-martingale decomposition in Proposition 1 to bound the fluctuations of M_A. The hitting time of $\lfloor \varepsilon K\rfloor$ and non-extinction event of Z_ε^* are denoted by:

$$T_\varepsilon^{*,K} = \inf\{t \geq 0, Z_\varepsilon^*(t) = \lfloor \varepsilon K\rfloor\}, \quad \text{and} \quad F_\varepsilon^* = \Big\{Z_\varepsilon^*(t) \geq 1, \forall\, t \geq 0\Big\}, \quad * \in \{-,+\}.$$

From (5.3) we get $\mathscr{D} = \{F_\varepsilon^-, T_\varepsilon^K \leq \tilde{T}_\varepsilon^K\} \subset \mathscr{C} = \{T_\varepsilon^K < \infty, T_\varepsilon^K \leq \tilde{T}_\varepsilon^K\} \subset \mathscr{F} = \{T_\varepsilon^{+,K} < \infty, T_\varepsilon^K \leq \tilde{T}_\varepsilon^K\}$. Lemma A.2 with $\mathscr{G} = \{T_\varepsilon^K > \tilde{T}_\varepsilon^K\} \cup \{T_\varepsilon^K = \infty\}$, and $\mathscr{H} = \{\sup_{t\leq T_\varepsilon^K} |P_{A,b_1}(t) - z_{Ab_1}/z_A| > \sqrt{\varepsilon}\}$ yields:

$$|\mathbb{P}(\mathscr{H}, F_\varepsilon^-, T_\varepsilon^K \leq \tilde{T}_\varepsilon^K) - \mathbb{P}(\mathscr{H}, T_\varepsilon^K < \infty)| \leq |\mathbb{P}(T_\varepsilon^{+,K} < \infty) - \mathbb{P}(F_\varepsilon^-)| + \mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K > \tilde{T}_\varepsilon^K). \tag{B.7}$$

But thanks to Equation (A.9) we get that $\mathbb{P}(T_\varepsilon^{+,K} < \infty) - \mathbb{P}(F_\varepsilon^-) = s_+(\varepsilon)/(1-(1-s_+(\varepsilon))^{\lfloor \varepsilon K\rfloor}) - s_-(\varepsilon)$, and (5.1) and (A.5) lead to:

$$\limsup_{K\to\infty} |\mathbb{P}(T_\varepsilon^{+,K} < \infty) - \mathbb{P}(F_\varepsilon^-)| + \mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K > \tilde{T}_\varepsilon^K) \leq c\varepsilon. \tag{B.8}$$

It allows us to focus on the probability $\mathbb{P}(\mathscr{H}, F_\varepsilon^-, T_\varepsilon^K \leq \tilde{T}_\varepsilon^K)$. We recall that $|N_{Ab_1}N_{ab_2} - N_{Ab_2}N_{ab_1}| \leq N_A N_a$, and that Assumption 5 holds. Then (3.1) and (3.11) imply for ε small enough

$$\sup_{t\leq T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K} \Big|P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} - M_A(t)\Big| \leq r_K f_a T_\varepsilon^K \sup_{t\leq T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K} \Big\{\frac{N_a(t)}{N_A(t)}\Big\} \leq \frac{r_K f_a \varepsilon T_\varepsilon^K}{\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A}} \leq \frac{c\varepsilon T_\varepsilon^K}{\log K}.$$

Moreover, $F_\varepsilon^- \cap \{T_\varepsilon^K \le \tilde{T}_\varepsilon^K\} \subset F_\varepsilon^- \cap \{T_\varepsilon^K \le T_\varepsilon^{-,K}\}$. Thus we get

$$\mathbb{P}\Big(\sup_{t\le T_\varepsilon^K} \Big| P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} - M_A(t)\Big| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \le \tilde{T}_\varepsilon^K\Big) \le \mathbb{P}\Big(\frac{c\varepsilon T_\varepsilon^{-,K}}{\log K} > \sqrt{\varepsilon}/2, F_\varepsilon^-\Big).$$

Finally, Equation (A.11) ensures that $\lim_{K\to\infty} T_\varepsilon^{-,K}/\log K = s_-(\varepsilon)^{-1}$ a.s. on the non-extinction event F_ε^-. Thus for $\varepsilon < s_-(\varepsilon)/2c$,

$$\lim_{K\to\infty} \mathbb{P}\Big(\sup_{t\le T_\varepsilon^K} \Big| P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} - M_A(t)\Big| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \le \tilde{T}_\varepsilon^K\Big) = 0. \tag{B.9}$$

To control the term $|M_A|$, we introduce the sequence of real numbers $t_K = (2f_a \log K)/S_{aA}$:

$$\mathbb{P}\Big(\sup_{t\le T_\varepsilon^K} |M_A(t)| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \le \tilde{T}_\varepsilon^K\Big) \le \mathbb{P}\Big(\sup_{t\le T_\varepsilon^K} |M_A(t)| > \frac{\sqrt{\varepsilon}}{2}, T_\varepsilon^K \le \tilde{T}_\varepsilon^K \wedge t_K\Big) + \mathbb{P}(T_\varepsilon^K > t_K, F_\varepsilon^-).$$

Equation (5.1) yields for ε small enough, $t_K.s_-(\varepsilon)/\log K > 3/2$. Thus thanks to (A.11) we get,

$$\lim_{K\to\infty} \mathbb{P}(T_\varepsilon^K > t_K, F_\varepsilon^-) \le \lim_{K\to\infty} \mathbb{P}(T_\varepsilon^{-,K} > t_K, F_\varepsilon^-) = 0.$$

Applying Doob's maximal inequality to the submartingale $|M_A|$ and (3.12) we get:

$$\begin{aligned} \mathbb{P}(\sup_{t\le T_\varepsilon^K} |M_A(t)| > \sqrt{\varepsilon}/2, T_\varepsilon^K \le \tilde{T}_\varepsilon^K \wedge t_K) &\le \mathbb{P}(\sup_{t\le t_K} |M_A(t\wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K)| > \sqrt{\varepsilon}/2) \\ &\le \frac{4}{\varepsilon}\mathbb{E}\Big[\langle M_A\rangle_{t_K\wedge T_\varepsilon^K \wedge \tilde{T}_\varepsilon^K}\Big] \le \frac{ct_K}{K}, \end{aligned}$$

which goes to 0 at infinity. Adding Equation (B.9) leads to:

$$\lim_{K\to\infty} \mathbb{P}\Big(\sup_{t\le T_\varepsilon^K} \Big| P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A}\Big| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \le \tilde{T}_\varepsilon^K\Big) = 0.$$

Finally, Equations (B.7) and (B.8) complete the proof of Lemma 7.5. □

Acknowledgements: *The author would like to thank Jean-François Delmas and Sylvie Méléard for their help and their careful reading of this paper. She also wants to thank Sylvain Billiard and Pierre Collet for fruitful discussions during her work and several suggestions. This work was partially funded by project MANEGE 'Modèles Aléatoires en Écologie, Génétique et Évolution' 09-BLAN-0215 of ANR (French national research agency) and Chaire Modélisation Mathématique et Biodiversité VEOLIA-Ecole Polytechnique-MNHN-F.X.*

References

[1] K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].

[2] R. D. Barrett and D. Schluter. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1):38–44, 2008.

[3] N. H. Barton. The effect of hitch-hiking on neutral genealogies. *Genetical Research*, 72(2):123–133, 1998.

[4] S. Billiard, R. Ferrière, S. Méléard, and V. C. Tran. Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks. *arXiv preprint arXiv:1310.6274*, 2013.

[5] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic Processes and their Applications*, 116(8):1127–1160, 2006.

[6] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theoretical population biology*, 69(3):297–321, 2006.

[7] N. Champagnat and A. Lambert. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *The Annals of Applied Probability*, 17(1):102–155, 2007.

[8] N. Champagnat and S. Méléard. Invasion and adaptive evolution for individual-based spatially structured populations. *Journal of Mathematical Biology*, 55(2):147–188, 2007.

[9] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*, 151(1-2):45–94, 2011.

[10] P. Collet, S. Méléard, and J. A. Metz. A rigorous model study of the adaptive dynamics of mendelian diploids. *Journal of Mathematical Biology*, pages 1–39, 2011.
[11] C. Coron. Stochastic modeling of density-dependent diploid populations and extinction vortex. *arXiv preprint arXiv:1207.4920, to appear in Advances in Applied Probabilities*, 2012.
[12] C. Coron. Slow-fast stochastic diffusion dynamics and quasi-stationary distributions for diploid populations. *arXiv preprint arXiv:1309.3405*, 2013.
[13] R. Durrett. *Probability models for DNA sequence evolution*. Springer, 2008.
[14] R. Durrett and J. Schweinsberg. Approximating selective sweeps. *Theoretical population biology*, 66(2):129–138, 2004.
[15] A. Eriksson, P. Fernström, B. Mehlig, and S. Sagitov. An accurate model for genetic hitchhiking. *Genetics*, 178(1):439–451, 2008.
[16] A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. An approximate sampling formula under genetic hitchhiking. *The Annals of Applied Probability*, 16(2):685–729, 2006.
[17] S. Ethier and T. Kurtz. Markov processes: Characterization and convergence, 1986, 1986.
[18] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, 14(4):1880–1919, 2004.
[19] N. Ikeda and S. Watanabe. Stochastic differential equations and diffusion processes. 1989.
[20] N. L. Kaplan, R. Hudson, and C. Langley. The" hitchhiking effect" revisited. *Genetics*, 123(4):887–899, 1989.
[21] S. Leocard. Selective sweep and the size of the hitchhiking set. *Advances in Applied Probability*, 41(3):731–764, 2009.
[22] G. McVean. The structure of linkage disequilibrium around a selective sweep. *Genetics*, 175(3):1395–1406, 2007.
[23] S. Méléard and V. C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of Mathematical Biology*, 58(6):881–921, 2009.
[24] J. A. Metz, S. A. Geritz, G. Meszéna, F. J. Jacobs, and J. Van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. *Stochastic and spatial structures of dynamical systems*, 45:183–231, 1996.
[25] B. M. Peter, E. Huerta-Sanchez, and R. Nielsen. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS genetics*, 8(10):e1003011, 2012.
[26] P. Pfaffelhuber and A. Studeny. Approximating genealogies for partially linked neutral loci under a selective sweep. *Journal of mathematical biology*, 55(3):299–330, 2007.
[27] M. Prezeworski, G. Coop, and J. D. Wall. The signature of positive selection on standing genetic variation. *Evolution*, 59(11):2312–2323, 2005.
[28] J. Schweinsberg and R. Durrett. Random partitions approximating the coalescence of lineages during a selective sweep. *The Annals of Applied Probability*, 15(3):1591–1651, 2005.
[29] J. M. Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genet Res*, 23(1):23–35, 1974.
[30] W. Stephan, T. H. Wiehe, and M. W. Lenz. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology*, 41(2):237–254, 1992.