



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

Charline Smadi

## ► To cite this version:

Charline Smadi. An Eco-Evolutionary approach of Adaptation and Recombination in a large population of varying size. 2014. hal-00948097v3

**HAL Id: hal-00948097**

**<https://hal.science/hal-00948097v3>**

Preprint submitted on 24 Nov 2014 (v3), last revised 24 Dec 2014 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 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), or on preexisting alleles that become advantageous after an environmental change (soft selective sweep from standing variation). New mutations are sources of diversity, and hard selective sweeps were until recently the only considered way of adaptation. Soft selective sweeps from standing variation allow a faster adaptation to novel environments, and their importance is growing in empirical and theoretical studies (Orr and Betancourt [30], Hermisson and Pennings [19], Prezeworski, Coop and Wall [35], Barrett and Schluter [2], Durand and al [12]). In particular Messer and Petrov [26] review a lot of evidence, from individual case studies as well as from genome-wide scans, that soft sweeps (from standing variation and from recurrent mutations) are common in a broad range of organisms. These distinct selective sweeps entail different genetic signatures in the vicinity of the novel fixed allele, and the multiplication of genetic data available allows one to detect these signatures in current populations as described by Peter, Huerta-Sanchez and Nielsen [33]. To do this in an effective way, it is necessary to identify accurately the signatures left by these two modes of adaptation. We will not consider in this work the soft selective sweeps from recurrent mutations. For a study of these sweeps we refer to [31, 32, 20].

In this work, we consider a sexual haploid population of varying size, modeled by a birth and death process with density dependent competition. Each individual's ability to survive and reproduce 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 entails progressive variations in the number of individuals carrying a given genotype. This process, called natural selection, is a key mechanism of evolution. Such an eco-evolutionary approach has been introduced by Metz and coauthors in [27] and made rigorous in the seminal paper of Fournier and Méléard [18]. Then it has been developed by Champagnat, Méléard and coauthors (see [5, 7, 6] and references therein) for the haploid asexual case and by Collet, Méléard and Metz [8] and Coron and coauthors [10, 11] for the diploid sexual case. The recent work of Billiard and coauthors [4] studies

---

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.

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 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 a soft selective sweep from standing variation 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 both represented 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 proportion 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 sweeps. 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 from standing variation. 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 where  $r_K \log K$  is large, and a weak one where  $r_K \log K$  is bounded. In both recombination regimes, we give the asymptotic value of the final neutral allele proportion 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 has been called genetic hitch-hiking by Maynard Smith and Haigh [24].

The first studies of hitch-hiking, initiated by Maynard Smith and Haigh [24], have modeled the mutant population size as the solution of a deterministic logistic equation [29, 22, 38, 37]. Kaplan and coauthors [22] described the neutral genealogies by a structured coalescent where the background was the frequency of the beneficial allele. Barton [3] was the first to point out the importance of the stochasticity of the mutant population size and the errors made by ignoring it. He divided the sweep in four periods: the two last ones are the analogues of the two last steps described in [5], and the two first ones correspond to the first one in [5]. Following the approaches of [22] and [3], a series of works studied the genealogies of neutral alleles sampled at the end of the sweep and took into account the randomness of the mutant population size during the sweep. In particular, Durrett and Schweinsberg [14, 36], Etheridge and coauthors [16], Pfaffelhuber and Studeny [34], and Leocard [23] described the population process by a structured coalescent and finely studied genealogies of neutral alleles during the sweep. Eriksson and coauthors [15] described a deterministic approximation for the growth of the beneficial 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 was constant and the individuals' "selective value" did 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 [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 semi-martingale decomposition of neutral proportions, a key tool in the different proofs. Section 4 is devoted to the proof for the soft sweep from standing variation. 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 used 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 [36]. Finally in the Appendix we state technical results.

This work stems from the papers of Champagnat [5] and Schweinsberg and Durrett [36]. 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  $\varepsilon$ . The set  $\mathbb{N} := \{1, 2, \dots\}$  denotes the set of positive integers.

## 2. MODEL AND MAIN RESULTS

We introduce the sets  $\mathcal{A} = \{A, a\}$ ,  $\mathcal{B} = \{b_1, b_2\}$ , and  $\mathcal{E} = \{A, a\} \times \{b_1, b_2\}$  to describe the genetic background of individuals. The state of the population will be given by the four dimensional Markov process  $N^{(z, K)} = (N_{\alpha\beta}^{(z, K)}(t), (\alpha, \beta) \in \mathcal{E}, t \geq 0)$  where  $N_{\alpha\beta}^{(z, K)}(t)$  denotes the number of individuals with alleles  $(\alpha, \beta)$  at time  $t$  when the carrying capacity is  $K \in \mathbb{N}$  and the initial state is  $\lfloor zK \rfloor$  with  $z = (z_{\alpha\beta}, (\alpha, \beta) \in \mathcal{E}) \in \mathbb{R}_+^{\mathcal{E}}$ . 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  $\alpha \in \mathcal{A}$ ,  $f_\alpha$  and  $D_\alpha$  denote the birth rate and the intrinsic death rate of an individual carrying allele  $\alpha$ .
- For  $(\alpha, \alpha') \in \mathcal{A}^2$ ,  $C_{\alpha, \alpha'}$  represents the competitive pressure felt by an individual carrying allele  $\alpha$  from an individual carrying allele  $\alpha'$ .
- $K \in \mathbb{N}$  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  $(\alpha, \beta)$  and  $(\alpha', \beta')$  in  $\mathcal{E}$  give birth, the newborn individual, either is a clone of one parent and carries alleles  $(\alpha, \beta)$  or  $(\alpha', \beta')$  each with probability  $(1 - r_K)/2$ , or has a mixed genotype  $(\alpha, \beta')$  or  $(\alpha', \beta)$  each with probability  $r_K/2$ .

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

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

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

$$(2.1) \quad d_{\alpha\beta}^K(n) = [D_\alpha + C_{\alpha, A} n_A / K + C_{\alpha, a} n_a / K] n_{\alpha\beta}.$$

An individual carrying allele  $\alpha \in \mathcal{A}$  produces gametes with rate  $f_\alpha$ , thus the relative frequencies of gametes available for reproduction are

$$p_{\alpha\beta}(n) = f_\alpha n_{\alpha\beta} / (f_A n_A + f_a n_a), \quad (\alpha, \beta) \in \mathcal{E}.$$

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  $((a, b_2), (a, b_1))$  is not able to generate an individual  $(A, b_1)$ ). We detail the computation of the cumulative birth rate of individuals  $(A, b_1)$ :

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

If we denote by  $\bar{\alpha}$  (resp.  $\bar{\beta}$ ) the complement of  $\alpha$  in  $\mathcal{A}$  (resp.  $\beta$  in  $\mathcal{B}$ ), we obtain in the same way the cumulative birth rate of individuals  $(\alpha, \beta)$ :

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

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  $(N_A^{(z,K)}, N_a^{(z,K)})$ , which is also a birth and death process with competition. It has been studied in [5] and its cumulative death and birth rates, which are direct consequences of (2.1) and (2.2), satisfy for  $\alpha \in \mathcal{A}$ :

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

It is proven in [5] that when  $N_A^{(z,K)}$  and  $N_a^{(z,K)}$  are of order  $K$ , the rescaled population process  $(N_A^{(z,K)}/K, N_a^{(z,K)}/K)$  is well approximated by the dynamical system:

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

More precisely Theorem 3 (b) in [5] states that for every compact subset

$$B \subset (\mathbb{R}_+^{A \times \mathcal{B}} \setminus (0, 0)) \times (\mathbb{R}_+^{a \times \mathcal{B}} \setminus (0, 0))$$

and finite real number  $T$ , we have for any  $\delta > 0$ ,

$$(2.5) \quad \lim_{K \rightarrow \infty} \sup_{z \in B} \mathbb{P} \left( \sup_{0 \leq t \leq T, \alpha \in \mathcal{A}} |N_\alpha^{(z,K)}(t)/K - n_\alpha^{(z)}(t)| \geq \delta \right) = 0.$$

Moreover, if we assume

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

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

$$(2.7) \quad \bar{n}_\alpha = \frac{f_\alpha - D_\alpha}{C_{\alpha,a}} > 0, \quad \alpha \in \mathcal{A}.$$

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

$$(2.8) \quad S_{\alpha\bar{a}} = f_\alpha - D_\alpha - C_{\alpha,\bar{a}} \bar{n}_{\bar{a}}, \quad \alpha \in \mathcal{A},$$

it corresponds to the per capita growth rate of a mutant  $\alpha$  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*

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

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

The case we are interested in is referred in population genetics as soft selection [39]: it is both frequency and density dependent. This kind of selection has no influence on the order of the total population size, which has the same order as the carrying capacity  $K$ . However, the factor multiplying the carrying capacity can be modified, as the way the individuals use the resources depends on the ecological parameters. We focus on strong selection coefficient, which are characterized by  $S_{aA} \gg 1/K$ . In this case the selection outcompetes the genetic drift. However we do not need to assume  $S_{aA} \ll 1$  to get approximations unlike [24, 3, 37]. To study the genealogy of the selected allele when the selection coefficient is weak ( $S_{aA}K$  moderate or small) we refer to the approach of Neuhauser and Krone [28].

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  $(z, K) \in \mathbb{R}_+^{\mathcal{E}} \times \mathbb{N}$ :

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

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

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

the proportion of alleles  $\beta$  in the  $\alpha$ -population at time  $t$ . 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:

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

We first focus on soft selective sweeps from standing variation. 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  $(z, r, t) \in (\mathbb{R}_+^{\mathcal{E}})^* \times [0, 1] \times \mathbb{R}_+$  by

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

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

$$h : (z, r, t) \in (\mathbb{R}_+^{\mathcal{E}})^* \times [0, 1] \times \mathbb{R}_+ \mapsto r f_A f_a \int_0^t n_A^{(z)}(s) / (f_A n_A^{(z)}(s) + f_a n_a^{(z)}(s)) ds$$

non-decreasing in time, then

$$0 \leq F(z, r, t) \leq \int_0^t \partial_s h(z, r, s) e^{-h(z, r, s)} ds = e^{-h(z, r, 0)} - e^{-h(z, r, t)} \leq 1.$$

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

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



Noticing that for every  $r \in [0, 1]$  and  $t \geq 0$ ,

$$0 \leq F(z, r) - F(z, r, t) \leq \int_t^\infty \frac{f_A f_a n_A^{(z)}(s)}{f_A n_A^{(z)}(s) + f_a n_a^{(z)}(s)} ds \xrightarrow{t \rightarrow \infty} 0,$$

we get that the convergence of  $(F(z, r, t), t \geq 0)$  is uniform for  $r \in [0, 1]$ .

In the case of a soft sweep from standing variation, the selected allele gets to fixation with high probability. More precisely, it is proven in [5] that under Assumption 1,

$$(2.14) \quad \lim_{K \rightarrow \infty} \mathbb{P}(\text{Fix}^{(z, K)}) = 1, \quad \forall z \in \mathbb{R}_+^{A \times \mathcal{B}} \times (\mathbb{R}_+^{a \times \mathcal{B}} \setminus (0, 0)).$$

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

**Theorem 1.** *Let  $z$  be in  $\mathbb{R}_+^{A \times \mathcal{B}} \times (\mathbb{R}_+^{a \times \mathcal{B}} \setminus (0, 0))$  and Assumption 1 hold. Then on the fixation event  $\text{Fix}^{(z, K)}$ , the proportion of alleles  $b_1$  when the  $A$ -population becomes extinct (time  $T_{\text{ext}}^{(z, K)}$ ) converges in probability:*

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(\mathbf{1}_{\text{Fix}^{(z, K)}} \left| \mathcal{P}_{a, b_1}^{(z, K)} - \left[ \frac{z_{Ab_1}}{z_A} F(z, r_K) + \frac{z_{ab_1}}{z_a} (1 - F(z, r_K)) \right] \right| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0.$$

The neutral proportion at the end of a soft sweep from standing variation is thus a weighted mean of initial proportions in populations  $A$  and  $a$ . In particular, a soft sweep from standing variation is responsible for a diminution of the number of neutral alleles with very low or very high proportions in the population, as remarked in [35]. We notice that the weight  $F(z, r_K)$  does not depend on the initial neutral proportions. It only depends on  $r_K$  and on the dynamical system (2.4) with initial condition  $(z_A, z_a)$ . The proof consists in comparing the population process with the four dimensional dynamical system,

$$(2.15) \quad \dot{n}_{\alpha\beta}^{(z, K)} = \left( f_\alpha - D_\alpha - C_{\alpha, A} n_A^{(z, K)} - C_{\alpha, a} n_a^{(z, K)} \right) n_{\alpha\beta}^{(z, K)} + r f_A f_a \frac{n_{\bar{\alpha}\beta}^{(z, K)} n_{\alpha\beta}^{(z, K)} - n_{\alpha\beta}^{(z, K)} n_{\bar{\alpha}\beta}^{(z, K)}}{f_A n_A^{(z, K)} + f_a n_a^{(z, K)}}, \quad (\alpha, \beta) \in \mathcal{E},$$

with initial condition  $n^{(z, K)}(0) = z \in \mathbb{R}_+^{\mathcal{E}}$ . Then by a change of variables, we can study the dynamical system (2.15) and prove that

$$\frac{n_{a, b_1}^{(z, K)}(\infty)}{n_a^{(z, K)}(\infty)} = \frac{z_{Ab_1}}{z_A} F(z, r_K) + \frac{z_{ab_1}}{z_a} (1 - F(z, r_K)),$$

which leads to the result.

Now we focus on hard selective sweeps: 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 2.** *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, it is proven in [5] that under Assumptions 1 and 2,

$$(2.16) \quad \lim_{K \rightarrow \infty} \mathbb{P}\left(\text{Fix}^{(z^{(K)}, K)}\right) = \frac{S_{aA}}{f_a}.$$

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

**Assumption 3.** *Strong recombination*

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

**Assumption 4.** *Weak recombination*

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

Recall (2.11) and introduce the real number

$$(2.17) \quad \rho_K := 1 - \exp\left(-\frac{f_a r_K \log K}{S_{aA}}\right).$$

Then we have the following results whose proofs are deferred to Sections 6 and 7:

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

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(\mathbf{1}_{\text{Fix}^{(z^{(K)}, K)}} \left| \mathcal{P}_{a, b_1}^{(z^{(K)}, K)} - \frac{z_{Ab_1}}{z_A} \right| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0,$$

and if Assumption 4 holds,

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(\mathbf{1}_{\text{Fix}^{(z^{(K)}, K)}} \left| \mathcal{P}_{a, b_1}^{(z^{(K)}, K)} - \left[(1 - \rho_K) + \rho_K \frac{z_{Ab_1}}{z_A}\right] \right| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0.$$

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 modify the neutral allele proportion. 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. More precisely, we will show that the probability for a neutral lineage to undergo a recombination and be descended from an individual of type  $A$  alive at the beginning of the sweep is close to  $\rho_K$ . Then to know the probability for such an allele to be a  $b_1$  or a  $b_2$ , we have to approximate the proportion of alleles  $b_1$  in the  $A$ -population when the recombination occurs. We will prove that this proportion stays close to the initial one  $z_{Ab_1}/z_A$  during the first phase. With probability  $1 - \rho_K$ , a neutral allele originates from the first mutant. In this case it is necessarily a  $b_1$ . This gives the result for the weak recombination regime. In fact the probability for a neutral lineage to undergo no recombination during the first phase is quite intuitive: broadly speaking, the probability to have no recombination at a birth event is  $1 - r_K$ , the birth rate is  $f_a$  and the duration of the first phase is  $\log K/S_{aA}$ . Hence as  $r_K$  is small for large  $K$ ,  $1 - r_K \sim \exp(-r_K)$  and the probability to undergo no recombination is approximately

$$(1 - r_K)^{f_a \log K/S_{aA}} \sim \exp(-r_K f_a \log K/S_{aA}) = 1 - \rho_K.$$

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

$$\lim_{r_K \log K \rightarrow \infty} \rho_K = 1.$$

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

**Remark 2.** *As it will appear in the proofs (see Sections 6 and 7), the final neutral proportion in the  $a$  population is already reached at the end of the first phase. In particular, the results are still valid if the sweep is not complete but the allele  $a$  only reaches a fraction  $0 < p < 1$  of the population at the end of the sweep. The fact that the final neutral proportion is mostly determined by the beginning of the sweep has already been noticed by Coop and Ralph in [9].*



## 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  $n_{\bar{a}\beta}n_{\alpha\bar{\beta}} - n_{\alpha\beta}n_{\bar{a}\bar{\beta}}$ . Proposition 1 states a semi-martingale representation of the neutral allele proportions and makes this interplay more precise.

**Proposition 1.** *Let  $(\alpha, z, K)$  be in  $\mathcal{A} \times (\mathbb{R}_+^S)^* \times \mathbb{N}$ . 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:*

$$(3.1) \quad P_{\alpha, b_1}^{(z, K)}(t) = P_{\alpha, b_1}^{(z, K)}(0) + M_{\alpha}^{(z, K)}(t) + r_K f_A f_a \int_0^t \mathbf{1}_{\{N_{\alpha}(s) \geq 1\}} \frac{N_{\bar{a}b_1}^{(z, K)}(s)N_{\alpha b_2}^{(z, K)}(s) - N_{ab_1}^{(z, K)}(s)N_{\bar{a}b_2}^{(z, K)}(s)}{(N_{\alpha}^{(z, K)}(s) + 1)(f_A N_A^{(z, K)}(s) + f_a N_a^{(z, K)}(s))} ds,$$

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

**Remark 3.** *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:*

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

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 [25] for a study of its structure around a sweep).

**Remark 4.** *By taking the expectations in Proposition 1 we can make a comparison with the results of Ohta and Kimura [29]. In their work the population size is infinite and the proportion of favorable allele  $(y_t, t \geq 0)$  evolves as a deterministic logistic curve:*

$$\frac{dy_t}{dt} = sy_t(1 - y_t).$$

Moreover,  $x_1$  and  $x_2$  denote the neutral proportions of a given allele in the selected and non selected populations respectively, and are modeled by a diffusion. By making the analogies

$$N_e(t) = N_A(t) + N_a(t), \quad y_t = \frac{N_a(t)}{N_A(t) + N_a(t)}, \quad x_1(t) = \frac{N_{ab_1}(t)}{N_a(t)}, \quad x_2(t) = \frac{N_{Ab_1}(t)}{N_A(t)},$$

where  $N_e$  is the effective population size, the results of [29] can be written

$$\frac{d\mathbb{E}[P_{\alpha, b_1}(t)]}{dt} = r \frac{\mathbb{E}[N_{\bar{a}b_1}(t)N_{\alpha b_2}(t) - N_{ab_1}(t)N_{\bar{a}b_2}(t)]}{N_{\alpha}(t)(N_A(t) + N_a(t))},$$

and

$$\frac{d\mathbb{E}[P_{\alpha, b_1}^2(t)]}{dt} = \frac{\mathbb{E}[P_{ab_1}(t)(1 - P_{\alpha b_2}(t))]}{2N_{\alpha}(t)} + 2r \frac{\mathbb{E}[N_{ab_1}(t)(N_{\bar{a}b_1}(t)N_{\alpha b_2}(t) - N_{ab_1}(t)N_{\bar{a}b_2}(t))]}{N_{\alpha}^2(t)(N_A(t) + N_a(t))}.$$

Hence the dynamics of the first moments are very similar to these that we obtain when we take equal birth rates  $f_A = f_a$  and a recombination  $r_K = r/f_a$ . In contrast, the second moments of neutral proportions are very different in the two models.

*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  $(e_{\alpha\beta}, (\alpha, \beta) \in \mathcal{E})$  the canonical basis of  $\mathbb{R}^{\mathcal{E}}$ . According to (2.3) a jump occurs at rate

$$\sum_{(\alpha, \beta) \in \mathcal{E}} (b_{\alpha\beta}^K(N) + d_{\alpha\beta}^K(N)) = f_a N_a + d_a^K(N) + f_A N_A + d_A^K(N).$$

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 [21] p. 66) yields for every function  $h$  measurable and bounded on  $\mathbb{R}_+^{\mathcal{E}}$ :

$$\begin{aligned} h(N(t)) &= h(N(0)) + \int_0^t \int_{\mathbb{R}_+} \left\{ \sum_{\alpha \in \mathcal{A}} \left( h(N(s^-) + e_{\alpha b_1}) \mathbf{1}_{0 < \theta - \mathbf{1}_{\alpha=A}(f_a N_a(s^-) + d_a^K(N(s^-)) \leq b_{ab_1}^K(N(s^-))} \right. \right. \\ &\quad + h(N(s^-) + e_{\alpha b_2}) \mathbf{1}_{b_{ab_1}^K(N(s^-)) < \theta - \mathbf{1}_{\alpha=A}(f_a N_a(s^-) + d_a^K(N(s^-)) \leq f_a N_a(s^-)} \\ &\quad + h(N(s^-) - e_{\alpha b_1}) \mathbf{1}_{0 < \theta - f_a N_a(s^-) - \mathbf{1}_{\alpha=A}(f_a N_a(s^-) + d_a^K(N(s^-)) \leq d_{ab_1}^K(N(s^-))} \\ &\quad + h(N(s^-) - e_{\alpha b_2}) \mathbf{1}_{d_{ab_1}^K(N(s^-)) < \theta - f_a N_a(s^-) - \mathbf{1}_{\alpha=A}(f_a N_a(s^-) + d_a^K(N(s^-)) \leq d_a^K(N(s^-))} \Big) \\ &\quad \left. - h(N(s^-)) \mathbf{1}_{\theta \leq f_a N_a(s^-) + d_a^K(N(s^-)) + f_A N_A(s^-) + d_A^K(N(s^-))} \right\} Q(ds, d\theta). \end{aligned} \quad (3.3)$$

Let us introduce the functions  $\mu_K^\alpha$  defined for  $\alpha \in \mathcal{A}$  and  $(s, \theta)$  in  $\mathbb{R}_+ \times \mathbb{R}_+$  by,

$$\begin{aligned} (3.4) \quad \mu_K^\alpha(N, s, \theta) &= \frac{\mathbf{1}_{N_\alpha(s) \geq 1} N_{\alpha b_2}(s)}{(N_\alpha(s) + 1) N_\alpha(s)} \mathbf{1}_{0 < \theta - \mathbf{1}_{\alpha=A}(f_a N_a(s) + d_a^K(N(s)) \leq b_{ab_1}^K(N(s))} \\ &\quad - \frac{\mathbf{1}_{N_\alpha(s) \geq 1} N_{\alpha b_1}(s)}{(N_\alpha(s) + 1) N_\alpha(s)} \mathbf{1}_{b_{ab_1}^K(N(s)) < \theta - \mathbf{1}_{\alpha=A}(f_a N_a(s) + d_a^K(N(s)) \leq f_a N_a(s)} \\ &\quad - \frac{\mathbf{1}_{N_\alpha(s) \geq 2} N_{\alpha b_2}(s)}{(N_\alpha(s) - 1) N_\alpha(s)} \mathbf{1}_{0 < \theta - f_a N_a(s) - \mathbf{1}_{\alpha=A}(f_a N_a(s) + d_a^K(N(s)) \leq d_{ab_1}^K(N(s))} \\ &\quad + \frac{\mathbf{1}_{N_\alpha(s) \geq 2} N_{\alpha b_1}(s)}{(N_\alpha(s) - 1) N_\alpha(s)} \mathbf{1}_{d_{ab_1}^K(N(s)) < \theta - f_a N_a(s) - \mathbf{1}_{\alpha=A}(f_a N_a(s) + d_a^K(N(s)) \leq d_a^K(N(s))}. \end{aligned}$$

Then we can represent the neutral allele proportions  $P_{\alpha, b_1}$  as,

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

A direct calculation gives

$$\int_0^\infty \mu_K^\alpha(N, s, \theta) d\theta = r_K f_A f_a \mathbf{1}_{\{N_\alpha(s) \geq 1\}} \frac{N_{\bar{a}b_1}(s) N_{ab_2}(s) - N_{ab_1}(s) N_{\bar{a}b_2}(s)}{(N_\alpha(s) + 1)(f_A N_A(s) + f_a N_a(s))}.$$

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

$$\begin{aligned} M_\alpha(t) &:= \int_0^t \int_0^\infty \mu_K^\alpha(N, s^-, \theta) \tilde{Q}(ds, d\theta) \\ &= P_{\alpha, b_1}(t) - P_{\alpha, b_1}(0) - r_K f_A f_a \int_0^t \mathbf{1}_{\{N_\alpha(s) \geq 1\}} \frac{N_{\bar{a}b_1}(s) N_{ab_2}(s) - N_{ab_1}(s) N_{\bar{a}b_2}(s)}{(N_\alpha(s) + 1)(f_A N_A(s) + f_a N_a(s))} ds \end{aligned}$$

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

$$(3.6) \quad \sup_{s \leq t} \left| r_K f_A f_a \int_0^s \mathbf{1}_{\{N_\alpha \geq 1\}} \frac{N_{\bar{a}b_1} N_{ab_2} - N_{ab_1} N_{\bar{a}b_2}}{(N_\alpha + 1)(f_A N_A + f_a N_a)} \right| \leq r_K f_a t \leq f_a t, \quad t \geq 0.$$

Thus  $M_\alpha$  is a square integrable pure jump martingale bounded on every finite interval with quadratic variation

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

This ends the proof of Proposition 1.  $\square$

**Remark 5.** By definition of the functions  $\mu_K^\alpha$  in (3.4) we have for all  $(s, \theta)$  in  $\mathbb{R}_+ \times \mathbb{R}_+$ ,

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

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

$$(3.9) \quad I_\varepsilon^K := \left[ K \left( \bar{n}_A - 2\varepsilon \frac{C_{A, a}}{C_{A, A}} \right), K \left( \bar{n}_A + 2\varepsilon \frac{C_{A, a}}{C_{A, A}} \right) \right] \cap \mathbb{N},$$

and the stopping times  $T_\varepsilon^K$  and  $S_\varepsilon^K$ , which denote respectively the hitting time of  $[\varepsilon K]$  by the mutant population and the exit time of  $I_\varepsilon^K$  by the resident population,

$$(3.10) \quad T_\varepsilon^K := \inf \left\{ t \geq 0, N_a^K(t) = \lfloor \varepsilon K \rfloor \right\}, \quad S_\varepsilon^K := \inf \left\{ t \geq 0, N_A^K(t) \notin I_\varepsilon^K \right\}.$$

Finally we introduce a constant depending on  $\alpha \in \mathcal{A}$  and  $v \in \mathbb{R}_+^*$ ,

$$(3.11) \quad C(\alpha, v) := 4D_\alpha + 2f_\alpha + 4(C_{\alpha, A} + C_{\alpha, a})v.$$

**Lemma 3.1.** For  $v < \infty$  and  $t \geq 0$  such that  $(N_A^{(z, K)}(t), N_a^{(z, K)}(t)) \in [0, vK]^2$ ,

$$(3.12) \quad \frac{d}{dt} \langle M_\alpha^{(z, K)} \rangle_t = \int_0^\infty \left( \mu_K^\alpha(N^{(z, K)}, t, \theta) \right)^2 d\theta \leq C(\alpha, v) \frac{\mathbf{1}_{N_\alpha(t) \geq 1}}{N_\alpha(t)}, \quad \alpha \in \mathcal{A}.$$

Moreover, under Assumptions 1 and 2, there exist  $k_0 \in \mathbb{N}$ ,  $\varepsilon_0 > 0$  and a pure jump martingale  $\bar{M}$  such that for  $\varepsilon \leq \varepsilon_0$  and  $t \geq 0$ ,

$$(3.13) \quad e^{\frac{S_{aA}}{2(k_0+1)} t \wedge T_\varepsilon^K \wedge S_\varepsilon^K} \int_0^\infty \left( \mu_K^a(N^{(z^{(K)}, K)}, t \wedge T_\varepsilon^K \wedge S_\varepsilon^K, \theta) \right)^2 d\theta \leq (k_0 + 1) C(a, 2\bar{n}_A) \bar{M}_{t \wedge T_\varepsilon^K \wedge S_\varepsilon^K},$$

and

$$(3.14) \quad \mathbb{E} \left[ \bar{M}_{t \wedge T_\varepsilon^K \wedge S_\varepsilon^K} \right] \leq \frac{1}{k_0 + 1}.$$

*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  $\varepsilon$  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  $t < T_\varepsilon^K \wedge S_\varepsilon^K$ ,

$$(3.15) \quad \frac{f_a N_a(t) (N_a(t) + k_0 - 1) - d_a^K(N(t)) (N_a(t) + k_0 + 1)}{(N_a(t) + k_0 - 1) (N_a(t) + k_0 + 1)} \geq \frac{S_{aA} N_a(t)}{2(N_a(t) + k_0 + 1)} \geq \frac{S_{aA} \mathbf{1}_{N_a(t) \geq 1}}{2(k_0 + 1)},$$

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

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

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

$$(3.16) \quad X(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K) = \bar{M}(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K) + \int_0^{t \wedge T_\varepsilon^K \wedge S_\varepsilon^K} \left( \frac{S_{aA}}{2(k_0+1)} - \frac{f_a N_a(s)(N_a(s) + k_0 - 1) - d_a^K(N(s))(N_a(s) + k_0 + 1)}{(N_a(s) + k_0 - 1)(N_a(s) + k_0 + 1)} \right) X(s) ds,$$

where the martingale  $\bar{M}$  has the following expression:

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

Thanks to (3.15) the integral in (3.16) is nonpositive. Moreover, according to (3.12), for  $t \leq T_\varepsilon^K \wedge S_\varepsilon^K$ , as  $2\varepsilon C_{A,a}/C_{A,A} \leq \bar{n}_A$  for  $\varepsilon$  small enough,

$$(3.18) \quad \int_0^\infty \left( \mu_K^a(N^{(z^{(K)}, K)}, t, \theta) \right)^2 d\theta \leq C(a, 2\bar{n}_A) \frac{\mathbf{1}_{N_a(t) \geq 1}}{N_a(t)} \leq (k_0+1)C(a, 2\bar{n}_A)X(t) \exp\left(-\frac{S_{aA}t}{2(k_0+1)}\right),$$

which ends the proof.  $\square$

#### 4. PROOF OF THEOREM 1

In this section we suppose that Assumption 1 holds. For  $\varepsilon \leq C_{a,a}/C_{A,A} \wedge 2|S_{Aa}|/C_{A,a}$  and  $z$  in  $\mathbb{R}_+^{A \times \mathcal{B}} \times (\mathbb{R}_+^{a \times \mathcal{B}} \setminus (0,0))$  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  $(0, \bar{n}_a)$ :

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

Once  $(n_A^{(z)}, n_a^{(z)})$  has reached the set  $[0, \varepsilon^2/2] \times [\bar{n}_a - \varepsilon/2, \infty)$  it never escapes from it. Moreover, according to Assumption 1 on the stable equilibrium,  $t_\varepsilon(z)$  is finite.

First we compare the population process with the four dimensional dynamical system (2.15) 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.** Recall that  $n^{(z,K)} = (n_{\alpha\beta}^{(z,K)}, (\alpha, \beta) \in \mathcal{E})$  is the solution of the dynamical system (2.15) with initial condition  $z$ . Then we have the following comparison result:

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

$$(4.2) \quad \lim_{K \rightarrow \infty} \sup_{s \leq t_\varepsilon(z)} \|N^{(z,K)}(s)/K - n^{(z,K)}(s)\| = 0 \quad \text{in probability}$$

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

*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

$$(4.3) \quad \tilde{b}_{\alpha\beta}^K(n) = \frac{1}{K} b_{\alpha\beta}^K(Kn) = f_a n_{\alpha\beta} + r_K f_a f_A \frac{n_{\bar{\alpha}\beta} n_{\alpha\bar{\beta}} - n_{\alpha\beta} n_{\bar{\alpha}\bar{\beta}}}{f_A n_A + f_a n_a}, \quad (\alpha, \beta) \in \mathcal{E}, n \in N^{\mathcal{E}},$$

and

$$(4.4) \quad \tilde{d}_{\alpha\beta}^K(n) = \frac{1}{K} d_{\alpha\beta}^K(Kn) = [D_\alpha + C_{\alpha,A} n_A + C_{\alpha,a} n_a] n_{\alpha\beta}, \quad (\alpha, \beta) \in \mathcal{E}, n \in N^{\mathcal{E}},$$

are Lipschitz and bounded on every compact subset of  $\mathbb{N}^{\mathcal{E}}$ . The only difference with [17] is that  $\tilde{b}_{\alpha\beta}^K$  depends on  $K$  via the term  $r_K$ . Let  $(Y_i^{(\alpha\beta)}, i \in \{1, 2\}, (\alpha, \beta) \in \mathcal{E})$  be eight independent standard Poisson processes. From the representation of the population process  $N^{(z,K)}$  in (3.3) we see that the process  $(\tilde{N}^{(z,K)}(t), t \geq 0)$  defined by

$$\tilde{N}^{(z,K)}(t) = \lfloor zK \rfloor + \sum_{(\alpha,\beta) \in \mathcal{E}} \left[ Y_1^{(\alpha\beta)} \left( \int_0^t b_{\alpha\beta}^K(\tilde{N}^{(z,K)}(s)) ds \right) - Y_2^{(\alpha\beta)} \left( \int_0^t d_{\alpha\beta}^K(\tilde{N}^{(z,K)}(s)) ds \right) \right],$$

has the same law as  $(N^{(z,K)}(t), t \geq 0)$ . Applying Definitions (4.3) and (4.4) we get:

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

where we recall that  $(e_{\alpha\beta}, (\alpha, \beta) \in \mathcal{E})$  is the canonical basis of  $\mathbb{R}_+^{\mathcal{E}}$  and the martingale  $\text{Mart}^{(z,K)}$  is defined by

$$\text{Mart}^{(z,K)} := \frac{1}{K} \sum_{(\alpha,\beta) \in \mathcal{E}} \left[ \tilde{Y}_1^{(\alpha\beta)} \left( K \int_0^t \tilde{b}_{\alpha\beta}^K \left( \frac{\tilde{N}^{(z,K)}(s)}{K} \right) ds \right) - \tilde{Y}_2^{(\alpha\beta)} \left( K \int_0^t \tilde{d}_{\alpha\beta} \left( \frac{\tilde{N}^{(z,K)}(s)}{K} \right) ds \right) \right]$$

and  $(\tilde{Y}_i^{(\alpha\beta)}(u) = Y_i^{(\alpha\beta)}(u) - u, u \geq 0, i \in \{1, 2\}, (\alpha, \beta) \in \mathcal{E})$  are the Poisson processes centered at their expectation. We also have by definition

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

Hence, for every  $t \leq t_{\varepsilon}(z)$ ,

$$\begin{aligned} \left| \frac{\tilde{N}^{(z,K)}(t)}{K} - n^{(z,K)}(t) \right| &\leq \left| \frac{\lfloor zK \rfloor}{K} - z \right| + \left| \text{Mart}^{(z,K)}(t) \right| \\ &\quad + \int_0^t \sum_{(\alpha,\beta) \in \mathcal{E}} \left| \left( \tilde{b}_{\alpha\beta}^K - \tilde{d}_{\alpha\beta} \right) \left( \frac{\tilde{N}^{(z,K)}(s)}{K} \right) - \left( \tilde{b}_{\alpha\beta}^K - \tilde{d}_{\alpha\beta} \right) (n^{(z,K)}(s)) \right| ds, \end{aligned}$$

and there exists a finite constant  $\mathcal{K}$  such that

$$\left| \frac{\tilde{N}^{(z,K)}(t)}{K} - n^{(z,K)}(t) \right| \leq \frac{1}{K} + \left| \text{Mart}^{(z,K)}(t) \right| + \mathcal{K} \int_0^t \left| \frac{\tilde{N}^{(z,K)}(s)}{K} - n^{(z,K)}(s) \right| ds.$$

But following Ethier and Kurtz, we get

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

and using Gronwall's Lemma we finally obtain

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

As the convergence in law to a constant is equivalent to the convergence in probability to the same constant, the result follows.  $\square$

Once we know that the rescaled population process is close to the solution of the dynamical system (2.15), we can study the dynamical system.

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

$$\left| \frac{n_{ab_1}^{(z,K)}(\infty)}{n_a^{(z,K)}(\infty)} - \frac{n_{ab_1}^{(z,K)}(t_{\varepsilon}(z))}{n_a^{(z,K)}(t_{\varepsilon}(z))} \right| \leq \frac{2f_a \varepsilon^2}{\bar{n}_A |S_{AA}|}.$$

*Proof.* First notice that by definition of the dynamical systems (2.4) and (2.15),  $n_\alpha^{(z,K)} = n_\alpha^{(z)}$  for  $\alpha \in \mathcal{A}$  and  $z \in \mathbb{R}^\mathcal{E}$ . Assumption 1 ensures that  $n_a^{(z)}(t)$  goes to  $\bar{n}_a$  at infinity. If we define the functions

$$p_{a,b_1}^{(z,K)} = n_{a,b_1}^{(z,K)} / n_a^{(z)}, \quad \alpha \in \mathcal{A}, \quad \text{and} \quad g^{(z,K)} = p_{A,b_1}^{(z,K)} - p_{a,b_1}^{(z,K)},$$

we easily check that  $\phi : (n_{Ab_1}^{(z,K)}, n_{Ab_2}^{(z,K)}, n_{ab_1}^{(z,K)}, n_{ab_2}^{(z,K)}) \mapsto (n_A^{(z)}, n_a^{(z)}, g^{(z,K)}, p_{a,b_1}^{(z,K)})$  defines a change of variables from  $(\mathbb{R}_+^*)^\mathcal{E}$  to  $\mathbb{R}_+^{2*} \times ]-1, 1[ \times ]0, 1[$ , and (2.15) is equivalent to:

$$(4.5) \quad \begin{cases} \dot{n}_\alpha^{(z)} = (f_\alpha - (D_\alpha + C_{\alpha,A} n_A^{(z)} + C_{\alpha,a} n_a^{(z)})) n_\alpha^{(z)}, & \alpha \in \mathcal{A} \\ \dot{g}^{(z,K)} = -g^{(z,K)} \left( r_K f_A f_a (n_A^{(z)} + n_a^{(z)}) / (f_A n_A^{(z)} + f_a n_a^{(z)}) \right) \\ \dot{p}_{a,b_1}^{(z,K)} = g^{(z,K)} \left( r_K f_A f_a n_A^{(z)} / (f_A n_A^{(z)} + f_a n_a^{(z)}) \right), \end{cases}$$

with initial condition

$$(n_A^{(z)}(0), n_a^{(z)}(0), g^{(z,K)}(0), p_{a,b_1}^{(z,K)}(0)) = (z_A, z_a, z_{Ab_1} / z_A - z_{ab_1} / z_a, z_{ab_1} / z_a).$$

Moreover, a direct integration yields

$$(4.6) \quad p_{a,b_1}^{(z,K)}(t) = p_{a,b_1}^{(z,K)}(0) - (p_{a,b_1}^{(z,K)}(0) - p_{A,b_1}^{(z,K)}(0)) F(z, r_K, t),$$

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

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

Recalling that  $r_K \leq 1$  and  $|g(t)| \leq 1$  for all  $t \geq 0$  we get:

$$(4.7) \quad \left| p_{a,b_1}^{(z,K)}(\infty) - p_{a,b_1}^{(z,K)}(t_\varepsilon(z)) \right| \leq \int_{t_\varepsilon(z)}^\infty \frac{f_A f_a n_A^{(z)}}{f_A n_A^{(z)} + f_a n_a^{(z)}} \leq \frac{f_A \varepsilon^2}{\bar{n}_a - \varepsilon/2} \int_0^\infty e^{S_{Aa}s/2} ds \leq \frac{2f_A \varepsilon^2}{(\bar{n}_a - \varepsilon/2)|S_{Aa}|},$$

which ends the proof.  $\square$

**4.2. A-population extinction.** The deterministic approximation (2.15) 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 the 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) and introduce the compact set  $\Theta$ :

$$(4.8) \quad \Theta := \{z \in \mathbb{R}_+^{A \times \mathcal{B}} \times \mathbb{R}_+^{a \times \mathcal{B}}, z_A \leq \varepsilon^2 \quad \text{and} \quad |z_a - \bar{n}_a| \leq \varepsilon\},$$

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

$$(4.9) \quad U_\varepsilon^K(z) := \inf \left\{ t \geq 0, N_A^{(z,K)}(t) > \varepsilon K \text{ or } |N_a^{(z,K)}(t) - \bar{n}_a K| > M'' \varepsilon K \right\}.$$

**Lemma 4.3.** *Let  $z$  be in  $\Theta$ . Under Assumption 1, there exist two positive finite constants  $c$  and  $\varepsilon_0$  such that for  $\varepsilon \leq \varepsilon_0$ ,*

$$\limsup_{K \rightarrow \infty} \mathbb{P} \left( \sup_{t \leq T_{\text{ext}}^{(z,K)}} \left| P_{a,b_1}^{(z,K)}(t) - P_{a,b_1}^{(z,K)}(0) \right| > \varepsilon \right) \leq c\varepsilon.$$

*Proof.* Let  $z$  be in  $\Theta$  and  $Z^1$  be a birth and death process with individual birth rate  $f_A$ , individual death rate  $D_A + (\bar{n}_a - M''\varepsilon)C_{A,a}$ , and initial state  $\lceil \varepsilon^2 K \rceil$ . Then on  $[0, U_\varepsilon^K(z)[$ ,  $N_A$  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:

$$(4.10) \quad N_A(t) \leq Z_t^1, \quad \forall t < U_\varepsilon^K(z).$$

Moreover, if we denote by  $T_0^1$  the extinction time of  $Z^1$ ,  $T_0^1 := \inf\{t \geq 0, Z_t^1 = 0\}$ , and recall that

$$(4.11) \quad f_A - D_A - (\bar{n}_a - M''\varepsilon)C_{A,a} = S_{Aa} + M''C_{A,a}\varepsilon < S_{Aa}/2 < 0, \quad \forall \varepsilon < |S_{Aa}| / (2M''C_{Aa}),$$



we get according to (A.10) that for  $z \leq \varepsilon^2$  and

$$L(\varepsilon, K) = 2 \log K / |S_{Aa} + M'' \varepsilon C_{A,a}|,$$

$$\mathbb{P}_{\lceil zK \rceil} \left( T_0^1 \leq L(\varepsilon, K) \right) \geq \exp \left( \lceil \varepsilon^2 K \rceil \left[ \log(K^2 - 1) - \log(K^2 - f_A(D_A + (\bar{n}_a - M'' \varepsilon) C_{A,a})^{-1}) \right] \right).$$

Thus:

$$(4.12) \quad \lim_{K \rightarrow \infty} \mathbb{P}_{\lceil zK \rceil} \left( T_0^1 < L(\varepsilon, K) \right) = 1.$$

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

$$(4.13) \quad \mathbb{P} \left( L(\varepsilon, K) < U_\varepsilon^K(z) \right) \geq 1 - c\varepsilon.$$

Equations (4.12) and (4.13) imply

$$(4.14) \quad \liminf_{K \rightarrow \infty} \mathbb{P} \left( T_0^1 < L(\varepsilon, K) < U_\varepsilon^K(z) \right) \geq 1 - c\varepsilon$$

for a finite  $c$  and  $\varepsilon$  small enough. According to Coupling (4.10) we have the inclusion

$$\{T_0^1 < L(\varepsilon, K) < U_\varepsilon^K(z)\} \subset \{T_{ext}^K < L(\varepsilon, K) < U_\varepsilon^K(z)\}.$$

Adding (4.14) we finally get:

$$(4.15) \quad \liminf_{K \rightarrow \infty} \mathbb{P}(T_{ext}^K < L(\varepsilon, K) < U_\varepsilon^K(z)) \geq 1 - c\varepsilon.$$

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:

$$(4.16) \quad \begin{aligned} \mathbb{P} \left( \sup_{t \leq L(\varepsilon, K) \wedge U_\varepsilon^K} |M_a(t)| > \frac{\varepsilon}{2} \right) &\leq \frac{4}{\varepsilon^2} \mathbb{E} \left[ \langle M_a \rangle_{L(\varepsilon, K) \wedge U_\varepsilon^K(z)} \right] \\ &\leq \frac{4C(a, \bar{n}_a + M'' \varepsilon) L(\varepsilon, K)}{\varepsilon^2 K (\bar{n}_a - M'' \varepsilon)} \\ &= \frac{8C(a, \bar{n}_a + M'' \varepsilon) \log K}{\varepsilon^2 K (\bar{n}_a - M'' \varepsilon) |S_{Aa} + M'' \varepsilon C_{A,a}|}. \end{aligned}$$

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  $t \geq 0$

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

Hence decomposition (3.1), Markov's Inequality, and Equations (4.10), (A.8) and (4.11) yield

$$(4.17) \quad \mathbb{P} \left( \left| (P_{a,b_1} - M_a)(t \wedge U_\varepsilon^K(z)) - P_{a,b_1}(0) \right| > \frac{\varepsilon}{2} \right) \leq \frac{2f_A \varepsilon^2}{\varepsilon(\bar{n}_a - \varepsilon M'')} \int_0^t e^{\frac{s_{Aa}s}{2}} ds \leq \frac{4f_A \varepsilon}{(\bar{n}_a - \varepsilon M'') |S_{Aa}|}.$$

Taking the limit of (4.16) when  $K$  goes to infinity and adding (4.17) end the proof.  $\square$

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

$$\begin{aligned} \left| P_{a,b_1}^{(z,K)}(T_{ext}^{(z,K)}) - p_{a,b_1}^{(z,K)}(\infty) \right| &\leq \left| P_{a,b_1}^{(z,K)}(T_{ext}^{(z,K)}) - P_{a,b_1}^{(z,K)}(t_\varepsilon(z)) \right| + \\ &\quad \left| P_{a,b_1}^{(z,K)}(t_\varepsilon(z)) - p_{a,b_1}^{(z,K)}(t_\varepsilon(z)) \right| + \left| p_{a,b_1}^{(z,K)}(t_\varepsilon(z)) - p_{a,b_1}^{(z,K)}(\infty) \right|. \end{aligned}$$

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

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

for a finite  $c$ , which is equivalent to the convergence in probability. Adding (4.6) completes the proof.

## 5. A COUPLING WITH TWO BIRTH AND DEATH PROCESSES

In Sections 6 and 7 we suppose that Assumptions 1 and 2 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  $T_\varepsilon^K$  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  $T_\varepsilon^K$ . 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. We recall Definition (3.10) and define for  $\varepsilon < S_{aA}/(2C_{a,A}C_{A,a}/C_{A,A} + C_{a,a})$ ,

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

Definitions (2.3) and (2.8) ensure that for  $t < T_\varepsilon^K \wedge S_\varepsilon^K$ ,

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

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

$$(5.3) \quad Z_\varepsilon^-(t) \leq N_a^K(t) \leq Z_\varepsilon^+(t), \quad \text{for all } t < T_\varepsilon^K \wedge S_\varepsilon^K,$$

where for  $* \in \{-, +\}$ ,  $Z_\varepsilon^*$  is a birth and death process with initial state 1, and individual birth and death rates  $f_a$  and  $f_a(1 - s_*(\varepsilon))$ .

## 6. PROOF OF THEOREM 2 IN THE STRONG RECOMBINATION REGIME

In this section, we suppose that Assumptions 1, 2 and 3 hold. 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  $T_\varepsilon^K$  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  $\varepsilon_0$  such that for  $\varepsilon \leq \varepsilon_0$ :*

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

*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:

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

$$(6.2) \quad Y(t) = G^2(t)e^{r_K f_a t/2}, \quad \forall t \geq 0.$$

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

$$(6.3) \quad Y(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K) = Y(0) + \hat{M}_{t \wedge T_\varepsilon^K \wedge S_\varepsilon^K} + r_K \int_0^t \mathbf{1}_{s < T_\varepsilon^K \wedge S_\varepsilon^K} (f_a/2 - H(s)) Y(s) ds \\ + \int_0^t \mathbf{1}_{s < T_\varepsilon^K \wedge S_\varepsilon^K} e^{r_K f_a s/2} ds \int_{\mathbb{R}_+} \left[ \left( \mu_A^K(N, s, \theta) \right)^2 + \left( \mu_a^K(N, s, \theta) \right)^2 \right] d\theta,$$

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

$$(6.4) \quad H(t) = \frac{2f_a f_A N_A(t) N_a(t)}{f_A N_A(t) + f_a N_a(t)} \left[ \frac{1}{N_A(t) + 1} + \frac{1}{N_a(t) + 1} \right] \geq \frac{f_a}{2}, \quad t < T_\varepsilon^K \wedge S_\varepsilon^K,$$

for  $\varepsilon$  small enough. In particular the first integral in (6.3) is non-positive. Applying Lemma 3.1 we obtain:

$$(6.5) \quad \begin{aligned} \mathbb{E}[Y(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K)] &\leq 1 + \frac{2C(A, 2\bar{n}_A)}{r_K f_a (\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A}) K} e^{\frac{r_K f_a t}{2}} \\ &\quad + \int_0^t (k_0 + 1) C(a, 2\bar{n}_A) \mathbb{E}[\tilde{M}_{s \wedge T_\varepsilon^K \wedge S_\varepsilon^K}] e^{(\frac{r_K f_a}{2} - \frac{S_{aA}}{2(k_0+1)})s} ds \\ &\leq c \left( 1 + \frac{1}{K r_K} e^{\frac{r_K f_a t}{2}} + e^{(\frac{r_K f_a}{2} - \frac{S_{aA}}{2(k_0+1)})t} \right), \end{aligned}$$

where  $c$  is a finite constant which can be chosen independently of  $\varepsilon$  and  $K$  if  $\varepsilon$  is small enough and  $K$  large enough. Combining the semi-martingale decomposition (3.1), the Cauchy-Schwarz Inequality, and Equations (3.12) and (6.5) we get for every  $t \geq 0$ ,

$$\begin{aligned} &\mathbb{E} \left[ \left| P_{A,b_1}(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K) - \frac{\lfloor z_{Ab_1} K \rfloor}{\lfloor z_A K \rfloor} \right| \right] \\ &\leq \mathbb{E} \left[ |M_A(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K)| \right] + \frac{r_K f_a \varepsilon}{\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A}} \int_0^t \mathbb{E} \left[ \mathbf{1}_{s < T_\varepsilon^K \wedge S_\varepsilon^K} |G(s)| \right] ds \\ &\leq \mathbb{E}^{1/2} \left[ \langle M_A \rangle_{t \wedge T_\varepsilon^K \wedge S_\varepsilon^K} \right] + c r_K \varepsilon \int_0^t \mathbb{E}^{1/2} \left[ Y(s \wedge T_\varepsilon^K \wedge S_\varepsilon^K) \right] e^{-r_K f_a s/4} ds \\ &\leq c \left( \sqrt{t/K} + \varepsilon r_K \int_0^t \left( e^{-r_K f_a s/2} + \frac{1}{K r_K} + e^{-S_{aA} s/2(k_0+1)} \right)^{1/2} ds \right), \end{aligned}$$

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

$$(6.6) \quad \mathbb{E} \left[ \left| P_{A,b_1}(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K) - \frac{\lfloor z_{Ab_1} K \rfloor}{\lfloor z_A K \rfloor} \right| \right] \leq c \left( \sqrt{\frac{t}{K}} + \varepsilon \left( 1 + \frac{t}{\sqrt{K}} \right) \right).$$

Let us introduce the sequences of times

$$t_K^{(-)} = (1 - c_1 \varepsilon) \frac{\log K}{S_{aA}}, \quad \text{and} \quad t_K^{(+)} = (1 + c_1 \varepsilon) \frac{\log K}{S_{aA}},$$

where  $c_1$  is a finite constant. Then according to Coupling (5.3) and limit (A.11),

$$(6.7) \quad \lim_{K \rightarrow \infty} \mathbb{P}(T_\varepsilon^K < t_K^{(-)} | T_\varepsilon^K \leq S_\varepsilon^K) = \lim_{K \rightarrow \infty} \mathbb{P}(T_\varepsilon^K > t_K^{(+)} | T_\varepsilon^K \leq S_\varepsilon^K) = 0.$$

Hence applying (6.6) at time  $t_K^{(+)}$  and using (A.3) and (6.7), we bound the first term in the expectation. To bound the second term in the expectation, we introduce the notation

$$A(\varepsilon, K) := \mathbb{E} \left[ \mathbf{1}_{t_K^{(-)} \leq T_\varepsilon^K \leq S_\varepsilon^K \wedge t_K^{(+)}} \left| P_{A,b_1}^K(T_\varepsilon^K \wedge S_\varepsilon^K) - P_{a,b_1}^K(T_\varepsilon^K \wedge S_\varepsilon^K) \right| \right].$$

From (6.7) we obtain

$$(6.8) \quad \limsup_{K \rightarrow \infty} \mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \left| P_{A,b_1}^K(T_\varepsilon^K) - P_{a,b_1}^K(T_\varepsilon^K) \right| \right] = \limsup_{K \rightarrow \infty} A(\varepsilon, K),$$

and by using (6.2), the Cauchy-Schwarz Inequality, and (6.5) we get

$$\begin{aligned}
A(\varepsilon, K) &\leq \mathbb{E} \left[ \sqrt{Y(t_K^{(+)} \wedge T_\varepsilon^K \wedge S_\varepsilon^K)} \right] e^{-\frac{r_K f_a}{4} t_K^{(-)}} \\
&\leq \mathbb{E}^{1/2} \left[ Y(t_K^{(+)} \wedge T_\varepsilon^K \wedge S_\varepsilon^K) \right] e^{-\frac{r_K f_a}{4} t_K^{(-)}} \\
&\leq c \left( 1 + \frac{1}{K r_K} e^{r_K f_a t_K^{(+)} / 2} + e^{(\frac{r_K f_a}{2} - \frac{S_{AA}}{2(k_0+1)}) t_K^{(+)}} \right)^{1/2} e^{-\frac{r_K f_a}{4} t_K^{(-)}} \\
&\leq c \left( e^{-\frac{r_K f_a}{4} t_K^{(-)}} + \frac{1}{\sqrt{K r_K}} e^{\frac{c_1 \varepsilon r_K f_a \log K}{2 S_{AA}}} + e^{(\frac{c_1 \varepsilon r_K f_a \log K}{2 S_{AA}} - \frac{S_{AA}}{4(k_0+1)}) t_K^{(+)}} \right),
\end{aligned}$$

where the value of the constant  $c$  can change from line to line. Assumption 3 then yields

$$\limsup_{K \rightarrow \infty} A(\varepsilon, K) = 0,$$

and we end the proof of the second bound by applying (6.8).  $\square$

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  $\varepsilon_0$  such that for  $\varepsilon \leq \varepsilon_0$ :*

$$\limsup_{K \rightarrow \infty} \mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \left| P_{a, b_1}^K(T_\varepsilon^K + t_\varepsilon(\frac{N(T_\varepsilon^K)}{K})) - \frac{z_{Ab_1}}{z_A} \right| \right] \leq c\varepsilon.$$

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

$$(6.9) \quad \Gamma := \left\{ z \in \mathbb{R}_+^\varepsilon, \left| z_A - \bar{n}_A \right| \leq 2\varepsilon \frac{C_{A,a}}{C_{A,A}}, \left| z_a - \varepsilon \right| \leq \frac{\varepsilon}{2} \right\}, \quad t_\varepsilon := \sup \{ t_\varepsilon(z), z \in \Gamma \},$$

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

$$(6.10) \quad L_\varepsilon^K(z) = \inf \{ t \geq 0, (N_A^{(z,K)}(t), N_a^{(z,K)}(t)) \notin [I(\Gamma, \varepsilon)K/2, (\bar{n}_A + \bar{n}_a)K]^2 \}.$$

Finally, we denote by  $(\mathcal{F}_t^K, t \geq 0)$  the canonical filtration of  $N^K$ . Notice that on the event  $\{T_\varepsilon^K \leq S_\varepsilon^K\}$ ,  $N(T_\varepsilon^K)/K \in \Gamma$ , 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:

$$\begin{aligned}
(6.11) \quad &\mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \left| P_{a, b_1} \left( T_\varepsilon^K + t_\varepsilon \left( \frac{N(T_\varepsilon^K)}{K} \right) \right) \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right) - P_{a, b_1}(T_\varepsilon^K) \right| \right] \\
&\leq \mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \mathbb{E} \left[ \left| M_a \left( T_\varepsilon^K + t_\varepsilon \left( \frac{N(T_\varepsilon^K)}{K} \right) \right) \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right) - M_a(T_\varepsilon^K) \right| + f_a \int_{T_\varepsilon^K}^{T_\varepsilon^K + t_\varepsilon \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right)} |G| \Big| \mathcal{F}_{T_\varepsilon^K} \right] \right] \\
&\leq \mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \left\{ \mathbb{E}^{1/2} \left[ \langle M_a \rangle_{T_\varepsilon^K + t_\varepsilon \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right)} - \langle M_a \rangle_{T_\varepsilon^K} \Big| \mathcal{F}_{T_\varepsilon^K} \right] + f_a \sqrt{t_\varepsilon} \mathbb{E}^{1/2} \left[ \int_{T_\varepsilon^K}^{T_\varepsilon^K + t_\varepsilon \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right)} G^2 \Big| \mathcal{F}_{T_\varepsilon^K} \right] \right\} \right].
\end{aligned}$$

To bound the first term of the right hand side we use the Strong Markov Property, Equation (3.12) and the definition of  $L_\varepsilon^K$  in (6.10). We get

$$(6.12) \quad \mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \mathbb{E}^{1/2} \left[ \langle M_a \rangle_{T_\varepsilon^K + t_\varepsilon \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right)} - \langle M_a \rangle_{T_\varepsilon^K} \Big| \mathcal{F}_{T_\varepsilon^K} \right] \right] \leq \sqrt{\frac{2t_\varepsilon C(a, \bar{n}_A + \bar{n}_a)}{I(\Gamma, \varepsilon)K}}.$$

Let us now focus on the second term. Itô's formula with jumps yields for every  $t \geq 0$ ,

$$\mathbb{E} \left[ G^2 \left( t \wedge L_\varepsilon^K \left( \frac{N(0)}{K} \right) \right) \right] \leq \mathbb{E}[G^2(0)] + \mathbb{E} \left[ \langle M_A \rangle_{t \wedge L_\varepsilon^K \left( \frac{N(0)}{K} \right)} \right] + \mathbb{E} \left[ \langle M_a \rangle_{t \wedge L_\varepsilon^K \left( \frac{N(0)}{K} \right)} \right],$$

and adding the Strong Markov Property we get

$$\begin{aligned}
& \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \mathbb{E}^{1/2} \left[ \int_{T_\varepsilon^K}^{T_\varepsilon^K + t_\varepsilon \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right)} G^2 \Big| \mathcal{F}_{T_\varepsilon^K} \right] \\
& \leq \sup_{z \in \Gamma} \mathbb{E}^{1/2} \left[ \int_0^{t_\varepsilon} \left( G^2 \left( s \wedge L_\varepsilon^K \left( \frac{N(0)}{K} \right) \right) - G^2(0) \right) ds \Big| N(0) = \lfloor zK \rfloor \right] + \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \sqrt{t_\varepsilon} |G(T_\varepsilon^K)| \\
& \leq \sup_{z \in \Gamma} \left[ \int_0^{t_\varepsilon} \mathbb{E} \left[ \langle M_A \rangle_{s \wedge L_\varepsilon^K \left( \frac{N(0)}{K} \right)} + \langle M_A \rangle_{s \wedge L_\varepsilon^K \left( \frac{N(0)}{K} \right)} \Big| N(0) = \lfloor zK \rfloor \right] ds \right]^{1/2} + \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \sqrt{t_\varepsilon} |G(T_\varepsilon^K)|.
\end{aligned}$$

Using again Equation (3.12) and the definition of  $L_\varepsilon^K$  in (6.10), and adding Lemma 6.1 finally lead to

$$(6.13) \quad \mathbb{E} \left[ \mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \mathbb{E}^{1/2} \left[ \int_{T_\varepsilon^K}^{T_\varepsilon^K + t_\varepsilon \wedge L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right)} G^2 \Big| \mathcal{F}_{T_\varepsilon^K} \right] \right] \leq c \left( \frac{1}{\sqrt{K}} + \varepsilon \right),$$

for  $\varepsilon$  small enough and  $K$  large enough, where  $c$  is a finite constant. Moreover (2.5) ensures that

$$\mathbb{P} \left( T_\varepsilon^K \leq S_\varepsilon^K, L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right) \leq t_\varepsilon \left( \frac{N(T_\varepsilon^K)}{K} \right) \right) \leq \mathbb{P} \left( \frac{N(T_\varepsilon^K)}{K} \in \Theta, L_\varepsilon^K \left( \frac{N(T_\varepsilon^K)}{K} \right) \leq t_\varepsilon \left( \frac{N(T_\varepsilon^K)}{K} \right) \right) \xrightarrow{K \rightarrow \infty} 0,$$

where  $\Theta$  has been defined in (4.8). Adding Equations (6.11), (6.12), (6.13) and Lemma 6.1, we finally end the proof of Lemma 6.2.  $\square$

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

$$\begin{aligned}
(6.14) \quad \mathbb{P} \left( \mathbf{1}_{N(T_\varepsilon^K + t_\varepsilon(N(T_\varepsilon^K)/K)) \in \Theta} \left| P_{a,b_1}(T_{\text{ext}}^K) - P_{a,b_1}(T_\varepsilon^K + t_\varepsilon \left( \frac{N(T_\varepsilon^K)}{K} \right)) \right| > \sqrt{\varepsilon} \right) \\
\leq \sup_{z \in \Theta} \mathbb{P} \left( |P_{a,b_1}(T_{\text{ext}}^K) - P_{a,b_1}(0)| > \sqrt{\varepsilon} \Big| N(0) = \lfloor zK \rfloor \right).
\end{aligned}$$

But Equation (2.5) yields  $\mathbb{P}(N(T_\varepsilon^K + t_\varepsilon(N(T_\varepsilon^K)/K))/K \in \Theta | N(T_\varepsilon^K)/K \in \Gamma) \xrightarrow{K \rightarrow \infty} 1$ , and  $\{T_\varepsilon^K \leq S_\varepsilon^K\} \subset \{N(T_\varepsilon^K)/K \in \Gamma\}$ . Adding Equation (A.6) and Lemmas 4.3 and 6.2, the triangle inequality allows us to conclude that for  $\varepsilon$  small enough

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

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

## 7. PROOF OF THEOREM 2 IN THE WEAK RECOMBINATION REGIME

**7.1. Coupling with a four dimensional population process and structure of the proof.** In this section we suppose that Assumptions 1, 2 and 4 hold. To lighten the proofs of Sections 7.2 to 7.5 we introduce a coupling of the population process  $N$  with a process  $\tilde{N} = (\tilde{N}_{\alpha\beta}, (\alpha, \beta) \in \mathcal{E})$  defined as follows for every  $t \geq 0$ :

$$\begin{aligned}
(7.1) \quad \tilde{N}(t) &= \mathbf{1}_{t < S_\varepsilon^K} N(t) + \mathbf{1}_{t \geq S_\varepsilon^K} \left( e_{Ab_1} N_{Ab_1}((S_\varepsilon^K)^-) + e_{Ab_2} N_{Ab_2}((S_\varepsilon^K)^-) \right. \\
&\quad + \int_0^t \int_{R_+} \left\{ e_{ab_1} \mathbf{1}_{\theta \leq b_{ab_1}^K(\tilde{N}(s^-))} + e_{ab_2} \mathbf{1}_{b_{ab_1}^K(\tilde{N}(s^-)) < \theta \leq f_a \tilde{N}_a(s^-)} \right. \\
&\quad \quad \left. - e_{ab_1} \mathbf{1}_{0 < \theta - f_a \tilde{N}_a(s^-) \leq d_{ab_1}^K(\tilde{N}(s^-))} \right. \\
(7.2) \quad &\quad \left. \left. - e_{ab_2} \mathbf{1}_{d_{ab_1}^K(\tilde{N}(s^-)) < \theta - f_a \tilde{N}_a(s^-) \leq d_a^K(\tilde{N}(s^-))} \right\} Q(ds, d\theta) \right),
\end{aligned}$$

where the Poisson random measure  $Q$  has been introduced in (3.3). From (A.5) we know that

$$(7.3) \quad \limsup_{K \rightarrow \infty} \mathbb{P}(\{\exists t \leq T_\varepsilon^K, N(t) \neq \tilde{N}(t)\}, T_\varepsilon^K < \infty) \leq c\varepsilon.$$

Hence we will study the process  $\tilde{N}$  and deduce from this study properties of the dynamics of the process  $N$  during the first phase. Moreover, as we want to prove convergences on the fixation event  $\text{Fix}^K$ , defined in (2.9), inequalities (A.6) and (7.3) allow us, to study the dynamics of  $\tilde{N}$  during the first phase, to restrict our attention to the conditional probability measure:

$$(7.4) \quad \hat{\mathbb{P}}(\cdot) = \mathbb{P}(\cdot | \tilde{T}_\varepsilon^K < \infty),$$

where  $\tilde{T}_\varepsilon^K$  is the hitting time of  $\lfloor \varepsilon K \rfloor$  by the process  $\tilde{N}_a$ :

$$(7.5) \quad \tilde{T}_\varepsilon^K := \inf \left\{ t \geq 0, \tilde{N}_a^K(t) = \lfloor \varepsilon K \rfloor \right\}.$$

Expectations and variances associated with this probability measure are denoted by  $\hat{\mathbb{E}}$  and  $\hat{\text{Var}}$  respectively.

Let us notice that, as by definition  $\tilde{N}_A(t) \in I_\varepsilon^K$  for all  $t \geq 0$ , Coupling (5.3) with birth and death processes  $Z_\varepsilon^-$  and  $Z_\varepsilon^+$  holds up to time  $\tilde{T}_\varepsilon^K$  for the process  $\tilde{N}$ :

$$(7.6) \quad Z_\varepsilon^-(t) \leq \tilde{N}_a^K(t) \leq Z_\varepsilon^+(t), \quad \text{for all } t < \tilde{T}_\varepsilon^K.$$

The sketch of the proof is the following. We first focus on the neutral proportion in the  $a$  population at time  $\tilde{T}_\varepsilon^K$ . The idea is to consider the neutral alleles of the  $a$  individuals at time  $\tilde{T}_\varepsilon^K$  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 events can happen to a neutral lineage: coalescences and m-recombinations (see Section 7.2); 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  $\rho_K$ . 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 approach presented by Schweinsberg and Durrett in [36]. In this paper, the authors described the population dynamics by a variation of Moran model with two loci and recombinations. In their model, the population size was constant and each individual has a constant selective advantage, 0 or  $s$ . In our model the size is varying and each individual's ability to survive and give birth 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.

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

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

Let us sample two individuals with the  $a$  allele uniformly at random at time  $\tilde{T}_\varepsilon^K$  and denote by  $\beta_p$  and  $\beta_q$  their neutral alleles. We want to follow their genealogy backward in time and know at each time between 0 and  $\tilde{T}_\varepsilon^K$  the types ( $A$  or  $a$ ) of the individuals carrying  $\beta_p$  and  $\beta_q$ .

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

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

We say that  $\beta_p$  m-recombines at time  $\tau_m^K$  if the individual carrying the allele  $\beta_p$  at time  $\tau_m^K$  is a newborn, carries the allele  $\alpha \in \mathcal{A}$ , and has inherited his allele  $\beta_p$  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:

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

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

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

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

$$(7.7) \quad p_{\alpha\alpha}^{c_K}(n) = \frac{2}{n_\alpha(n_\alpha + 1)} \left( 1 - \frac{r_K f_{\bar{a}} n_{\bar{a}}}{f_A n_A + f_a n_a} \right) \quad \text{and} \quad p_{\alpha\bar{a}}^{c_K}(n) = \frac{r_K f_{\bar{a}}}{(n_\alpha + 1)(f_A n_A + f_a n_a)}.$$

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

**Remark 6.** *A m-recombination for a neutral allele associated with an  $\alpha$  allele is a coalescence with an  $\bar{\alpha}$  individual. Thus if we denote by  $p_\alpha^{r_K}(n)$  the probability that an  $\alpha$ -individual, chosen uniformly at time  $\tau_m^K$ , is the newborn and underwent a m-recombination at his birth, conditionally on  $(\tilde{N}_A^K(\tau_{m-1}^K), \tilde{N}_a^K(\tau_{m-1}^K)) = n \in \mathbb{N}^2$  and on the birth of an individual  $\alpha$  at time  $\tau_m^K$  we get*

$$(7.8) \quad p_\alpha^{r_K}(n) = n_{\bar{a}} p_{\alpha\bar{a}}^{c_K}(n) = \frac{n_{\bar{a}} r_K f_{\bar{a}}}{(n_\alpha + 1)(f_A n_A + f_a n_a)}.$$

Moreover, if we recall the definition of  $I_\varepsilon^K$  in (3.9), we notice that there exists a finite constant  $c$  such that for  $k < \lfloor \varepsilon K \rfloor$ ,

$$(7.9) \quad (1 - c\varepsilon) \frac{r_K}{k+1} \leq \inf_{n_A \in I_\varepsilon^K} p_a^{r_K}(n_A, k) \leq \sup_{n_A \in I_\varepsilon^K} p_a^{r_K}(n_A, k) \leq \frac{r_K}{k+1}.$$

**7.3. 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 uniformly sampled neutral alleles  $\beta_p$  and  $\beta_q$ . By definition, these events can only occur at a birth time. Thus we need to study the upcrossing number of the process  $\tilde{N}_a^K$  before  $\tilde{T}_\varepsilon^K$  (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 is 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 (7.31)). 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  $\zeta_k^K$  the jump number of last visit to  $k$  before the hitting of  $\lfloor \varepsilon K \rfloor$ ,

$$(7.10) \quad \zeta_k^K := \sup\{m \leq J^K, \tilde{N}_a^K(\tau_m^K) = k\}, \quad 1 \leq k \leq \lfloor \varepsilon K \rfloor.$$

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

$$(7.11) \quad U_{j,k}^{(K,1)} := \#\{m \in \{0, \dots, \zeta_j^K - 1\}, (\tilde{N}_a^K(\tau_m^K), \tilde{N}_a^K(\tau_{m+1}^K)) = (k, k+1)\},$$

$$(7.12) \quad U_{j,k}^{(K,2)} := \#\{m \in \{\zeta_j^K, \dots, J^K - 1\}, (\tilde{N}_a^K(\tau_m^K), \tilde{N}_a^K(\tau_{m+1}^K)) = (k, k+1)\}.$$

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  $\tilde{T}_\varepsilon^K$ :

$$(7.13) \quad H_k^K := \#\{m < J^K, \tilde{N}_a^K(\tau_m^K) = \tilde{N}_a^K(\tau_{m+1}^K) = k\},$$

$$(7.14) \quad U_k^K := U_{j,k}^{(K,1)} + U_{j,k}^{(K,2)} = \#\{m < J^K, (\tilde{N}_a^K(\tau_m^K), \tilde{N}_a^K(\tau_{m+1}^K)) = (k, k+1)\}.$$

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

$$(7.15) \quad \lambda_\varepsilon := \frac{(1 - s_-(\varepsilon))^3}{(1 - s_+(\varepsilon))^2},$$

which belongs to  $(0, 1)$  for  $\varepsilon$  small enough, we have

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

$$(7.16) \quad \hat{\mathbb{E}}[H_j^K] \leq \frac{12f_A \bar{n}_A K}{s_-^4(\varepsilon)f_A j}, \quad \hat{\mathbb{E}}[(U_{j,k}^{(K,1)})^2] \leq \frac{4\lambda_\varepsilon^{k-j}}{s_-^7(\varepsilon)(1 - s_+(\varepsilon))},$$

$$(7.17) \quad \hat{\mathbb{E}}[(U_j^K)^2] \leq \frac{2}{s_-^2(\varepsilon)}, \quad \left| \hat{\text{Cov}}(U_{j,k}^{(K,2)}, U_j^K) \right| \leq c(\varepsilon + (1 - s_-(\varepsilon))^{k-j}),$$

and

$$(7.18) \quad r_K \left| \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{\hat{\mathbb{E}}[U_k^K]}{k+1} - \frac{f_A \log K}{S_{AA}} \right| \leq c\varepsilon.$$

This Lemma is widely used in Sections 7.4 and 7.5. Indeed, we shall decompose on the possible states of the population when a birth occurs, and apply Equations (7.7) and (7.8) 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.

**7.4. 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  $\varepsilon_0$  such that for  $K \in \mathbb{N}$  and  $\varepsilon \leq \varepsilon_0$ ,*

$$\hat{\mathbb{P}}\left(TR_2^K(\beta_p) \neq -\infty\right) \leq \frac{c}{\log K}, \quad \text{and} \quad \hat{\mathbb{P}}\left(0 \leq TR_1^K(\beta_p) \leq TC^K(\beta_p, \beta_q)\right) \leq \frac{c}{\log K}.$$

*Proof.* By definition, the neutral allele  $\beta_p$  is associated with an allele  $a$  at time  $\tilde{T}_\varepsilon^K$ . If there are at least two m-recombinations it implies that there exists a time between 0 and  $\tilde{T}_\varepsilon^K$  at which  $\beta_p$  has undergone a m-recombination when it was associated with an allele  $A$ . We shall work conditionally on the stopped process  $((\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K)), m \leq J^K)$  and decompose according to the  $a$ -population size when this m-recombination occurs. We get the inclusion:

$$\{TR_2^K(\beta_p) \neq -\infty\} \subset \bigcup_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \bigcup_{m=1}^{J^K} \left\{ TR_2^K(\beta_p) = m, \tilde{N}_a(\tau_{m-1}^K) = \tilde{N}_a(\tau_m^K) = k \right\}.$$

We recall the definition of  $I_\varepsilon^K$  in (3.9). Thanks to Equations (7.8) and (7.16), we get:

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

Assumption 4 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, for  $(\alpha, \alpha') \in \mathcal{A}^2, m \leq J^K$  the notations

$$(\alpha\beta_p)_m := \{\text{the neutral allele } \beta_p \text{ is associated with the allele } \alpha \text{ at time } \tau_m^K\},$$

$$(\alpha\beta_p, \alpha'\beta_q)_m := (\alpha\beta_p)_m \cap (\alpha'\beta_q)_m.$$

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

$$\{TC^K(\beta_p, \beta_q) \neq -\infty, (A\beta_p)_{TC^K(\beta_p, \beta_q)}\} \subset \bigcup_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \bigcup_{m=1}^{J^K} \left\{ TC^K(\beta_p, \beta_q) = m, \tilde{N}_a(\tau_{m-1}^K) = k, (A\beta_p)_m \right\},$$

and decompose on the possible selected alleles associated with  $\beta_q$  and on the type of the newborn at the coalescence time. Using Lemma 7.1, Equations (7.16) and (7.17), and  $r_K \leq 1$ , we get

$$(7.19) \quad \hat{\mathbb{P}}(TC^K(\beta_p, \beta_q) \neq -\infty, (A\beta_p)_{TC^K(\beta_p, \beta_q)}) \leq \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \left[ \sup_{n_A \in I_\varepsilon^K} p_{AA}^{c_K}(n_A, k) + \sup_{n_A \in I_\varepsilon^K} p_{Aa}^{c_K}(n_A, k) \right] \hat{\mathbb{E}}[H_k^K] + \sup_{n_A \in I_\varepsilon^K} p_{aA}^{c_K}(n_A, k) \hat{\mathbb{E}}[U_k^K] \leq \frac{c}{K} \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{1}{k},$$

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

*Second step:* Then, we focus on the case where  $\beta_p$  and  $\beta_q$  are associated with an allele  $a$  at the coalescence time. The inclusion

$$\left\{ \tilde{N}_a\left(\tau_{TC^K(\beta_p, \beta_q)-1}^K\right) = k, (a\beta_p, a\beta_q)_{TC^K(\beta_p, \beta_q)} \right\} \subset \bigcup_{m=1}^{J^K} \left\{ TC^K(\beta_p, \beta_q) = m, \tilde{N}_a(\tau_{m-1}^K) = k, (a\beta_p, a\beta_q)_m \right\},$$

and Equations (7.7) and (7.17) yield for every  $k \in \{1, \dots, \lfloor \varepsilon K \rfloor - 1\}$ :

$$\hat{\mathbb{P}}\left(\tilde{N}_a\left(\tau_{TC^K(\beta_p, \beta_q)-1}^K\right) = k, (a\beta_p, a\beta_q)_{TC^K(\beta_p, \beta_q)}\right) \leq \sup_{n_A \in I_\varepsilon^K} p_{aa}^{c_K}(n_A, k) \hat{\mathbb{E}}[U_k^K] \leq \frac{4}{s_-^2(\varepsilon)k(k+1)}.$$

If  $\beta_p$  and  $\beta_q$  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  $\zeta_k^K$ -th jump when we look forward in time. For  $k, l < \lfloor \varepsilon K \rfloor$ ,

$$\begin{aligned} \hat{\mathbb{P}}\left(\tilde{N}_a\left(\tau_{TR_1^K(\beta_p)}^K\right) = l, 0 \leq TR_1^K(\beta_p) \leq TC^K(\beta_p, \beta_q) \mid \tilde{N}_a\left(\tau_{TC^K(\beta_p, \beta_q)-1}^K\right) = k, (a\beta_p, a\beta_q)_{TC^K(\beta_p, \beta_q)}\right) \\ \leq \sup_{n_A \in I_\varepsilon^K} p_a^{r_K}(n_A, l) \left( \mathbf{1}_{k>l} \hat{\mathbb{E}}[U_l^K] + \mathbf{1}_{k \leq l} \hat{\mathbb{E}}[U_{k,l}^{(K,1)}] \right) \leq \frac{2r_K}{(l+1)s_-^2(\varepsilon)} \left( \mathbf{1}_{k>l} + \frac{2\mathbf{1}_{k \leq l} \lambda_\varepsilon^{l-k}}{s_-^5(\varepsilon)(1-s_+(\varepsilon))} \right), \end{aligned}$$

where the last inequality is a consequence of (7.9), (7.16) and (7.17). The two last equations finally yield the existence of a finite  $c$  such that for every  $K \in \mathbb{N}$ :

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

which completes the proof of Lemma 7.3 with Assumption 4.  $\square$

**7.5. Probability to be descended from the first mutant.** We want to estimate the probability for the neutral lineage of  $\beta_p$  to undergo no m-recombination. Recall Definition (2.17):

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

$$\limsup_{K \rightarrow \infty} \left| \hat{\mathbb{P}}(TR_1^K(\beta_p) = -\infty) - (1 - \rho_K) \right| \leq c\varepsilon^{1/2}.$$

*Proof.* We introduce  $\rho_m^K$ , the conditional probability that the neutral lineage of  $\beta_p$  m-recombines at time  $\tau_m^K$ , given  $(\tilde{N}_A(\tau_n^K), \tilde{N}_a(\tau_n^K), n \leq J^K)$  and given that it has not m-recombined during the time interval  $]\tau_m^K, T_\varepsilon^K]$ . The last condition implies that  $\beta_p$  is associated with an allele  $a$  at time  $\tau_m^K$ .

$$(7.20) \quad \rho_m^K := \mathbf{1}_{\{\tilde{N}_a^K(\tau_m^K) - \tilde{N}_a^K(\tau_{m-1}^K) = 1\}} p_a^{r_K}(\tilde{N}_A^K(\tau_{m-1}^K), \tilde{N}_a^K(\tau_{m-1}^K)).$$

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

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

We want to give a rigorous meaning to the sequence of equivalencies:

$$\hat{\mathbb{P}}\left(TR_1^K(\beta_p) = -\infty \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}\right) = \prod_{m=1}^{J^K} (1 - \rho_m^K) \sim \prod_{m=1}^{J^K} e^{-\rho_m^K} \sim e^{-\mathbb{E}[\eta^K]},$$

when  $K$  goes to infinity. Jensen's Inequality, the triangle inequality, and the Mean Value Theorem imply

$$(7.21) \quad \begin{aligned} & \mathbb{E} \left| \hat{\mathbb{P}}\left(TR_1^K(\beta_p) = -\infty \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}\right) - (1 - \rho_K) \right| \leq \\ & \mathbb{E} \left| \hat{\mathbb{P}}\left(TR_1^K(\beta_p) = -\infty \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}\right) - e^{-\eta^K} \right| + \left| e^{-\mathbb{E}[\eta^K]} - (1 - \rho_K) \right| + \mathbb{E} \left| \eta^K - \mathbb{E}[\eta^K] \right|. \end{aligned}$$

We aim to bound the right hand side of (7.21). The bounding of the first term follows the method developed in Lemma 3.6 in [36]. We refer to this proof, and get the following Poisson approximation

$$(7.22) \quad \begin{aligned} \mathbb{E} \left| \hat{\mathbb{P}}\left(TR_1^K(\beta_p) = -\infty \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}\right) - e^{-\eta^K} \right| & \leq \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \sup_{n_A \in I_\varepsilon^K} \left( p_a^{r_K}(n_A, k) \right)^2 \mathbb{E}[U_k^K] \\ & \leq \frac{\pi^2 r_K^2}{3s_-^2(\varepsilon)}, \end{aligned}$$

where  $I_\varepsilon^K$  has been defined in (3.9) and the last inequality follows from (7.9) and (7.17). To bound the second term, we need to estimate  $\mathbb{E}[\eta^K]$ . Inequality (7.9) implies

$$(7.23) \quad (1 - c\varepsilon)r_K \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{U_k^K}{k+1} \leq \eta^K \leq r_K \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{U_k^K}{k+1}.$$

Adding (7.18) we get that for  $\varepsilon$  small enough,

$$(7.24) \quad \limsup_{K \rightarrow \infty} \left| \exp(-\mathbb{E}[\eta^K]) - (1 - \rho_K) \right| \leq c\varepsilon.$$

The bounding of the last term of (7.21) 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 (7.17). We observe that  $\mathbb{E}|\eta^K - \mathbb{E}[\eta^K]| \leq (\hat{\text{Var}} \eta^K)^{1/2}$ , but the variance of  $\eta^K$  is quite involved to study and according to Assumption 4 and Equations (7.23) and (7.17),

$$(7.25) \quad \begin{aligned} \left| \hat{\text{Var}} \eta^K - \hat{\text{Var}} \left( r_K \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{U_k^K}{k+1} \right) \right| & \leq c\varepsilon \mathbb{E} \left[ \left( r_K \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{U_k^K}{k+1} \right)^2 \right] \\ & \leq c\varepsilon r_K^2 \sum_{k,l=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{\mathbb{E}[(U_k^K)^2] + \mathbb{E}[(U_l^K)^2]}{(k+1)(l+1)} \leq c\varepsilon, \end{aligned}$$

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

$$\left| \widehat{\text{Cov}}(U_k^K, U_l^K) \right| \leq \left( \widehat{\mathbb{E}}[(U_k^K)^2] \widehat{\mathbb{E}}[(U_{k,l}^{(K,1)})^2] \right)^{1/2} + \left| \widehat{\text{Cov}}(U_k^K, U_{k,l}^{(K,2)}) \right|.$$

Applying Inequalities (7.16) and (7.17) and noticing that  $(1 - s_-(\varepsilon)) < \lambda_\varepsilon^{1/2} < 1$  (recall the definition of  $\lambda_\varepsilon$  in (7.15)) lead to

$$(7.26) \quad \left| \widehat{\text{Cov}}(U_k^K, U_l^K) \right| \leq c(\lambda_\varepsilon^{(l-k)/2} + \varepsilon + (1 - s_-(\varepsilon))^{l-k}) \leq c(\lambda_\varepsilon^{(l-k)/2} + \varepsilon)$$

for a finite  $c$  and  $\varepsilon$  small enough. We finally get:

$$(7.27) \quad \begin{aligned} \text{Var} \left( r_K \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{U_k^K}{k+1} \right) &\leq 2r_K^2 \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{1}{k+1} \sum_{l=k}^{\lfloor \varepsilon K \rfloor - 1} \frac{\widehat{\text{Cov}}(U_k^K, U_l^K)}{l+1} \\ &\leq cr_K^2 \sum_{k=1}^{\lfloor \varepsilon K \rfloor - 1} \frac{1}{k+1} \sum_{l=k}^{\lfloor \varepsilon K \rfloor - 1} \frac{\lambda_\varepsilon^{(l-k)/2} + \varepsilon}{l+1} \leq cr_K^2 \varepsilon \log^2 K, \end{aligned}$$

where we used (7.26) for the second inequality. Applying Jensen's Inequality to the left hand side of (7.21) and adding Equations (7.22), (7.24), (7.25) and (7.27) we obtain

$$(7.28) \quad \begin{aligned} \widehat{\mathbb{E}} \left| \widehat{\mathbb{P}} \left( TR_1^K(\beta_p) = -\infty \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K} \right) - (1 - \rho_K) \right| \\ \leq \frac{\pi^2 r_K^2}{3s_-^2(\varepsilon)} + c\varepsilon + c(\varepsilon + r_K^2 \varepsilon \log^2 K)^{1/2}. \end{aligned}$$

This completes the proof of Lemma 7.4.  $\square$

We finally focus on the dependence between the genealogies of  $\beta_p$  and  $\beta_q$ , and to this aim follow [36] pp. 1622 to 1624 in the case  $J = 1$ . We define for  $m \leq J^K$  the random variable

$$K_m = \mathbf{1}_{\{TR_1^K(\beta_p) \geq m\}} + \mathbf{1}_{\{TR_1^K(\beta_q) \geq m\}},$$

which counts the number of neutral lineages which recombine after the  $m$ -th jump (forward in time) among the lineages of  $\beta_p$  and  $\beta_q$ . First we will show that for  $d \in \{0, 1, 2\}$ ,

$$(7.29) \quad \left| \widehat{\mathbb{P}}(K_0 = d) - \binom{d}{2} \widehat{\mathbb{E}} \left[ \widehat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K})^d \right. \right. \\ \left. \left. (1 - \widehat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}))^{2-d} \right] \right| \leq \frac{c}{\log K},$$

for  $\varepsilon$  small enough and  $K$  large enough, where  $c$  is a finite constant. The proof of this inequality can be found in [36] pp. 1622-1624 and relies on Equation (A.13). The idea is to couple the process  $(K_m, 0 \leq m \leq J^K)$  with a process  $(K'_m, 0 \leq m \leq J^K)$  satisfying for every  $m \leq J^K$ ,

$$\mathcal{L} \left( K'_{m-1} - K'_m \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}, (K'_u)_{m \leq u \leq J^K} \right) = \text{Bin}(2 - K'_m, \rho_m^K),$$

where  $\text{Bin}(n, p)$  denotes the binomial distribution with parameters  $n$  and  $p$ , and  $\rho_m^K$  has been defined in (7.20). This implies

$$\mathcal{L} \left( K'_0 \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K} \right) = \text{Bin}(2, \widehat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 \mid (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K})),$$

and the coupling yields

$$\widehat{\mathbb{P}}(K'_m \neq K_m \text{ for some } 0 \leq m \leq J^K) \leq c/\log K,$$

for  $\varepsilon$  small enough and  $K$  large enough, where  $c$  is a finite constant. In particular, the weak dependence between two neutral lineages stated in Lemma 7.3 is needed in this proof. We now aim at proving that

$$(7.30) \quad \left| \hat{\mathbb{E}}[\hat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 | (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K})^d (1 - \hat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 | (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K})^{2-d} - \rho_K^d(1 - \rho_K)^{2-d}] \right| \leq c\varepsilon^{1/2},$$

where we recall the definition of  $\rho_K$  in (2.17). Equation (A.12) involves

$$\left| \hat{\mathbb{E}}[\hat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 | (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K})^d (1 - \hat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 | (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K})^{2-d} - \rho_K^d(1 - \rho_K)^{2-d}] \right| \leq 2\hat{\mathbb{E}}[\hat{\mathbb{P}}(TR_1^K(\beta_p) \geq 0 | (\tilde{N}_A(\tau_m^K), \tilde{N}_a(\tau_m^K))_{m \leq J^K}) - \rho_K].$$

Applying Equation (7.28) and adding (7.29), we finally get for  $d$  in  $\{0, 1, 2\}$ ,

$$(7.31) \quad \limsup_{K \rightarrow \infty} \left| \hat{\mathbb{P}}(\mathbf{1}_{TR_1^K(\beta_p) \geq 0} + \mathbf{1}_{TR_1^K(\beta_p) \geq 0} = d) - \binom{d}{2} \rho_K^d(1 - \rho_K)^{2-d} \right| \leq c\varepsilon^{1/2}.$$

**7.6. Neutral proportion at time  $T_\varepsilon^K$ .** Let us again focus on the population process  $N$ . By abuse of notation, we still use  $(TR_i^K(\beta_p), i \in \{1, 2\})$  and  $TC^K(\beta_p, \beta_q)$  to denote recombination and coalescence times of the neutral genealogies for the process  $N$ . According to Lemma 7.3, Equation (7.31), and Coupling (7.3),

$$\limsup_{K \rightarrow \infty} \mathbb{P}\left(\{TR_2^K(\beta_p) \geq 0\} \cup \{0 \leq TR_1^K(\beta_p) \leq TC^K(\beta_p, \beta_q)\} \mid T_\varepsilon^K < \infty\right) \leq c\varepsilon,$$

and

$$\limsup_{K \rightarrow \infty} \left| \mathbb{P}(\mathbf{1}_{TR_1^K(\beta_p) \geq 0} + \mathbf{1}_{TR_1^K(\beta_p) \geq 0} = d \mid T_\varepsilon^K < \infty) - \binom{d}{2} \rho_K^d(1 - \rho_K)^{2-d} \right| \leq c\varepsilon^{1/2},$$

for a finite  $c$  and  $\varepsilon$  small enough. Hence, it is enough to distinguish two cases for the randomly chosen neutral allele  $\beta_p$ : either its lineage has undergone one m-recombination, or no m-recombination. In the second case,  $\beta_p$  is a  $b_1$ . In the first one, the probability that  $\beta_p$  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  $\varepsilon_0$  such that for  $\varepsilon \leq \varepsilon_0$ ,*

$$\limsup_{K \rightarrow \infty} \mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} \left| P_{A, b_1}^K(t) - \frac{Z_{Ab_1}}{Z_A} \right| > \sqrt{\varepsilon}, T_\varepsilon^K < \infty\right) \leq c\varepsilon.$$

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  $\varepsilon_0$  such that for  $\varepsilon \leq \varepsilon_0$ ,*

$$\limsup_{K \rightarrow \infty} \hat{\mathbb{P}}\left(\left| P_{a, b_2}^K(T_\varepsilon^K) - \frac{Z_{Ab_2}}{Z_A} \rho_K \right| > \varepsilon^{1/6}\right) \leq c\varepsilon^{1/6}.$$

*Proof.* The sequence  $(\beta_i, i \leq \lfloor \varepsilon K \rfloor)$  denotes the neutral alleles carried by the  $a$ -individuals at time  $T_\varepsilon^K$  and

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

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

$$0 \leq N_{ab_2}^K(T_\varepsilon^K) - \sum_{i=1}^{\lfloor \varepsilon K \rfloor} \mathbf{1}_{A_2^K(i)} \leq \sum_{i=1}^{\lfloor \varepsilon K \rfloor} \mathbf{1}_{\{TR_2^K(\beta_i) \neq -\infty\}}.$$



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

$$(7.32) \quad \left| \hat{\mathbb{P}}\left(A_2^K(i) \mid TR_1^K(\beta_i) \geq 0, TR_2^K(\beta_i) = -\infty, \sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}^K(t) - \frac{z_{Ab_1}}{z_A} \right| \leq \sqrt{\varepsilon} \right) - \left( 1 - \frac{z_{Ab_1}}{z_A} \right) \right| \leq \sqrt{\varepsilon}.$$

Lemma 7.5 and Equation (A.5) ensure that  $\limsup_{K \rightarrow \infty} \hat{\mathbb{P}}(\sup_{t \leq T_\varepsilon^K} |P_{A,b_1}^K(t) - z_{Ab_1}/z_A| > \sqrt{\varepsilon}) \leq c\varepsilon$ , and Lemmas 7.3 and 7.4, and Coupling (7.3) that  $|\hat{\mathbb{P}}(TR_1^K(\beta_i) \geq 0, TR_2^K(\beta_i) = -\infty) - \rho_K| \leq c\varepsilon$ . It yields:

$$\left| \hat{\mathbb{P}}\left(TR_1^K(\beta_i) \geq 0, TR_2^K(\beta_i) = -\infty, \sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}^K(t) - \frac{z_{Ab_1}}{z_A} \right| \leq \sqrt{\varepsilon} \right) - \rho_K \right| \leq c\sqrt{\varepsilon}$$

for a finite  $c$  and  $\varepsilon$  small enough. Adding (7.32) we get:

$$(7.33) \quad \limsup_{K \rightarrow \infty} \left| \hat{\mathbb{E}}[P_{a,b_2}^K(T_\varepsilon^K)] - \rho_K \left( 1 - \frac{z_{Ab_1}}{z_A} \right) \right| \leq c\sqrt{\varepsilon}.$$

In the same way, using the weak dependence between lineages stated in (7.31) and Coupling (7.3), we prove that  $\limsup_{K \rightarrow \infty} |\hat{\mathbb{E}}[P_{a,b_2}^K(T_\varepsilon^K)^2] - \rho_K^2(1 - z_{Ab_1}/z_A)^2| \leq c\sqrt{\varepsilon}$ . This implies, adding (7.33) that  $\limsup_{K \rightarrow \infty} \hat{\text{Var}}(P_{a,b_2}^K(T_\varepsilon^K)) \leq c\sqrt{\varepsilon}$ . We end the proof by using Chebyshev's Inequality.  $\square$

**7.7. Second and third periods.** Thanks to Lemma 4.3 we already know that with high probability the neutral proportion in the  $a$ -population stays nearly constant during the third phase. We will prove that this is also the case during the second phase. 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  $\varepsilon_0$  such that for  $\varepsilon \leq \varepsilon_0$ ,*

$$(7.34) \quad \limsup_{K \rightarrow \infty} \hat{\mathbb{P}}\left(\left| P_{a,b_1}^K(T_{\text{ext}}^K) - P_{a,b_1}^K(T_\varepsilon^K) \right| > \varepsilon^{1/3}\right) \leq c\varepsilon^{1/3}.$$

*Proof.* Let us introduce the stopping time  $V_\varepsilon^K$ :

$$V_\varepsilon^K := \inf \left\{ t \leq t_\varepsilon, \sup_{\alpha \in \mathcal{A}} \left| N_\alpha^K(T_\varepsilon^K + t)/K - n_\alpha^{(N(T_\varepsilon^K)/K)}(t) \right| > \varepsilon^3 \right\},$$

where  $t_\varepsilon$  has been introduced in (6.9). Recall that  $(\mathcal{F}_t^K, t \geq 0)$  denotes the canonical filtration of  $N^K$ . The Strong Markov Property, Doob's Maximal Inequality and Equation (3.12) yield:

$$\begin{aligned} \mathbb{P}\left(T_\varepsilon^K \leq S_\varepsilon^K, \sup_{t \leq t_\varepsilon} |M_a^K(T_\varepsilon^K + t \wedge V_\varepsilon^K) - M_a^K(T_\varepsilon^K)| > \sqrt{\varepsilon}\right) \\ = \mathbb{E}\left[\mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \mathbb{P}\left(\sup_{t \leq t_\varepsilon} |M_a^K(T_\varepsilon^K + t \wedge V_\varepsilon^K) - M_a^K(T_\varepsilon^K)| > \sqrt{\varepsilon} \mid \mathcal{F}_{T_\varepsilon^K}^K\right)\right] \\ \leq \frac{1}{\varepsilon} \mathbb{E}\left[\mathbf{1}_{T_\varepsilon^K \leq S_\varepsilon^K} \left(\langle M_a \rangle_{T_\varepsilon^K + t_\varepsilon \wedge V_\varepsilon^K} - \langle M_a \rangle_{T_\varepsilon^K}\right)\right] \leq \frac{t_\varepsilon C(a, \bar{n}_a + \bar{n}_A)}{\varepsilon(I(\Gamma, \varepsilon) - \varepsilon^3)K}, \end{aligned}$$

for  $\varepsilon$  small enough, where  $I(\Gamma, \varepsilon)$  has been defined in (6.9). But according to Equation (2.5) with  $\delta = \varepsilon^3$ ,  $\limsup_{K \rightarrow \infty} \mathbb{P}(V_\varepsilon^K < t_\varepsilon \mid T_\varepsilon^K \leq S_\varepsilon^K) = 0$ . 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 4, we finally get:

$$(7.35) \quad \limsup_{K \rightarrow \infty} \mathbb{P}\left(\sup_{t \leq t_\varepsilon} |P_{a,b_1}^K(T_\varepsilon^K + t) - P_{a,b_1}^K(T_\varepsilon^K)| > \sqrt{\varepsilon}, T_\varepsilon^K \leq S_\varepsilon^K\right) = 0.$$

Adding Lemma 4.3 ends the proof of Lemma 7.7.  $\square$

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

$$\limsup_{K \rightarrow \infty} \mathbb{P} \left( \left| P_{a,b_2}^K(T_{\text{ext}}^K) - \rho_K \frac{z_{Ab_2}}{z_A} \right| > 2\varepsilon^{1/6} \right) \leq c\varepsilon^{1/6}.$$

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

$$\limsup_{K \rightarrow \infty} \mathbb{P} \left( \mathbf{1}_{\text{Fix}^K} \left| P_{a,b_2}^K(T_{\text{ext}}^K) - \rho_K \frac{z_{Ab_2}}{z_A} \right| > 2\varepsilon^{1/6} \right) \leq c\varepsilon^{1/6}.$$

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

#### A. TECHNICAL RESULTS

This section is dedicated to technical results needed in the proofs. We first present some results stated in [5]. We recall Definitions (2.8), (2.9), (4.8), (4.9), (3.10) and (6.9) and that the notation  $\cdot^K$  refers to the processes that satisfy Assumption 2. Proposition 2 is a direct consequence of Equations (42), (71), (72) and (74) in [5]:

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

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

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

$$(A.2) \quad F_\varepsilon^K = \left\{ T_\varepsilon^K \leq S_\varepsilon^K, N_a^K(T_\varepsilon^K + t_\varepsilon) < \frac{\varepsilon^2 K}{2}, |N_a^K(T_\varepsilon^K + t_\varepsilon) - \bar{n}_a K| < \frac{\varepsilon K}{2} \right\}$$

satisfies

$$(A.3) \quad \liminf_{K \rightarrow \infty} \mathbb{P}(T_\varepsilon^K \leq S_\varepsilon^K) \geq \liminf_{K \rightarrow \infty} \mathbb{P}(F_\varepsilon^K) \geq \frac{S_{aA}}{f_a} - M_2\varepsilon,$$

and if  $z \in \Theta$ , then there exist two positive finite constants  $V$  and  $c$  such that:

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

Thanks to these results we can state the following Lemma, which motivates the coupling of  $N$  and  $\tilde{N}$  and allows us to focus on the event  $\{\tilde{T}_\varepsilon^K < \infty\}$  rather than on  $\text{Fix}^K$  in Section 7.

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

$$(A.5) \quad \limsup_{K \rightarrow \infty} \mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K > S_\varepsilon^K) \leq c\varepsilon,$$

and

$$(A.6) \quad \limsup_{K \rightarrow \infty} \left[ \mathbb{P}(\{T_\varepsilon^K \leq S_\varepsilon^K\} \setminus \text{Fix}^K) + \mathbb{P}(\text{Fix}^K \setminus \{T_\varepsilon^K \leq S_\varepsilon^K\}) \right] \leq c\varepsilon.$$

*Proof.* We have the following equality

$$\begin{aligned} \mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K > S_\varepsilon^K) &= \mathbb{P}(T_\varepsilon^K < \infty) - \mathbb{P}(T_\varepsilon^K < \infty, T_\varepsilon^K \leq S_\varepsilon^K) \\ &= \mathbb{P}(T_\varepsilon^K < \infty) - \mathbb{P}(T_\varepsilon^K \leq S_\varepsilon^K), \end{aligned}$$

where we used the inclusion  $\{T_\varepsilon^K \leq S_\varepsilon^K\} \subset \{T_\varepsilon^K < \infty\}$ , as  $S_\varepsilon^K$  is almost surely finite (a birth and death process with competition has a finite extinction time). Equations (A.1) and (A.3) ends the proof of (A.5). From Equation (A.1), we also have that for  $\varepsilon < \bar{n}_a/2$

$$(A.7) \quad \lim_{K \rightarrow \infty} \mathbb{P}(T_\varepsilon^K = \infty | \text{Fix}^K) \leq \lim_{K \rightarrow \infty} \mathbb{P} \left( \left| N_a^K(T_{\text{ext}}^K) - K\bar{n}_a \right| > \varepsilon K \mid \text{Fix}^K \right) = 0.$$

This implies that

$$\mathbb{P}(T_\varepsilon^K > S_\varepsilon^K, \text{Fix}^K) \leq \mathbb{P}(T_\varepsilon^K > S_\varepsilon^K, T_\varepsilon^K < \infty) + \mathbb{P}(T_\varepsilon^K = \infty, \text{Fix}^K) \leq c\varepsilon,$$

where we used (A.5). Moreover,

$$\begin{aligned}\mathbb{P}(T_\varepsilon^K \leq S_\varepsilon^K, (\text{Fix}^K)^c) &\leq \mathbb{P}(T_\varepsilon^K < \infty, (\text{Fix}^K)^c) \\ &= \mathbb{P}(T_\varepsilon^K < \infty) - \mathbb{P}(T_\varepsilon^K < \infty | \text{Fix}^K) \mathbb{P}(\text{Fix}^K) \leq c\varepsilon,\end{aligned}$$

where we used (A.1), (A.7) and (2.16).  $\square$

We also recall some results on birth and death processes whose proofs can be found in Lemma 3.1 in [36] and in [1] p 109 and 112.

**Proposition 3.** *Let  $Z = (Z_t)_{t \geq 0}$  be a birth and death process with individual birth and death rates  $b$  and  $d$ . For  $i \in \mathbb{Z}^+$ ,  $T_i = \inf\{t \geq 0, Z_t = i\}$  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  $t \geq 0$ ,

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

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

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

- If  $d \neq b \in \mathbb{R}_+^*$ , for every  $i \in \mathbb{Z}_+$  and  $t \geq 0$ ,

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

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

$$(A.11) \quad T_N / \log N \xrightarrow[N \rightarrow \infty]{} (b - d)^{-1}, \quad a.s.$$

Finally, we recall Lemma 3.4.3 in [13] and Lemma 5.1 in [36]. Let  $d \in \mathbb{N}$ . Then

**Lemma A.2.** • Let  $a_1, \dots, a_d$  and  $b_1, \dots, b_d$  be complex numbers of modulus smaller than 1. Then

$$(A.12) \quad \left| \prod_{i=1}^d a_i - \prod_{i=1}^d b_i \right| \leq \sum_{i=1}^d |a_i - b_i|.$$

- Let  $V$  and  $V'$  be  $\{0, 1, \dots, d\}$ -valued random variables such that  $\mathbb{E}[V] = \mathbb{E}[V']$ . Then, there exist random variables  $\tilde{V}$  and  $\tilde{V}'$  on some probability space such that  $V$  and  $\tilde{V}$  have the same distribution,  $V'$  and  $\tilde{V}'$  have the same distribution, and

$$(A.13) \quad \mathbb{P}(\tilde{V} \neq \tilde{V}') \leq d \max\{\mathbb{P}(\tilde{V} \geq 2), \mathbb{P}(\tilde{V}' \geq 2)\}.$$

For  $0 < s < 1$ , if  $\tilde{Z}^{(s)}$  denotes a random walk with jumps  $\pm 1$  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

$$(A.14) \quad \tau_a := \inf\{n \in \mathbb{Z}_+, \tilde{Z}_n^{(s)} = \lfloor a \rfloor\}.$$

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

$$(A.15) \quad q_{j,k}^{(s_1, s_2)} := \frac{\mathbb{P}_{k+1}^{(s_1)}(\tau_{\varepsilon K} < \tau_k)}{\mathbb{P}_{k+1}^{(s_2)}(\tau_{\varepsilon K} < \tau_j)} = \frac{s_1}{1 - (1 - s_1)^{\lfloor \varepsilon K \rfloor - k}} \frac{1 - (1 - s_2)^{\lfloor \varepsilon K \rfloor - j}}{1 - (1 - s_2)^{k+1-j}}, \quad 0 < s_1, s_2 < 1,$$

whose expressions are direct consequences of (A.9). Let us now state a technical result, which helps us to control upcrossing numbers of the process  $\tilde{N}_a^K$  before reaching the size  $\lfloor \varepsilon K \rfloor$  (see Appendix B).

**Lemma A.3.** For  $a \in ]0, 1/2[$ ,  $(s_1, s_2) \in [a, 1-a]^2$ , and  $0 \leq j \leq k < l < \lfloor \varepsilon K \rfloor$ ,

$$(A.16) \quad q_{0,k}^{(s_1 \wedge s_2, s_1 \vee s_2)} \geq s_1 \wedge s_2 \quad \text{and} \quad \left| \frac{1}{q_{k,l}^{(s_1, s_2)}} - \frac{1}{q_{j,l}^{(s_2, s_1)}} \right| \leq \frac{4(1+1/s_2)}{ea^2 |\log(1-a)|} |s_2 - s_1| + \frac{(1-s_2)^{l+1-k}}{s_2^3}.$$

*Proof.* The first part of (A.16) is a direct consequence of Definition (A.15). 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  $x \in [a, 1-a]$ ,

$$(A.17) \quad \|f'_{\alpha,\beta}\|_\infty \leq 4(ea^2 |\log(1-a)|)^{-1}.$$

Indeed, the first derivative of  $f_{\alpha,\beta}$  is:

$$f'_{\alpha,\beta}(x) = \frac{\beta x^{\beta-1}(1-x^\alpha) - \alpha x^{\alpha-1}(1-x^\beta)}{(1-x^\beta)^2}.$$

Hence, for  $x \in [a, 1-a]$ ,

$$|f'_{\alpha,\beta}(x)| \leq \frac{\beta(1-a)^\beta + \alpha(1-a)^\alpha}{(1-a)a^2} \leq 2 \frac{\beta(1-a)^\beta + \alpha(1-a)^\alpha}{a^2},$$

where we used that  $1-x^\beta \geq 1-(1-a)$  and that  $1-a \geq 1/2$ . Adding the following inequality

$$\sup_{k \in \mathbb{N}} k(1-a)^k \leq \sup_{x \in \mathbb{R}^+} x(1-a)^x = (e|\log(1-a)|)^{-1},$$

completes the proof of (A.17). From (A.9), we get for  $0 < s < 1$  and  $0 \leq j \leq k < \lfloor \varepsilon K \rfloor$ ,

$$(A.18) \quad \begin{aligned} \left| \mathbb{P}_{l+1}^{(s)}(\tau_{\varepsilon K} < \tau_k) - \mathbb{P}_{l+1}^{(s)}(\tau_{\varepsilon K} < \tau_j) \right| &= \frac{(1-(1-s)^{k-j})((1-s)^{l+1-k} - (1-s)^{\lfloor \varepsilon K \rfloor - k})}{(1-(1-s)^{\lfloor \varepsilon K \rfloor - k})(1-(1-s)^{\lfloor \varepsilon K \rfloor - j})} \\ &\leq (1-s)^{l+1-k} s^{-2}. \end{aligned}$$

The 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.18) and the Mean Value Theorem with (A.17), we get the second part of (A.16).  $\square$

## B. PROOFS OF LEMMAS 7.2 AND 7.5

*Proof of Equation (7.17).* In the whole proof, the integer  $n_A$  denotes the state of  $\tilde{N}_A$  and thus belongs to  $I_\varepsilon^K$  which has been defined in (3.9).  $\mathbb{P}_{(n_A, n_a)}$  (resp.  $\hat{\mathbb{P}}_{(n_A, n_a)}$ ) denotes the probability  $\mathbb{P}$  (resp.  $\hat{\mathbb{P}}$ ) when  $(\tilde{N}_A(0), \tilde{N}_a(0)) = (n_A, n_a) \in \mathbb{Z}_+^2$ . We introduce for  $u \in \mathbb{R}_+$  the hitting time of  $\lfloor u \rfloor$  by the process  $\tilde{N}_a$ :

$$(B.1) \quad \sigma_u^K := \inf\{t \geq 0, \tilde{N}_a^K(t) = \lfloor u \rfloor\}.$$

Let  $(i, j, k)$  be in  $\mathbb{Z}_+^3$  with  $j < k < \lfloor \varepsilon K \rfloor$ . Between jumps  $\zeta_j^K$  and  $J^K$  the process  $\tilde{N}_a$  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  $\tilde{N}_A$  when  $\tilde{N}_a$  hits  $k+1$  for the first time, we take the maximum over all the possible values in  $I_\varepsilon^K$ . Recall Definition (7.4). We get, as  $\{\tilde{T}_\varepsilon^K < \sigma_j^K\} \subset \{\tilde{T}_\varepsilon^K < \sigma_k^K\} \subset \{\tilde{T}_\varepsilon^K < \infty\}$ :

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

where we used that on the events  $\{\tilde{T}_\varepsilon^K < \sigma_j^K\}$  and  $\{\tilde{T}_\varepsilon^K < \sigma_k^K\}$  the jumps from  $j$  to  $j+1$  belong to the past, and Markov Property. Coupling (7.6) allows us to compare these conditional probabilities with the probabilities of the same events under  $\mathbb{P}^{(s_-(\varepsilon))}$  and  $\mathbb{P}^{(s_+(\varepsilon))}$ , and recalling (A.15) we get

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

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  $\{U_j^K = i\}$ ,

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

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

$$(B.3) \quad G'_1 \leq U_{j,k}^{(K,2)} \leq G'_2.$$

Recall that  $U_{0,k}^{(K,2)} = U_k^K$ . Hence taking  $j = 0$  and adding the first part of Equation (A.16) give the first inequality of (7.17). According to Definition (5.1),  $|s_+(\varepsilon) - s_-(\varepsilon)| \leq c\varepsilon$  for a finite  $c$ . Hence Equations (B.2), (B.3) and (A.16) entail the existence of a finite  $c$  such that for  $\varepsilon$  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 (7.17),

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

where we use that  $U_j^K \leq (U_j^K)^2$ . This ends the proof of (7.17).  $\square$

*Proof of Equation (7.16).* Definitions (2.3) and Coupling (7.6) ensure that for  $n_A \in I_\varepsilon^K$ ,  $\varepsilon$  small enough and  $K$  large enough,

$$\begin{aligned} \hat{\mathbb{P}}_{(n_A, k)}(\tilde{N}_a(dt) = k+1) &= \frac{\mathbb{P}_{(n_A, k+1)}(\tilde{T}_\varepsilon^K < \infty)}{\mathbb{P}_{(n_A, k)}(\tilde{T}_\varepsilon^K < \infty)} \mathbb{P}_{(n_A, k)}(\tilde{N}_a(dt) = k+1) \\ &\geq \frac{\mathbb{P}_{k+1}^{(s_-(\varepsilon))}(\tau_{\varepsilon K} < \tau_0)}{\mathbb{P}_k^{(s_+(\varepsilon))}(\tau_{\varepsilon K} < \tau_0)} f_a k dt \\ &= \frac{1 - (1 - s_-(\varepsilon))^{k+1}}{1 - (1 - s_-(\varepsilon))^{\lfloor \varepsilon K \rfloor}} \frac{1 - (1 - s_+(\varepsilon))^{\lfloor \varepsilon K \rfloor}}{1 - (1 - s_+(\varepsilon))^k} f_a k dt \\ &\geq s_-^2(\varepsilon) f_a k dt, \end{aligned}$$

and

$$\begin{aligned} \hat{\mathbb{P}}_{(n_A, k)}(\tilde{N}_A(dt) \neq n_A) &\leq \frac{\mathbb{P}_k^{(s_+(\varepsilon))}(\tau_{\varepsilon K} < \tau_0)}{\mathbb{P}_k^{(s_-(\varepsilon))}(\tau_{\varepsilon K} < \tau_0)} \mathbb{P}_{(n_A, k)}(\tilde{N}_A(dt) \neq n_A) \\ &\leq (1 + c\varepsilon) 2f_A \bar{n}_A K dt \end{aligned}$$

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

$$\hat{\mathbb{P}}(\tilde{N}_a(\tau_{m+1}^K) \neq \tilde{N}_a(\tau_m^K) | \tilde{N}_a(\tau_m^K) = k) \geq \frac{s_-^2(\varepsilon) f_a k}{3f_A \bar{n}_A K}.$$

If  $D_k^K$  denotes the downcrossing number from  $k$  to  $k-1$  before  $\tilde{T}_\varepsilon^K$ , 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  $s_-^2(\varepsilon) f_a k / 3f_A \bar{n}_A K$  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 (7.17) twice we get

$$\hat{\mathbb{E}}[H_k^K] \leq \left( \frac{4}{s_-^2(\varepsilon)} - 1 \right) \left( \frac{3f_A \bar{n}_A K}{s_-^2(\varepsilon) f_a k} - 1 \right),$$

which ends the proof of the first inequality in (7.16).

As the mutant population size is not Markovian we cannot use symmetry and the Strong Markov Property to control the dependence of jumps before and after the last visit to a given state as in [36]. Hence we describe the successive excursions of  $\tilde{N}_a^K$  above a given level to get the last inequality in (7.16). Let  $\tilde{U}_{j,k}^{(i)}$  be the number of jumps from  $k$  to  $k+1$  during the  $i$ th excursion above  $j$ . We first bound the expectation  $\hat{\mathbb{E}}[(\tilde{U}_{j,k}^{(i)})^2]$ . During an excursion above  $j$ ,  $\tilde{N}_a$  hits  $j+1$ , but we do not know the value of  $\tilde{N}_A$  at this time. Thus we take the maximum value for the probability when  $n_A$  belongs to  $I_\varepsilon^K$ , and

$$\hat{\mathbb{P}}(\tilde{U}_{j,k}^{(i)} \geq 1) \leq \sup_{n_A \in I_\varepsilon^K} \hat{\mathbb{P}}_{(j+1, n_A)}(\sigma_{k+1}^K < \sigma_j^K | \sigma_j^K < \tilde{T}_\varepsilon^K).$$

Then using Coupling (7.6) and Definition (7.4) we obtain

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

Adding Equation (A.9) we finally get

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



Moreover if  $\tilde{U}_{j,k}^{(i)} \geq 1$ ,  $\tilde{N}_a$  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:

$$\begin{aligned} \hat{\mathbb{P}}\left(\tilde{U}_{j,k}^{(i)} = 1 \mid \tilde{U}_{j,k}^{(i)} \geq 1\right) &\geq \inf_{n_A \in I_\varepsilon^K} \hat{\mathbb{P}}_{(n_A, k)}(\sigma_j^K < \sigma_{k+1}^K \mid \sigma_j^K < \tilde{T}_\varepsilon^K) \\ &\geq \frac{\mathbb{P}_j^{(s_-(\varepsilon))}(\tau_{\varepsilon K} < \tau_0) \mathbb{P}_k^{(s_+(\varepsilon))}(\tau_j < \tau_{k+1})}{\mathbb{P}_j^{(s_+(\varepsilon))}(\tau_{\varepsilon K} < \tau_0) \mathbb{P}_k^{(s_-(\varepsilon))}(\tau_j < \tau_{\varepsilon K})}, \end{aligned}$$

which yields

$$(B.6) \quad \hat{\mathbb{P}}\left(\tilde{U}_{j,k}^{(i)} = 1 \mid \tilde{U}_{j,k}^{(i)} \geq 1\right) \geq s_-(\varepsilon) s_+(\varepsilon) \left(\frac{1-s_+(\varepsilon)}{1-s_-(\varepsilon)}\right)^{k-j} \geq s_-^2(\varepsilon) \left(\frac{1-s_+(\varepsilon)}{1-s_-(\varepsilon)}\right)^{k-j} =: q.$$

Hence, given that  $\tilde{U}_{j,k}^{(i)}$  is non-null,  $\tilde{U}_{j,k}^{(i)}$  is smaller than a geometrically distributed random variable with parameter  $q$ . In particular,

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

Adding Equation (B.5) and recalling that  $|s_+(\varepsilon) - s_-(\varepsilon)| \leq c\varepsilon$  for a finite  $c$  yield

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

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  $\tilde{T}_\varepsilon^K$  is  $U_j^K - 1$ , we get

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

where we used the first part of Equation (7.17). This ends the proof of Equation (7.16).  $\square$

*Proof of Equation (7.18).* Definition (A.15), Inequality (B.3) and Equation (A.9) yield:

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

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  $K$ ,  $A = \log(\varepsilon K) + O(1)$ , and for every  $u > 1$  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  $\varepsilon$  small enough,  $|s_+(\varepsilon) - s_{aA}/f_a| \leq c\varepsilon$  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 ends the proof of Lemma 7.2.  $\square$

*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  $Z_\varepsilon^*$  are denoted by:

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

Let us introduce the difference of probabilities

$$B_\varepsilon^K := \mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}^K(t) - \frac{Z_{Ab_1}}{Z_A} \right| > \sqrt{\varepsilon}, T_\varepsilon^K < \infty\right) - \mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}^K(t) - \frac{Z_{Ab_1}}{Z_A} \right| > \sqrt{\varepsilon}, F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\right).$$

Then  $B_\varepsilon^K$  is nonnegative and we have

$$\begin{aligned} B_\varepsilon^K &\leq \mathbb{P}(T_\varepsilon^K < \infty) - \mathbb{P}(F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K) \\ &= \mathbb{P}(T_\varepsilon^K < \infty) - \mathbb{P}(T_\varepsilon^K \leq S_\varepsilon^K) + \mathbb{P}(T_\varepsilon^{(+,K)} < \infty, T_\varepsilon^K \leq S_\varepsilon^K) - \mathbb{P}(F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K), \end{aligned}$$

where the inequality comes from the inclusion  $\{F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\} \subset \{T_\varepsilon^K < \infty\}$ , as  $S_\varepsilon^K$  is almost surely finite. The equality is a consequence of Coupling (5.3) which ensures that on the event  $\{T_\varepsilon^K \leq S_\varepsilon^K\}$ ,  $\{T_\varepsilon^{(+,K)} < \infty\}$  holds. By noticing that

$$\{F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\} \subset \{T_\varepsilon^{(-,K)} < \infty, T_\varepsilon^K \leq S_\varepsilon^K\} \subset \{T_\varepsilon^{(+,K)} < \infty, T_\varepsilon^K \leq S_\varepsilon^K\}$$

we get the bound

$$(B.7) \quad B_\varepsilon^K \leq \mathbb{P}(T_\varepsilon^K < \infty) - \mathbb{P}(T_\varepsilon^K \leq S_\varepsilon^K) + \mathbb{P}(T_\varepsilon^{(+,K)} < \infty) - \mathbb{P}(F_\varepsilon^-).$$

The values of the two first probabilities are approximated in (A.1) and (A.3), and (A.9) implies that  $\mathbb{P}(T_\varepsilon^{(+,K)} < \infty) - \mathbb{P}(F_\varepsilon^-) = s_+(\varepsilon)/(1 - (1 - s_+(\varepsilon))^{\lfloor \varepsilon K \rfloor}) - s_-(\varepsilon)$ . Hence

$$(B.8) \quad \limsup_{K \rightarrow \infty} B_\varepsilon^K \leq c\varepsilon,$$

where  $c$  is finite for  $\varepsilon$  small enough, which allows us to focus on the intersection with the event  $\{F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\}$ . We recall that  $|N_{Ab_1}N_{Ab_2} - N_{Ab_2}N_{Ab_1}| \leq N_A N_a$ , and that Assumption 4 holds. Then (3.1) and (3.10) imply for  $\varepsilon$  small enough

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

for a finite  $c$ . Moreover,  $F_\varepsilon^- \cap \{T_\varepsilon^K \leq S_\varepsilon^K\} \subset F_\varepsilon^- \cap \{T_\varepsilon^K \leq T_\varepsilon^{(-,K)}\}$ . Thus we get

$$\mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} - M_A(t) \right| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\right) \leq \mathbb{P}\left(\frac{c\varepsilon T_\varepsilon^{(-,K)}}{\log K} > \sqrt{\varepsilon}/2, F_\varepsilon^-\right).$$

Finally, Equation (A.11) ensures that  $\lim_{K \rightarrow \infty} T_\varepsilon^{(-,K)}/\log K = s_-(\varepsilon)^{-1}$  a.s. on the non-extinction event  $F_\varepsilon^-$ . Thus for  $\varepsilon$  small enough,

$$(B.9) \quad \lim_{K \rightarrow \infty} \mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} - M_A(t) \right| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\right) = 0.$$

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

$$\mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} |M_A(t)| > \frac{\sqrt{\varepsilon}}{2}, F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\right) \leq \mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} |M_A(t)| > \frac{\sqrt{\varepsilon}}{2}, T_\varepsilon^K \leq S_\varepsilon^K \wedge t_K\right) + \mathbb{P}(T_\varepsilon^K > t_K, F_\varepsilon^-).$$

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

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

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

$$\begin{aligned} \mathbb{P}(\sup_{t \leq T_\varepsilon^K} |M_A(t)| > \sqrt{\varepsilon}/2, T_\varepsilon^K \leq S_\varepsilon^K \wedge t_K) &\leq \mathbb{P}(\sup_{t \leq t_K} |M_A(t \wedge T_\varepsilon^K \wedge S_\varepsilon^K)| > \sqrt{\varepsilon}/2) \\ &\leq \frac{4}{\varepsilon} \mathbb{E}\left[\langle M_A \rangle_{t_K \wedge T_\varepsilon^K \wedge S_\varepsilon^K}\right] \\ &\leq \frac{4}{\varepsilon} \frac{C(A, 2\bar{n}_A) t_K}{(\bar{n}_A - 2\varepsilon C_{A,a}/C_{A,A})K}, \end{aligned}$$

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

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(\sup_{t \leq T_\varepsilon^K} \left| P_{A,b_1}(t) - \frac{z_{Ab_1}}{z_A} - M_A(t) \right| > \sqrt{\varepsilon}, F_\varepsilon^-, T_\varepsilon^K \leq S_\varepsilon^K\right) = 0.$$

Finally, Equation (B.8) complete the proof of Lemma 7.5.  $\square$

**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, so as the two anonymous referees for several corrections and improvements. 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 Chair Modélisation Mathématique et Biodiversité VEOLIA-Ecole Polytechnique-MNHN-FX.

## REFERENCES

- [1] K. B. Athreya and P. E. Ney. *Branching processes*, volume 28. Springer-Verlag Berlin, 1972.
- [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, P.-E. Jabin, and S. Méléard. Adaptation in a stochastic multi-resources chemostat model. *Journal de Mathématiques Pures et Appliquées*, 101(6):755–788, 2014.
- [7] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*, 151(1-2):45–94, 2011.
- [8] 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.
- [9] G. Coop and P. Ralph. Patterns of neutral diversity under general models of selective sweeps. *Genetics*, 192(1):205–224, 2012.
- [10] C. Coron. Slow-fast stochastic diffusion dynamics and quasi-stationary distributions for diploid populations. *arXiv preprint arXiv:1309.3405*, 2013.
- [11] C. Coron. Stochastic modeling of density-dependent diploid populations and the extinction vortex. *Advances in Applied Probability*, 46(2):446–477, 2014.
- [12] E. Durand, M. I. Tenaillon, C. Ridel, D. Coubriche, P. Jamin, S. Jouanne, A. Ressayre, A. Charcosset, and C. Dillmann. Standing variation and new mutations both contribute to a fast response to selection for flowering time in maize inbreds. *BMC evolutionary biology*, 10(1):2, 2010.
- [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] J. Hermisson and P. S. Pennings. Soft sweeps molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352, 2005.
- [20] J. Hermisson and P. Pfaffelhuber. The pattern of genetic hitchhiking under recurrent mutation. *Electron J Probab*, 13(68):2069–2106, 2008.
- [21] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. 1989.
- [22] N. L. Kaplan, R. Hudson, and C. Langley. The "hitchhiking effect" revisited. *Genetics*, 123(4):887–899, 1989.
- [23] S. Leocard. Selective sweep and the size of the hitchhiking set. *Advances in Applied Probability*, 41(3):731–764, 2009.
- [24] J. Maynard Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genet Res*, 23(1):23–35, 1974.
- [25] G. McVean. The structure of linkage disequilibrium around a selective sweep. *Genetics*, 175(3):1395–1406, 2007.
- [26] P. W. Messer and D. A. Petrov. Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*, 28(11):659–669, 2013.
- [27] 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.
- [28] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 145(2):519–534, 1997.

- [29] T. Ohta and M. Kimura. The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genetical research*, 25(03):313–325, 1975.
- [30] H. A. Orr and A. J. Betancourt. Haldane's sieve and adaptation from the standing genetic variation. *Genetics*, 157(2):875–884, 2001.
- [31] P. S. Pennings and J. Hermisson. Soft sweeps ii-molecular population genetics of adaptation from recurrent mutation or migration. *Molecular biology and evolution*, 23(5):1076–1084, 2006.
- [32] P. S. Pennings and J. Hermisson. Soft sweeps iii: the signature of positive selection from recurrent mutation. *PLoS genetics*, 2(12):e186, 2006.
- [33] 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.
- [34] 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.
- [35] M. Prezeworski, G. Coop, and J. D. Wall. The signature of positive selection on standing genetic variation. *Evolution*, 59(11):2312–2323, 2005.
- [36] 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.
- [37] W. Stephan, Y. S. Song, and C. H. Langley. The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics*, 172(4):2647–2663, 2006.
- [38] 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.
- [39] B. Wallace. Hard and soft selection revisited. *Evolution*, pages 465–473, 1975.