



## The Effect of Recombination on the Speed of Evolution

Nantawat Udomchatpitak

Department of Mathematics, Faculty of Science, Mahidol University  
272 Rama VI Road, Ratchathewi District,  
Bangkok, 10400, Thailand.

*E-mail address:* [nantawat.udo@mahidol.ac.th](mailto:nantawat.udo@mahidol.ac.th)

*URL:* <https://mathematics.sc.mahidol.ac.th/faculty-members/>

**Abstract.** It has been a puzzling question of why sexual reproduction has been so successful. Fisher and Muller hypothesized that reproducing by sex can speed up evolution; recombination could combine beneficial alleles on different chromosomes and speeds up the time that the beneficial alleles become fixated population. We consider a population of  $2N$  chromosomes and focus on two loci on these chromosomes. The allele at each locus can mutate to a beneficial allele at rate  $\mu_N$ . Chromosomes with 0, 1, and 2 beneficial alleles die at rates  $1$ ,  $1 - s_N$ , and  $1 - 2s_N$ , respectively, and they are replaced immediately after death events. With probability  $1 - r_N$ , the replacement inherits both alleles from one parent chosen at random from the population at the time of replacement. With probability  $r_N$ , recombination occurs, and the replacement receives its two alleles from two randomly chosen parents. Under certain assumptions on the parameters  $N$ ,  $\mu_N$ ,  $s_N$ , and  $r_N$ , we obtain an asymptotic approximation to the time that both beneficial alleles are fixated in the population. When the recombination probability is small, recombination does not speed up the time that the two beneficial alleles become fixated. In contrast, when the recombination probability is significant, recombination shorten this time. The result agrees with the Fisher-Muller hypothesis and confirms the advantage of reproducing by sex.

### 1. Introduction

Sexual reproduction has been so successful that it has persisted and survived natural selection. Several organisms nowadays can reproduce by sex, and many of

---

*Received by the editors* *July 4th, 2019*; *accepted* *November 28th, 2020*.

2010 *Mathematics Subject Classification.* 92D15, 60J27, 60J75, 60J85.

*Key words and phrases.* beneficial mutations, evolution, fixation time, recombination, selection.

Research supported in part by NSF grant DMS-1707953.

these have lost the ability to reproduce asexually altogether. Evolutionary theorists have been searching for reasons behind the success of sexual reproduction.

While sexual reproduction sounds like a success, it has huge costs. First, organisms that reproduce sexually can only transmit half of their genes to the offspring. Thus, the offspring might not receive all beneficial alleles from the parent. In contrast, if the parent reproduces asexually, the offspring will be genetically identical to the parent and receive the full advantage of beneficial alleles from the parent.

A sexually reproducing population must have both sexes, male and female, to be able to reproduce the next generation. To make matters worst, only the females can carry and produce offspring. Unlike asexually reproducing populations, every individual has the capability of producing offspring on its own. This disadvantage of sex has been known as the cost of males; males are necessary for the survival of the population, but they cannot carry offspring.

Even though sexual reproduction, when it is compared with asexual reproduction, seems to be a disadvantage, several benefits might outweigh the drawbacks. In asexually reproducing populations, beneficial mutations usually arise in different backgrounds. Their descendants compete against each other for survival in the population, hindering the evolutionary process. This effect is known as clonal interference. Several authors have been studying the rate of adaptation and the fate of beneficial alleles in an asexual population (see [Gerrish and Lenski, 1998](#); [Rouzine et al., 2003](#); [Wilke, 2004](#); [Desai and Fisher, 2007](#); [Rouzine et al., 2008](#); [Fogle et al., 2008](#); [Park et al., 2010](#); [Hallatschek, 2011](#); [Good et al., 2012](#); [Fisher, 2013](#)).

[Fisher \(1999\)](#) and [Muller \(1932\)](#) proposed that sexual reproduction can speed up evolution. This is also known as the Fisher-Muller hypothesis. In an asexually reproducing population, the only way to avoid competition among beneficial alleles is that a new beneficial mutation must occur in the background that had previous beneficial mutations. However, in a sexually reproducing population, recombination can combine beneficial mutations in different backgrounds, resulting in a faster evolutionary rate.

The Fisher-Muller hypothesis has been a subject of theoretical research and has caused much debate over this benefit of sexual reproduction over asexual reproduction. [Crow and Kimura \(1965\)](#) used a simple quantitative argument to support that recombination can increase the rate of evolution in a large population with frequent small-effect beneficial mutations. [Maynard Smith \(1968\)](#) believed otherwise; his study on an infinite two-loci population model in which beneficial mutations are recurrent showed that if the beneficial mutants are in equilibrium frequencies, which balance the effects of recurrent mutations and selection, this equilibrium will persist through time. Hence, recombination cannot speed up the evolution because it can only act on the population by destroying the linkage disequilibrium. We note that the assumption on recurrent mutations is different from the model of [Crow and Kimura \(1965\)](#), which assumes that every new beneficial mutation is different from all past mutations. Several other works testing the validity of the Fisher-Muller hypothesis are discussed by [Felsenstein \(1974\)](#).

Another benefit of recombination arises from its ability to break linked loci. [Hill and Robertson \(1966\)](#) discovered that the selection at one locus could interfere with the selection at other linked locus, even when there is no interaction between the genes at these loci. Hence, the overall effect of selection can be reduced by the linkage among loci. Recombination could be favored by selection because of

its ability to break linkages. Thus, selection should favor higher recombination rate. Several authors study the fate of alleles that can modify recombination rates, also known as recombination-modifier alleles. (see Charlesworth, 1993a; Otto and Feldman, 1997; Barton and Otto, 2005; Keightley and Otto, 2006; Roze and Barton, 2006; Hartfield et al., 2010).

Recombination can also help a beneficial mutation escape from extinction if the beneficial mutation is originated in a deleterious background. Most mutations that occur in a population are believed to produce negative effects. Haigh (1978) studied the distribution of the number of deleterious genes in a finite population in which deleterious mutations occur at a constant rate. Later, Peck (1994) studied the fate of a beneficial mutation that occurs in this population and found that the survival probability of the beneficial mutant is enhanced if the population reproduces sexually.

Lastly, sexual reproduction creates genetic diversity, which allows the population to adapt to the changing environment that may be shaped by the evolutionary racing between predator and prey, or between host and parasite. This benefit of sexual reproduction is known as the red queen hypothesis (see Van Valen, 1973). Several works study the evolution of sex and recombination in changing environments (see Charlesworth, 1993b; Bürger, 1999; Waxman and Peck, 1999).

In this article, we consider a two-loci model in which there are two types of alleles, neutral and beneficial, at each locus. Each neutral allele at each locus can mutate to the beneficial version, but not backward. We present an asymptotic result for the fixation time of the beneficial alleles, which is the first time that the whole population has only beneficial alleles. Then, we use the result to compare the fixation time in an asexually reproducing population with the fixation time in a sexually reproducing population to confirm the Fisher-Muller hypothesis.

1.1. *The model.* We consider a population of fixed size  $2N$ , representing  $2N$  chromosomes in a diploid population. These chromosomes will be referred to as individuals. We will focus on two loci on which an  $a$  or  $A$  allele resides at the first locus, and a  $b$  or  $B$  allele resides at the second locus. Hence, each chromosome can be one of the following four types:  $ab, Ab, aB$ , and  $AB$ . The  $a$  and  $b$  alleles are neutral alleles, while the  $A$  and  $B$  alleles are beneficial alleles. Each  $a$  allele and each  $b$  allele mutates to  $A$  allele and  $b$  allele, respectively, at an exponential rate  $\mu_N$ , and we assume that the mutations occur independently. With this assumption, beneficial mutations will not be exhausted. However, backward mutations are not allowed. Selection acts on each individual through the death rate: individuals with 0, 1, and 2 beneficial alleles die independently at exponential rates  $1, 1 - s_N$ , and  $1 - 2s_N$ , respectively. We also assume that the deaths occur independently from each other and independently from the mutations. After an individual dies, a new individual is created immediately to keep the population size fixed. With probability  $1 - r_N$ , no recombination occurs, and the new individual receives both alleles from a randomly chosen individual in the population at that time. With probability  $r_N$ , recombination occurs, and the new individual receives each allele from two independently randomly selected individuals. The goal is to understand the asymptotic behavior of the time that both beneficial alleles have fixated, assuming that the population starts with no beneficial allele.

1.2. *Previous works.* For the study on the fixation time of two beneficial mutations, [Takahata \(1982\)](#) considered a model similar to ours, but the population process is a discrete-time process with discrete generations. Also, the fitnesses of individuals with 0, 1, and 2 mutations are assumed to be  $1, 1 + s, 1 + s,$  and  $1 + t,$  respectively. Several parameter regimes for  $s$  and  $t$  were considered, i.e.,  $t = s = 0, t = 2s > 0, t = 2s < 0, t > 2s > 0,$  and  $t > 0 > s.$  (Note that in some of these parameter regimes, the mutations are neutral or deleterious.) The numerical fixation time of both beneficial mutations was obtained through simulation, which leads to the conclusion that recombination can decrease the fixation time of two beneficial mutations. This model shares some resemblance to our model in the parameter regime  $t = 2s > 0.$

Some works in quantitative biological literature consider similar models that allows backward mutations with more than two loci. For asexual reproduction, [Rouzine et al. \(2003\)](#) proposed that the distribution of fitness over time behaves like a solitary wave on the fitness space. [Cohen et al. \(2006\)](#) consider another similar model in continuous time that incorporates recombination. Their result suggested that recombination can speed up evolution in a population that is large enough.

Several other works obtained results that suggested relations between recombination rate and the speed of adaptation. [Neher et al. \(2010\)](#) considered a population of size  $N$  with a large number of loci. The fitnesses of all allele combinations are described through a distribution on the fitness space. Beneficial mutations occur at a rate  $U_b.$  Individuals in the population are capable of reproducing asexually and sexually; sexual reproduction occurs at rate  $r.$  Assuming weak selection with large population size ( $Ns \gg 1$ ), and weak interference between loci ( $r \gg s$ ), the result suggested that the speed of adaptation increases as  $r^2.$  [Weissman and Barton \(2012\)](#) considered a similar model but with a linear genetic map of length  $R,$  which means that there are  $R$  recombination events on average in each generation. Unlike the previous model, the population cannot reproduce asexually. Under the assumption that  $U_b$  is small, their results suggested that the rate of adaptation is proportional to  $R.$  [Weissman and Hallatschek \(2014\)](#) considered a larger value of  $U_b,$  but not as large as  $R,$  and obtained the same result. Lastly, [Neher et al. \(2013\)](#) considered the regime of a large mutation rate, which was not considered in the other two works, but still assumed a small selection effect. Their result suggested that the rate of adaptation is proportional to the square root of the recombination rate.

Some rigorous results can be found in the mathematics literature. [Berestycki and Zhao \(2018\)](#) studied branching Brownian motion in a particular fitness space and showed that the population would ultimately move toward a fitter landscape; the result is similar to the solitary wave description by [Rouzine et al. \(2003\)](#). Furthermore, this result also shows that the fitnesses on two loci are negatively correlated. Recombination can reduce this negative correlation, pushing the population toward the direction of fitter background in the fitness landscape.

[Cuthbertson et al. \(2012\)](#) considered a two loci model similar to our model. The first difference between the two models is the selection mechanism. In their model,  $A$  and  $B$  alleles increase fitness by  $s_1$  and  $s_2,$  respectively, with the assumption that  $s_1 < s_2.$  The second difference is that the mutation from  $ab$  to  $aB$  occurs only once randomly during the spreading of  $Ab.$  For both  $A$  and  $B$  to spread to the entire population, three requirements must be met. First, type  $aB$  should become a significant fraction of the population. Second, recombination between  $A$  and

$B$  must occur. Lastly, the number of type  $AB$  should become substantial, after which  $AB$  is almost sure to fixate. The essence of their result is that the fixation probability is approximately  $2s_2/(1 - s_2)$  multiplied by a certain probability that involves the solution to a specific system of ODEs. Here, we note that this solution to the system of ODEs was not given in an explicit form.

Bossert and Pfaffelhuber (2018) considered a diffusion model with four types:  $ab, Ab, aB,$  and  $AB$ , where the fitnesses of  $ab, Ab, aB,$  and  $AB$  are in increasing order. The frequencies of these four types evolve according to a system of SDEs. Initially, the frequencies of types  $Ab$  and  $aB$  are assumed to be small, and there is no type  $AB$  yet. They obtained formulas for the approximations of the fixation probability and fixation time of type  $AB$ . Their result can also be used to approximate the fixation time for finite populations: the fixation probability of  $AB$  is approximately  $2s_2$  multiplied by a positive constant, which takes a value less than one and depends on the fitness parameters and the recombination rate. Hence, this result shares some similar features with the result by Cuthbertson et al. (2012).

Both Cuthbertson et al. (2012) and Bossert and Pfaffelhuber (2018) assume that at least one beneficial mutation is present at the beginning and do not allow an unlimited supply of new mutations. In contrast, our model assumes that all individuals do not have any beneficial mutations in the beginning, and both beneficial mutations occur according to a Poisson process.

1.3. *Conditions of the parameters.* There are four parameters in our model:  $N, \mu_N, r_N,$  and  $s_N$ . We assume that  $\mu_N \in (0, 1), s_N \in (0, 1/2]$  and  $r_N \in [0, 1)$ .

For two non-negative sequences  $(a_N)_{N=1}^\infty$  and  $(b_N)_{N=1}^\infty$ , we write  $a_N \ll b_N$  if

$$\lim_{N \rightarrow \infty} \frac{a_N}{b_N} = 0.$$

The four parameters are assumed to satisfy the following conditions:

$$s_N \ll 1, \tag{1.1}$$

$$1 \ll N\mu_N, \tag{1.2}$$

$$N\mu_N^2 \ll s_N, \tag{1.3}$$

and

$$r_N \ln_+(Nr_N) \ll s_N, \tag{1.4}$$

where  $\ln_+(x)$  is defined to be  $\ln(x)$  if  $x \in (1, \infty)$ , and 0 if  $x \in [0, 1]$ . Note that (1.2) and (1.3) imply that

$$\mu_N \ll s_N.$$

Many population models in mathematical literature assume that  $Ns_N = O(1), N\mu_N = O(1),$  and  $Nr_N = O(1),$  since these assumptions allow one to use scaling arguments to obtain diffusion limits. Here, we explore the parameter in the complementary regime. The condition  $s_N \ll 1$  says that each beneficial allele has only a weak positive effect, and the presence of beneficial mutations at the beginning does not have a considerable impact on the population.

The condition  $1 \ll N\mu_N$  ensures that mutations occur fast enough for both  $A$  and  $B$  alleles to start spreading in the population around the same time. If  $\mu_N$  is too small, one of the beneficial alleles become fixated before the second beneficial allele becomes a significant portion of the population. Hence, the low mutation rate is not of much interest, since recombination cannot improve the fixation time of  $AB$ .

For the two beneficial alleles to fixate separately without affecting one another, the second beneficial mutation should not start spreading until the first mutation has fixated. The first beneficial allele will take time about  $\frac{2}{s_N} \ln(2Ns_N)$  to become fixated in the population (see section 6.1 of [Durrett, 2008](#)). During this time, a mutation that gives rise to the second beneficial allele occurs at a total rate of  $2N\mu_N$ . Here, we can use an asymmetric random walk to approximate the number of descendants of the individuals that carry the second beneficial allele; the survival probability of each second beneficial allele is about  $s$ . Hence, the number of second beneficial mutations that eventually survive in the population during the spreading of the first beneficial allele is approximately

$$2N\mu_N \cdot s_N \cdot \frac{2}{s_N} \ln(2Ns_N) = 4N\mu_N \ln(2Ns_N).$$

So, if  $N\mu_N \ln(2Ns_N) \ll 1$ , then no second beneficial allele survives during the spreading of the first allele. Our assumption  $1 \ll N\mu_N$  is slightly stronger than  $1 \ll N\mu_N \ln(2Ns_N)$ .

Our third assumption,  $N\mu_N^2 \ll s_N$ , ensures that once both beneficial mutations have been established in the population, the selection will be the force that causes both mutations to spread to the entire population. We note that if  $N\mu_N \gg s_N$ , mutations will occur so fast that the fixation of both beneficial alleles might arise solely from the influx of mutations.

Lastly, the condition  $r_N \ln(Nr_N) \ll s_N$  says that the recombination does not occur too frequently that it overwhelms the effect of the selection. This condition is similar to the condition  $r \ll s[L/\ln(Nr)]^{1/2}$ , which appeared in [Rouzine and Coffin \(2010\)](#) and was described as the ‘‘infrequent recombination’’ condition.

#### 1.4. Main theorem.

**Theorem 1.1.** *Let  $T$  be the first time that all individuals in the population are type  $AB$ , which we also call the fixation time of  $AB$ . For every positive integer  $N$ , and  $r \in [0, 1]$ , we define*

$$t_N^*(r) = \frac{1}{s_N} \ln \left( \frac{Ns_N^3}{\mu_N \cdot (N\mu_N^2 \vee r \ln_+(Nr))} \right). \quad (1.5)$$

Then, for every  $\theta \in (0, 1)$ ,

$$\lim_{N \rightarrow \infty} P((1 - \theta)t_N^*(r_N) \leq T \leq (1 + \theta)t_N^*(r_N)) = 1.$$

This theorem suggests that the fixation time of both beneficial alleles is approximately  $t_N^*(r_N)$  when  $N$  is large. From (1.5), when there is no recombination, the fixation time of the two beneficial alleles is approximately

$$t_N^*(0) = \frac{1}{s_N} \ln \left( \frac{s_N^3}{\mu_N^3} \right).$$

When  $r_N \ln_+(Nr_N) > N\mu_N^2$ , we observe that  $t_N^*(r_N) < t_N^*(0)$ . Therefore, when  $r_N$  is sufficiently large, recombination decreases the fixation time of  $AB$ .

From (1.3) and (1.4), we have that  $N\mu_N^2 \vee r_N \ln_+(Nr_N) < s_N$  for sufficiently large  $N$ . Hence,

$$t_N^*(r_N) \leq \frac{1}{s_N} \ln \left( \frac{s_N^3}{\mu_N^3} \right) = t_N^*(0),$$

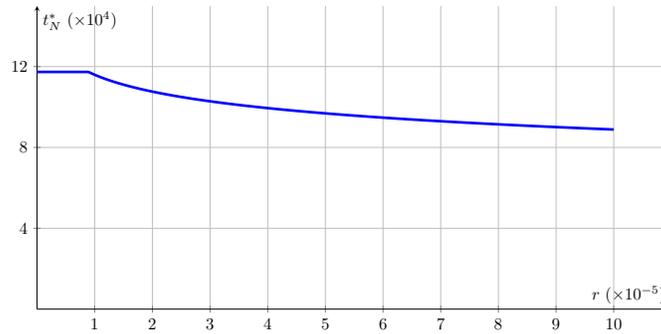


FIGURE 1.1. The graph shows the relation between  $t_N^*$  and  $r$  when  $N = 10^7$ ,  $\mu = 2 \times 10^{-6}$  and  $s = 10^{-4}$ . The  $r$ -axis is scaled by  $10^{-5}$ , and the  $t_N^*$ -axis is scaled by  $10^4$ .

and

$$t_N^*(r_N) > \frac{1}{s_N} \ln \left( \frac{Ns_N^2}{\mu_N} \right) = \frac{2}{3} \cdot \frac{1}{s_N} \ln \left( \frac{s_N^3}{\mu_N^3} \right) + \frac{1}{s_N} \ln(N\mu_N) > \frac{2}{3}t_N^*(0).$$

These two inequalities imply that recombination can decrease the fixation time of  $AB$  by no more than a factor of one-third.

One may think that recombination might be able to reduce the fixation time of type  $AB$  more than a factor of one-third. This limiting factor of  $2/3$  on the fixation time may arise from two possible sources: the recombination rate is not large enough, or two loci is too few to see the substantial effect of recombination. With a greater recombination rate, both loci are acting almost independently of each other, and the fixation time might reduce further beyond the factor of one-third. However, we are not certain and further study is needed. With more loci, we might be able to see a more significant effect of recombination on the population.

At this point, we briefly mention that our assumptions on the parameters are attainable. For example, when  $\mu_N = N^{-a}$ ,  $r_N = N^{-b}$ , and  $s_N = N^{-c}$  for some positive numbers  $a, b$ , and  $c$ , one can check that (1.1), (1.2), (1.3), and (1.4) are equivalent to  $0 < c < b$  and  $(1 + c)/2 < a < 1$ .

## 2. Overview of the proof

From now on, we will refer to individuals with  $ab$ ,  $Ab$ ,  $aB$ , and  $AB$  as types 0, 1, 2, and 3, respectively, and we will omit writing the subscript  $N$  in  $\mu_N$ ,  $s_N$ , and  $r_N$ .

For  $i = 0, 1, 2, 3$  and  $t \geq 0$ , we define  $X_i(t)$  as the number of type  $i$  individuals at time  $t$  and define  $\bar{X}_i(t) = X_i(t)/(2N)$ , which is the fraction of type  $i$  individuals in the population at time  $t$ .

Now, we consider our process  $((X_0(t), X_1(t), X_2(t), X_3(t)), t \geq 0)$ . The behavior of this process is essentially divided into two cases. In the first case, which we will call the *recombination dominating case*, we assume that

$$N\mu^2 \ll r \ln(Nr) \ll s. \tag{2.1}$$

In the second case, which we name the *mutation dominating case*, we assume that there is a positive constant  $C$  such that for sufficiently large  $N$ ,

$$r \ln_+(Nr) \leq CN\mu^2. \quad (2.2)$$

The reason behind these names is as follow: in the recombination dominating case, type 3 individuals start to appear from recombination between  $A$  alleles from type 1 individuals and  $B$  alleles from type 2 individuals, while in the mutation dominating case, the type 3 individuals start to appear from mutations from type 1 and type 2 individuals.

Before we proceed to discuss the behavior of the process, we will prove some relations between our parameters.

**Lemma 2.1.** *The following statements hold.*

(1) *In the recombination dominating case,*

$$1 \ll Nr. \quad (2.3)$$

(2) *In the mutation dominating case,*

$$r \ll N\mu^2.$$

(3) *In both cases,*

$$r \ll s, \quad (2.4)$$

$$\frac{r}{s} \ln(Ns) \ll 1, \quad (2.5)$$

and

$$\frac{r}{s} \ln\left(\frac{s}{\mu}\right) \ll 1. \quad (2.6)$$

*Proof:* We will first prove statement (1). In the recombination dominating case, from conditions (1.2) and (2.1),

$$1 \ll (N\mu)^2 \ll Nr \ln(Nr),$$

which implies that  $1 \ll Nr$ .

Next, we prove statement (2) by contradiction. Suppose there is a constant  $c > 0$  and an increasing sequence  $\{N_k\}_{k=1}^\infty$  of natural numbers such that for all  $k = 1, 2, 3, \dots$ , we have

$$r_{N_k} > cN_k\mu_{N_k}^2.$$

From (2.2), we have that for all  $k = 1, 2, 3, \dots$ ,

$$cN_k\mu_{N_k}^2 \ln_+(cN_k^2\mu_{N_k}^2) \leq r_{N_k} \ln_+(N_k r_{N_k}) \leq CN_k\mu_{N_k}^2.$$

This leads to a contradiction, because  $1 \ll N\mu$  implies that

$$\ln_+(cN_k^2\mu_{N_k}^2) \rightarrow \infty,$$

as  $k \rightarrow \infty$ .

Lastly, we show statement (3). We will first consider the recombination dominating case. By (1.4) and (2.3),

$$r \ll r \ln(Nr) \ll s.$$

From (2.1) and (2.4), it follows that

$$\frac{r}{s} \ln(Ns) = \frac{r}{s} \ln(Nr) + \frac{r}{s} \ln\left(\frac{s}{r}\right) \ll 1.$$

Also, because of (1.2), for sufficiently large  $N$ ,

$$\frac{r}{s} \ln\left(\frac{s}{\mu}\right) \leq \frac{r}{s} \ln(Ns) \ll 1,$$

which implies (2.6). For the mutation dominating case, we define  $r_N^*$  such that  $Nr_N^*$  is the solution of

$$x \ln(x) = \sqrt{(N\mu)^2 \cdot Ns}.$$

It follows that  $N\mu^2 \ll r_N^* \ln(Nr_N^*) \ll s$ . Therefore, by the same argument above,

$$r_N^* \ll s, \tag{2.7}$$

$$\frac{r_N^*}{s} \ln(Ns) \ll 1, \tag{2.8}$$

and

$$\frac{r_N^*}{s} \ln\left(\frac{s}{\mu}\right) \ll 1. \tag{2.9}$$

From (2.2) and the fact that  $N\mu^2 \ll r_N^* \ln(Nr_N^*)$ , we have  $r_N \leq r_N^*$  for sufficiently large  $N$ . Thus, (2.4), (2.5), and (2.6) hold.  $\square$

We note that  $\mu \ll r \ll s \ll 1$  in the recombination dominating case. So, in this case, the effect from the recombination is not as strong as that from selection, yet stronger than that from mutation. In the mutation dominating case, we still have  $\mu \ll s$  and  $r \ll s$ , but we cannot conclude that  $r \ll \mu$ . Nevertheless, we still have  $r \ll N\mu^2$ , which tells us that recombination has the weakest effect in the mutation dominating case.

Now, in the following table, we define times when we see significant changes in the behavior of the process.

Time	Recombination dominating	Mutation dominating	Behavior
$t_0$	$\frac{1}{s} \ln\left(\frac{s}{\mu\sqrt{Nr}}\right) - \frac{C_{0,r}}{s}$	$\frac{1}{s} \ln\left(\frac{s}{N\mu^2}\right) - \frac{C_{0,m}}{s}$	Type 3 appears in the population.
$t_1$	$\frac{1}{s} \ln\left(\frac{s}{\mu}\right) - \frac{C_1}{s}$	$\frac{1}{s} \ln\left(\frac{s}{\mu}\right) - \frac{C_1}{s}$	Types 1 and 2 become established.
$t_2$	$\frac{1}{s} \ln\left(\frac{s}{\mu}\right) + \frac{C_2}{s}$	$\frac{1}{s} \ln\left(\frac{s}{\mu}\right) + \frac{C_2}{s}$	Most individuals become type 1 or 2.
$t_3$	$\frac{1}{s} \ln\left(\frac{s^2}{\mu r \ln(Nr)}\right) - \frac{C_3}{s}$	$\frac{1}{s} \ln\left(\frac{s^2}{N\mu^3}\right) - \frac{C_3}{s}$	Type 3 becomes established in the population.
$t_4$	$\frac{1}{s} \ln\left(\frac{s^2}{\mu r \ln(Nr)}\right) + \frac{c_4}{s}$	$\frac{1}{s} \ln\left(\frac{s^2}{N\mu^3}\right) + \frac{c_4}{s}$	Most individuals has become type 3.

Here, becoming established means that the number of that type has reached the order  $N$  but remains a small fraction of the population. The constants  $C_{0,r}$ ,  $C_{0,m}$ ,  $C_1$ ,  $C_2$ ,  $C_3$ , and  $c_4$  depend on  $\epsilon$  and  $\delta$  and can be arbitrarily large when  $\epsilon$  and  $\delta$  are sufficiently small. All of these constants except  $c_4$  are positive. Their precise definitions can be found in (5.7), (5.5), (5.4), (6.12), (7.1), and (8.4). However, the reader does not need to know what these constants are precisely at this point. It is

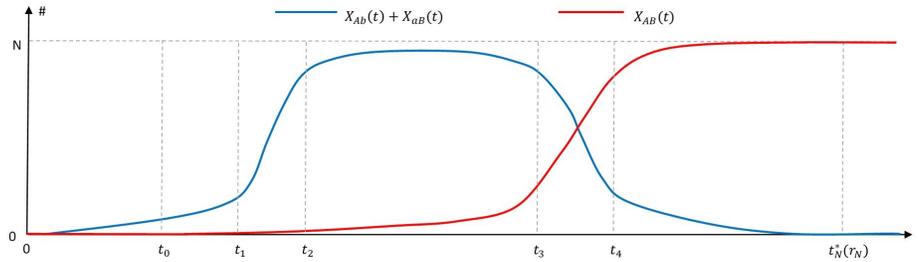


FIGURE 2.2. The graphs of approximate numbers of individuals with one beneficial mutation (blue) and two beneficial mutations (red).

enough to notice that  $C_i/s$  is the lower order term in the definition of the  $t_i$ . Also,  $t_0 < t_1 < t_2 < t_3 < t_4$  for sufficiently large  $N$  because of our assumptions on the parameters and Lemma 2.1.

From now on, all statements are assumed to be correct in the recombination dominating case and the mutation dominating case unless stated otherwise.

Overall, the behavior of the numbers of types 1, 2, and 3 individuals are similar in the sense that they first grow exponentially, then grow logistically. From the start, the numbers of type 1 and type 2 individuals both grow simultaneously, but the number of type 3 individuals will start to grow late due to the late appearance of type 3 individuals. The behavior of the process can be split according to the time into five time intervals.

During the time interval  $[0, t_1]$ , which we will call *phase 1*, most individuals are type 0. Type 1 and type 2 individuals appear from mutations from type 0 individuals. Since type 1 and type 2 individuals die at rate  $1 - s$ , while the majority of the population, which is type 0, dies at rate 1, the numbers of type 1 and 2 individuals grow exponentially at a rate approximately  $s$ . Since the total rate of mutation from type 0 to type 1 is approximately  $2N\mu$ , we have

$$X_1(t) \approx \int_0^t 2N\mu \cdot e^{s(t-u)} du \approx \frac{2N\mu}{s} e^{st}.$$

By analogy, this approximation is also valid for the number of type 2. Here, we are going to emphasize that this approximation depends on the relative sizes of our parameters  $\mu, r$ , and  $s$  given in subsection 1.3. It does not hold in the diffusion scaling nor in the scaling in Cuthbertson et al. (2012).

The type 3 individuals appear around time  $t_0$ . From this time, the number of type 3 individuals will grow exponentially at a rate of about  $2s$  because each type 3 individual dies at rate  $1 - 2s$ , and most individuals in the population die at rate 1. The following proposition describes the process at time  $t_1$ .

**Proposition 2.2.** *For  $\epsilon > 0$  and  $\delta \in (0, 1)$ , there is an event  $A_{(1)}$ , such that for sufficiently large  $N$ , we have that  $P(A_{(1)}) \geq 1 - 17\epsilon$ , and the following statements hold:*

- (1) *On the event  $A_{(1)}$ , when  $N$  is sufficiently large, for  $i = 1, 2$ ,*

$$2(1 - \delta^2)e^{-C_1 N} \leq X_i(t_1) \leq 2(1 + \delta^2)e^{-C_1 N}. \tag{2.10}$$

- (2) In the recombination dominating case, on the event  $A_{(1)}$ , there are positive constants  $K_{1r}^+$  and  $K_{1r}^-$  such that for sufficiently large  $N$ ,

$$\frac{K_{1r}^-Nr \ln(Nr)}{s} \leq X_3(t_1) \leq \frac{K_{1r}^+Nr \ln(Nr)}{s}. \tag{2.11}$$

- (3) In the mutation dominating case, on the event  $A_{(1)}$ , there are positive constants  $K_{1m}^+$  and  $K_{1m}^-$  such that for sufficiently large  $N$ ,

$$\frac{K_{1m}^-N^2\mu^2}{s} \leq X_3(t_1) \leq \frac{K_{1m}^+N^2\mu^2}{s}. \tag{2.12}$$

This proposition says that when  $N$  is sufficiently large, at time  $t_1$ , both type 1 and type 2 individuals have established themselves in the population. However,  $\tilde{X}_3(t_1)$  is only of order  $r \ln(Nr)/s$  in the recombination dominating case and is only of order  $N\mu^2/s$  in the mutation dominating case: the order is smaller than  $N$ . Hence, the number of type 3 at time  $t_1$  is not yet comparable to the numbers of type 1 and type 2.

During the time interval  $[t_1, t_2]$ , which we will call *phase 2*, the numbers of type 1 and 2 now grow logistically; more precisely,

$$\tilde{X}_i(t) \approx \frac{1}{2} \left( \frac{1}{1 + Be^{-s(t-t_1)}} \right),$$

for  $i = 1, 2$ , where  $B$  is some positive constant. A basic fact on the spreading of a single beneficial allele has already told us that this logistic growth phase should last for a time of order  $1/s$ , which is exactly the difference between  $t_1$  and  $t_2$ . The following proposition describes the process at time  $t_2$ .

**Proposition 2.3.** *For  $\epsilon > 0$  and  $\delta \in (0, 1)$ , there is an event  $A_{(2)}$ , such that for sufficiently large  $N$ , we have that  $P(A_{(2)}) \geq 1 - 21\epsilon$ , and the following statements hold:*

- (1) On the event  $A_{(2)}$ , for sufficiently large  $N$ , for  $i = 1, 2$ ,

$$(1 - 3\delta^2)N \leq X_i(t_2) \leq \left(1 - \frac{\delta^4}{2}\right)N.$$

- (2) In the recombination dominating case, on the event  $A_{(2)}$ , there are positive constants  $K_{2r}^+$  and  $K_{2r}^-$  such that for sufficiently large  $N$ ,

$$\frac{K_{2r}^-Nr \ln(Nr)}{s} \leq X_3(t_2) \leq \frac{K_{2r}^+Nr \ln(Nr)}{s}.$$

- (3) In the mutation dominating case, on the event  $A_{(2)}$ , there are positive constants  $K_{2m}^+$  and  $K_{2m}^-$  such that for sufficiently large  $N$ ,

$$\frac{K_{2m}^-N^2\mu^2}{s} \leq X_3(t_2) \leq \frac{K_{2m}^+N^2\mu^2}{s}.$$

This proposition says that at time  $t_2$ , almost half of the population becomes type 1, and most of the other half becomes type 2. In contrast, the number of type 3 individuals doesn't change much from time  $t_1$  and has the same order as it was at time  $t_1$ .

During the time interval  $[t_2, t_3]$ , which we will call *phase 3*, the majority of the population has become type 1 or type 2. The number of type 3 individuals continues to grow exponentially from time  $t_2$ . Since the majority of the population dies at rate  $1 - s$ , and a type 3 individual dies at rate  $1 - 2s$ , the type 3 population

grows exponentially at approximately rate  $s$ . The following proposition describes the behavior of the process at time  $t_3$ .

**Proposition 2.4.** *For  $\epsilon > 0$  and  $\delta \in (0, 1)$ , there is an event  $A_{(3)}$ , such that for sufficiently large  $N$ , we have that  $P(A_{(3)}) \geq 1 - 25\epsilon - 7\delta - \delta^2$ , and the following statements hold:*

(1) *For sufficiently large  $N$ , on the event  $A_{(3)}$ , we have*

$$X_0(t_3) < 2\delta e^{-(1-3\delta)(C_3-C_2)} N \cdot \left(\frac{r \ln(Nr)}{s}\right)^{1-3\delta}$$

*in the recombination dominating case, and*

$$X_0(t_3) < 2\delta e^{-(1-3\delta)(C_3-C_2)} N \cdot \left(\frac{N\mu^2}{s}\right)^{1-3\delta}$$

*in the mutation dominating case.*

(2) *In both cases, there is a positive constant  $K_3$  such that for sufficiently large  $N$ , on the event  $A_{(3)}$ , we have*

$$K_3 N \leq X_3(t_3) \leq \delta^2 N.$$

The first part of this proposition is slightly technical, but the main idea is that the number of type 0 individuals at time  $t_3$  has an order smaller than  $N$ ; only a small number of type 0 individuals remain in the population. The second part says that the number of type 3 individuals has reached order  $N$ ; type 3 has been established.

During the time interval  $[t_3, t_4]$ , which we will call *phase 4*, the number of type 3 individuals now grows logistically, similar to the behavior of type 1 and type 2 populations during phase 2. The following proposition describes the behavior of the process at time  $t_4$ .

**Proposition 2.5.** *For  $\epsilon > 0$  and  $\delta \in (0, 1)$ , there is an event  $A_{(4)}$ , such that for sufficiently large  $N$ , we have that  $P(A_{(4)}) \geq 1 - 26\epsilon - 7\delta - \delta^2$ , and on the event  $A_{(4)}$ ,*

$$\left(2 - \frac{5\delta^2}{2}\right) N \leq X_3(t_4) \leq \left(2 - \frac{3K_3}{2}\right) N,$$

*and*

$$X_1(t_4) + X_2(t_4) \geq K_3 N.$$

This proposition implies that by time  $t_4$ , almost all individuals have become type 3, and only small fractions of type 1 and 2 individuals remain in the population.

After time  $t_4$ , which we will call *phase 5*, the number of individuals that are not of type 3 can be approximated by a subcritical branching process. The non-type 3 population is heading toward extinction, and type 3 becomes fixated in the population. From the basic fact of the fixation time of one beneficial allele, the time that the beneficial allele takes to fixate, starting from the end of its logistic growth phase, is approximately  $\frac{1}{s} \ln(2Ns)$ . This should hold in our situation as well. Thus, the fixation of type 3 occurs around time  $t_N^*(r_N)$ .

The proofs of these propositions will be split into sections corresponding to the phases. However, there are several facts that we need to establish before the proof. In section 3, we will discuss transition rates related to our process. In section 4, we construct martingales and submartingales, and give expectation and variance

formulas. We will use these results intensively in the proofs of results in phases 1, 2, and 3. In section 5, we prove several lemmas on the process during phase 1 and give a proof of Proposition 2.2 at the end of the section. Proposition 2.3, 2.4, and 2.5 will be proved in sections 6, 7, and 8, respectively. Finally, the proof of Theorem 1.1 will be given at the end of section 9. More details will be given in each section of the proof.

### 3. Transition rates

For the proof, we will separate type 1 individuals into two groups: one for those that come from mutation from type 0 individuals and another for individuals that come from recombination between type 0 and type 3 individuals. We need to do the same for the other three types. The precise definitions are given below.

- (1) A type 1 (or 2) individual is called a **type 1m (or 2m) ancestor** if it appears by mutation from a type 0 individual.
- (2) A type 1 (or 2) individual is called a **type 1r (or 2r) ancestor** if it appears by recombination between a  $b$  (or an  $a$ ) allele from a type 0 individual and an  $A$  (or a  $B$ ) allele from a type 3 individual.
- (3) A type 1 individual  $x$  is called an offspring of another type 1 individual  $y$  if
  - $x$  receives the  $A$  allele from  $y$ , or
  - $x$  receives the  $b$  allele from  $y$  and receives the  $A$  allele from a type 3 individual.
- (4) A type 2 individual  $x$  is called an offspring of another type 2 individual  $y$  if a
  - $x$  receives the  $B$  allele from  $y$ , or
  - $x$  receives the  $a$  allele from  $y$  and receives the  $B$  allele from a type 3 individual.
- (5) A type 1 (or 2) individual is called **type 1m (or 2m)** if it descends from a type 1m (or 2m) ancestor. A type 1 (or 2) individual is called **type 1r (or 2r)** if it descends from a type 1r (or 2r) ancestor.
- (6) A type 3 individual is called a **type 3m ancestor** if it appears from a mutation from a type 1 individual or a type 2 individual.
- (7) A type 3 individual is called a **type 3r ancestor** if it appears by recombination between an  $A$  allele from a type 1 individual and a  $B$  allele from a type 2 individual.
- (8) A type 3 individual  $x$  is called an offspring of another type 3 individual  $y$  if
  - $x$  receives the  $A$  allele from  $y$ , or
  - $x$  receives the  $B$  allele from  $y$  and receives the  $A$  allele from a type 1 individual.
- (9) A type 3 individual is called **type 3m** if it descends from a type 3m ancestor. A type 3 individual is called **type 3r** if it descends from a type 3r ancestor.
- (10) A type 0 individual is called a **type 0r ancestor** if it appears from recombination between an  $a$  allele from a type 1 individual and a  $b$  allele from a type 2 individual.
- (11) A type 0 individual  $x$  is called an offspring of another type 0 individual  $y$  if
  - $x$  receives the  $a$  allele from  $y$ , or
  - $x$  receives the  $b$  allele from  $y$  and receives the  $a$  allele from a type 2.

(12) A type 0 individual is called a **type 0r** if it descends from a type 0r ancestor.

The reasons that we have to define ancestors and descendants are as follows. Early in the process, both  $A$  and  $B$  alleles appear in the population as type 1 and type 2 individuals through mutations. The numbers of these early mutations are large, and the majority of type 1 and 2 individuals at time  $t_1$  would be traced back to these ancestors. This is not the scenario where one mutation event spreads the mutation to the entire population. Also, type 3 individuals that appear around time  $t_0$  come from mutations of type 1 and type 2 individuals or from recombination of  $A$  and  $B$  alleles from type 1 and 2 individuals. Whether these early type 3 individuals appear from mutation or recombination will affect the number of type 3 at time  $t_1$ . By defining ancestors, one may view the birth of each ancestor as an immigration event.

Here, we define notations and terms that will be used throughout the proof. For  $i = 1, 2, 3$ , we define  $X_{im}(t)$  as the number of type  $im$  at time  $t$ , and for  $i = 0, 1, 2, 3$ , we define  $X_{ir}(t)$  as the number of type  $ir$  at time  $t$ . Note that  $X_i(t) = X_{im}(t) + X_{ir}(t)$  for  $i = 1, 2, 3$  and  $t \geq 0$ . Next, we define  $X_{im}^{(a,b]}(t)$  to be the number, at time  $t$ , of type  $im$  individuals whose ancestor appears in the time interval  $(a, b]$ . We also define  $X_{ir}^{(a,b]}(t)$  for type  $ir$ . It follows that if  $0 \leq t \leq b$ , we have that  $X_{im}^{(0,b]}(t) = X_{im}(t)$  for  $i = 1, 2, 3$ , and  $X_{ir}^{(0,b]}(t) = X_{ir}(t)$  for  $i = 0, 1, 2, 3$ . We will call an individual type **im(a,b]** (or **ir(a,b]**), if it is of type  $im$  (or type  $ir$ ) and its ancestor appears in the time interval  $(a, b]$ . Lastly, we define  $\tilde{X}_{im}(t)$ ,  $\tilde{X}_{ir}(t)$ ,  $\tilde{X}_{im}^{(a,b]}(t)$ , and  $\tilde{X}_{ir}^{(a,b]}(t)$  to be the fractions of type  $im$ ,  $ir$ ,  $im(a,b]$  and  $ir(a,b]$  in the population at time  $t$ , respectively.

Now, we first consider the rate that  $X_{1m}^{(a,b]}(t)$  increases by 1. There are only two scenarios that increase  $X_{1m}^{(a,b]}(t)$ . The first scenario is when a type 0 individual mutates to a type 1 individual during the time interval  $(a, b]$ , creating a type  $1m(a,b]$  ancestor, which occurs at the total rate

$$M_1^{(a,b]}(t) := \mu X_0(t) 1_{(a,b]}(t). \tag{3.1}$$

The second scenario is when an individual that is not of type  $1m(a,b]$  die, which occurs at the total rate

$$X_0(t) + (1 - s)(X_1(t) - X_{1m}^{(a,b]}(t)) + (1 - s)X_2(t) + (1 - 2s)X_3(t), \tag{3.2}$$

and the replacement individual must have type  $1m(a,b]$ . The probability that recombination doesn't occur and the new individual has type  $1m(a,b]$  is  $(1 - r)\tilde{X}_{1m}^{(a,b]}(t)$ . If recombination occurs, the new type  $1m(a,b]$  individual may come from combining an  $A$  allele from a type  $1m(a,b]$  individual with a  $b$  allele from a type 0 or 1 individual, or combining an  $A$  allele from a type 3 individual with a  $b$  allele from a type  $1m(a,b]$  individual. (Note that recombination between an  $A$  allele from a type 3 individual and a  $b$  allele from a type 0 individual creates an ancestor of type  $1r$  instead.) So, the probability that recombination occurs and the new individual has type  $1m(a,b]$  is

$$\begin{aligned} & r\left(\tilde{X}_{1m}^{(a,b]}(t)\tilde{X}_0(t) + \tilde{X}_{1m}^{(a,b]}(t)\tilde{X}_1(t) + \tilde{X}_3(t)\tilde{X}_{1m}^{(a,b]}(t)\right) \\ & = r\tilde{X}_{1m}^{(a,b]}(t)\left(\tilde{X}_0(t) + \tilde{X}_1(t) + \tilde{X}_3(t)\right). \end{aligned} \tag{3.3}$$

Hence, the total rate that the number of descendants of type 1m(a,b] increases by 1 is

$$\begin{aligned} & \left( X_0(t) + (1-s)(X_1(t) - X_{1m}^{(a,b]}(t)) + (1-s)X_2(t) + (1-2s)X_3(t) \right) \\ & \cdot \left( (1-r)\tilde{X}_{1m}^{(a,b]}(t) + r\tilde{X}_{1m}^{(a,b]}(t)(\tilde{X}_0(t) + \tilde{X}_1(t) + \tilde{X}_3(t)) \right). \end{aligned}$$

Let us define

$$\begin{aligned} B_{1m}^{(a,b]}(t) &= \left( \tilde{X}_0(t) + (1-s)(\tilde{X}_1(t) - \tilde{X}_{1m}^{(a,b]}(t)) + (1-s)\tilde{X}_2(t) + (1-2s)\tilde{X}_3(t) \right) \\ & \cdot \left( 1 - r\tilde{X}_2(t) \right). \end{aligned} \tag{3.4}$$

We will refer to  $B_{1m}^{(a,b]}(t)$  as the birth rate of type 1m(a,b] at time  $t$ . Note that  $X_{1m}^{(a,b]}(t)$  increases by 1 at rate  $M_1^{(a,b]}(t) + B_{1m}^{(a,b]}(t)X_{1m}^{(a,b]}(t)$ .

Similarly, the rate that the number of type 1m(a,b] individuals decreases by 1 is given by

$$(1-s)X_{1m}^{(a,b]}(t) \left( 1 - (1-r)\tilde{X}_{1m}^{(a,b]}(t) - r\tilde{X}_{1m}^{(a,b]}(t)(\tilde{X}_0(t) + \tilde{X}_1(t) + \tilde{X}_3(t)) \right) + \mu X_{1m}^{(a,b]}(t). \tag{3.5}$$

We note that  $(1-s)X_{1m}^{(a,b]}(t)$  is the total rate that type 1m(a,b] individuals die at time  $t$ . The term

$$1 - (1-r)\tilde{X}_{1m}^{(a,b]}(t) - r\tilde{X}_{1m}^{(a,b]}(t)(\tilde{X}_0(t) + \tilde{X}_1(t) + \tilde{X}_3(t))$$

is the probability that the new individual is not a type 1m(a,b]. The last term  $\mu X_{1m}^{(a,b]}(t)$  corresponds to the total rate that type 1m(a,b] mutates to type 3. We define

$$D_{1m}^{(a,b]}(t) = (1-s) \left( 1 - \tilde{X}_{1m}^{(a,b]}(t) + r\tilde{X}_2(t)\tilde{X}_{1m}^{(a,b]}(t) \right) + \mu, \tag{3.6}$$

which we will refer to as the death rate of type 1m(a,b] at time  $t$ . Again, note that the number of type 1m(a,b] individuals decreases by 1 at rate  $D_{1m}^{(a,b]}(t)X_{1m}^{(a,b]}(t)$ .

Next, we consider the rate that  $X_{1r}^{(a,b]}(t)$  increases by 1. There are two ways to increase  $X_{1r}^{(a,b]}(t)$  by 1. First, an individual that is not of type 1r(a,b] dies, and the recombination between an  $A$  allele from a type 3 individual and a  $b$  allele from a type 0 individual occurs during the time interval  $(a, b]$ , creating a type 1r(a,b] ancestor. This occurs at the total rate of

$$\begin{aligned} R_1^{(a,b]}(t) &= \left( X_0(t) + (1-s)(X_1(t) - X_{1r}^{(a,b]}(t)) + (1-s)X_2(t) + (1-2s)X_3(t) \right) \\ & \cdot \left( r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t) \right). \end{aligned}$$

Second, an individuals that is not of type 1r(a,b] dies, and a new type 1r(a,b] individual is born from the type 1r(a,b] individuals at that time. Similar to the way we obtain (3.2) and (3.3), we define

$$\begin{aligned} B_{1r}^{(a,b]}(t) &= \left( \tilde{X}_0(t) + (1-s)(\tilde{X}_1(t) - \tilde{X}_{1r}^{(a,b]}(t)) + (1-s)\tilde{X}_2(t) + (1-2s)\tilde{X}_3(t) \right) \\ & \cdot \left( 1 - r\tilde{X}_2(t) \right), \end{aligned} \tag{3.7}$$

which we will refer to as the birth rate of type 1r(a,b] at time  $t$ . One can see that the rate that  $X_{1r}^{(a,b]}(t)$  increases by 1 is  $R_1^{(a,b]}(t) + B_{1r}^{(a,b]}(t)X_{1r}^{(a,b]}(t)$ .

We now consider the rate that  $X_{1r}^{(a,b)}(t)$  decreases by 1. One way that  $X_{1r}^{(a,b)}(t)$  decreases by 1 is when a type 1r(a,b] individual dies, and the new individual is not of type 1r(a,b] (i.e, the new individual is not born from a type 1r(a,b] individual, and it is not a type 1r(a,b] ancestor). Another way is when a type 1r(a,b] individual mutates to a type 3 individual. By the same reason we used to obtain (3.5), the rate that  $X_{1r}^{(a,b)}(t)$  decreases by 1 is

$$(1-s)X_{1r}^{(a,b)}(t)\left(1-(1-r)\tilde{X}_{1r}^{(a,b)}(t)-r\tilde{X}_{1r}^{(a,b)}(t)(\tilde{X}_0(t)+\tilde{X}_1(t)+\tilde{X}_3(t))\right. \\ \left.-r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t)\right)+\mu X_{1r}^{(a,b)}(t).$$

We note that the term  $r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t)$  is precisely the probability that a type 1r(a,b] ancestor is created. By defining

$$D_{1r}^{(a,b)}(t)=(1-s)\left(1-\tilde{X}_{1r}^{(a,b)}(t)+r\tilde{X}_2(t)\tilde{X}_{1r}^{(a,b)}(t)-r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t)\right)+\mu, \quad (3.8)$$

which we will refer to as the death rate of type 1r(a,b] at time  $t$ , one can see that the rate that  $X_{1r}^{(a,b)}(t)$  decreases by 1 is  $D_{1r}^{(a,b)}(t)X_{1r}^{(a,b)}(t)$ .

By analogy, we should define the rates for the other types as follows.

$$B_{2m}^{(a,b)}(t)=\left(\tilde{X}_0(t)+(1-s)\tilde{X}_1(t)+(1-s)(\tilde{X}_2(t)-\tilde{X}_{2m}^{(a,b)}(t))+(1-2s)\tilde{X}_3(t)\right) \\ \cdot\left(1-r\tilde{X}_1(t)\right),$$

$$D_{2m}^{(a,b)}(t)=(1-s)\left(1-\tilde{X}_{2m}^{(a,b)}(t)+r\tilde{X}_1(t)\tilde{X}_{2m}^{(a,b)}(t)\right)+\mu,$$

$$M_2^{(a,b)}(t)=\mu X_0(t)1_{(a,b]}(t),$$

$$B_{2r}^{(a,b)}(t)=\left(\tilde{X}_0(t)+(1-s)\tilde{X}_1(t)+(1-s)(\tilde{X}_2(t)-\tilde{X}_{2r}^{(a,b)}(t))+(1-2s)\tilde{X}_3(t)\right) \\ \cdot\left(1-r\tilde{X}_1(t)\right),$$

$$D_{2r}^{(a,b)}(t)=(1-s)\left(1-\tilde{X}_{2r}^{(a,b)}(t)+r\tilde{X}_1(t)\tilde{X}_{2r}^{(a,b)}(t)-r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t)\right)+\mu,$$

$$R_2^{(a,b)}(t)=\left(X_0(t)+(1-s)X_1(t)+(1-s)(X_2(t)-X_{2r}^{(a,b)}(t))+(1-2s)X_3(t)\right) \\ \cdot\left(r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t)\right),$$

$$B_{3m}^{(a,b)}(t)=\left(\tilde{X}_0(t)+(1-s)(\tilde{X}_1(t)+\tilde{X}_2(t))+(1-2s)(\tilde{X}_3(t)-\tilde{X}_{3m}^{(a,b)}(t))\right) \\ \left(1-r\tilde{X}_0(t)\right), \quad (3.9)$$

$$D_{3m}^{(a,b)}(t)=(1-2s)\left(1-\tilde{X}_{3m}^{(a,b)}(t)+r\tilde{X}_0(t)\tilde{X}_{3m}^{(a,b)}(t)\right), \quad (3.10)$$

$$M_3^{(a,b)}(t)=\mu(X_1(t)+X_2(t))1_{(a,b]}(t), \quad (3.11)$$

$$B_{3r}^{(a,b)}(t)=\left(\tilde{X}_0(t)+(1-s)(\tilde{X}_1(t)+\tilde{X}_2(t))+(1-2s)(\tilde{X}_3(t)-\tilde{X}_{3r}^{(a,b)}(t))\right) \\ \left(1-r\tilde{X}_0(t)\right), \quad (3.12)$$

$$D_{3r}^{(a,b)}(t)=(1-2s)\left(1-\tilde{X}_{3r}^{(a,b)}(t)+r\tilde{X}_0(t)\tilde{X}_{3r}^{(a,b)}(t)-r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t)\right), \quad (3.13)$$

$$\begin{aligned}
R_3^{(a,b]}(t) &= \left( X_0(t) + (1-s)(X_1(t) + X_2(t)) + (1-2s)(X_3(t) - X_{3r}^{(a,b]}(t)) \right) \\
&\quad \cdot \left( r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t) \right), \tag{3.14} \\
B_{0r}^{(a,b]}(t) &= \left( (\tilde{X}_0(t) - \tilde{X}_{0r}^{(a,b]}(t)) + (1-s)(\tilde{X}_1(t) + \tilde{X}_2(t)) + (1-2s)\tilde{X}_3(t) \right) \\
&\quad \cdot \left( 1 - r\tilde{X}_3(t) \right), \\
D_{0r}^{(a,b]}(t) &= \left( 1 - \tilde{X}_{0r}^{(a,b]}(t) + r\tilde{X}_3(t)\tilde{X}_{0r}^{(a,b]}(t) - r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t) \right) + 2\mu, \\
R_0^{(a,b]}(t) &= \left( (X_0(t) - X_{0r}^{(a,b]}(t)) + (1-s)(X_1(t) + X_2(t)) + (1-2s)X_3(t) \right) \\
&\quad \cdot \left( r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t) \right).
\end{aligned}$$

Again, one can check that  $X_{im}^{(a,b]}(t)$  increases by 1 at rate  $M_i^{(a,b]}(t) + B_{im}^{(a,b]}(t)X_{im}^{(a,b]}(t)$  and decreases by 1 at rate  $D_{im}^{(a,b]}(t)X_{im}^{(a,b]}(t)$  for  $i = 2, 3$ . Also,  $X_{ir}^{(a,b]}(t)$  increases by 1 at rate  $R_i^{(a,b]}(t) + B_{ir}^{(a,b]}(t)X_{ir}^{(a,b]}(t)$  and decreases by 1 at rate  $D_{ir}^{(a,b]}(t)X_{ir}^{(a,b]}(t)$  for  $i = 0, 2$ , and 3.

Here, we briefly mention an upper bound of  $R_i^{(a,b]}(t)$ . Since we know that  $s \ll 1$  and  $\tilde{X}_0(t) + \tilde{X}_1(t) + \tilde{X}_2(t) + \tilde{X}_3(t) = 2N$ , for sufficiently large  $N$ ,

$$R_0^{(a,b]}(t) \leq 2Nr\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t), \tag{3.15}$$

$$R_1^{(a,b]}(t) \leq 2Nr\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t), \tag{3.16}$$

$$R_2^{(a,b]}(t) \leq 2Nr\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b]}(t),$$

$$R_3^{(a,b]}(t) \leq 2Nr\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t). \tag{3.17}$$

The upper bound for each  $R_i^{(a,b]}(t)$  essentially captures the dominant term in the expression of  $R_i^{(a,b]}(t)$ .

From these birth and death rates, for  $i = 1, 2, 3$  and  $0 \leq a < b \wedge t$ , we define  $G_i(t) = B_{im}^{(a,b]}(t) - D_{im}^{(a,b]}(t)$ , which represents the growth rate of the type im(a,b] population at time  $t$ . Similarly, for  $i = 0, 1, 2, 3$  and  $0 \leq a < b \wedge t$ , we define  $G_{ir}^{(a,b]}(t) = B_{ir}^{(a,b]}(t) - D_{ir}^{(a,b]}(t)$ . This is the growth rate of the type ir(a,b] population at time  $t$ . Note that  $G_i(t)$  does not depend on the interval  $(a, b]$  because

$$\begin{aligned}
G_1(t) &= B_{1m}^{(a,b]}(t) - D_{1m}^{(a,b]}(t) \\
&= \left( 1 - (1-s)\tilde{X}_{1m}^{(a,b]}(t) - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t) \right) \left( 1 - r\tilde{X}_2(t) \right) \\
&\quad - (1-s) \left( 1 - \tilde{X}_{1m}^{(a,b]}(t) + r\tilde{X}_2(t)\tilde{X}_{1m}^{(a,b]}(t) \right) - \mu \\
&= s \left( 1 - \tilde{X}_1(t) - \tilde{X}_2(t) - 2\tilde{X}_3(t) \right) \\
&\quad - r\tilde{X}_2(t) \left( 1 - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t) \right) - \mu. \tag{3.18}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
G_2(t) &= s \left( 1 - \tilde{X}_1(t) - \tilde{X}_2(t) - 2\tilde{X}_3(t) \right) \\
&\quad - r\tilde{X}_1(t) \left( 1 - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t) \right) - \mu,
\end{aligned}$$

and

$$G_3(t) = s \left( 2 - \tilde{X}_1(t) - \tilde{X}_2(t) - 2\tilde{X}_3(t) \right) - r\tilde{X}_0(t) \left( 1 - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t) \right). \quad (3.19)$$

Also, by similar calculation,

$$G_{1r}^{(a,b)}(t) = G_1(t) + (1-s)r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b)}(t) \quad (3.20)$$

$$G_{2r}^{(a,b)}(t) = G_2(t) + (1-s)r\tilde{X}_0(t)\tilde{X}_3(t)1_{(a,b)}(t) \quad (3.21)$$

$$G_{3r}^{(a,b)}(t) = G_3(t) + (1-2s)r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b)}(t) \quad (3.22)$$

$$\begin{aligned} G_{0r}^{(a,b)}(t) &= -s \left( \tilde{X}_1(t) + \tilde{X}_2(t) + 2\tilde{X}_3(t) \right) \\ &\quad - r\tilde{X}_3(t) \left( 1 - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t) \right) \\ &\quad - 2\mu + r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b)}(t). \end{aligned} \quad (3.23)$$

Now, we introduce a new notation. For  $i = 0, 1, 2, 3$  and  $0 \leq a \leq t$ , we define  $X_i^{[a]}(t)$  to be the number of type  $i$  individuals, at time  $t$ , that descend from one of the type  $i$  individuals at time  $a$ . It follows that for  $0 \leq a \leq t \leq b$  and  $i = 1, 2, 3$ ,

$$X_i(t) = X_i^{[a]}(t) + X_{im}^{(a,b)}(t) + X_{ir}^{(a,b)}(t),$$

and

$$X_0(t) = X_0^{[a]}(t) + X_{0r}^{(a,b)}(t).$$

We will need these notions in the proofs of phases 2 and 3. The reason is that during phase 2, the main contribution to the number of type 3 individuals comes from the type 3 individuals that descend from the type 3 individuals at time  $t_2$ . Thus,  $X_3^{[t_2]}(t)$  will be used to control  $X_3(t)$  in phase 2.

Following the argument we used to obtain  $B_{im}^{(a,b)}(t)$  and  $D_{im}^{(a,b)}(t)$ , for  $0 \leq a \leq t$ , we define

$$\begin{aligned} B_1^{[a]}(t) &= \left( \tilde{X}_0(t) + (1-s)(\tilde{X}_1(t) - X_1^{[a]}(t)) + (1-s)\tilde{X}_2(t) + (1-2s)\tilde{X}_3(t) \right) \\ &\quad \cdot \left( 1 - r\tilde{X}_2(t) \right), \\ D_1^{[a]}(t) &= (1-s) \left( 1 - \tilde{X}_1^{[a]}(t) + r\tilde{X}_2(t)\tilde{X}_1^{[a]}(t) \right) + \mu, \\ B_2^{[a]}(t) &= \left( \tilde{X}_0(t) + (1-s)\tilde{X}_1(t) + (1-s)(\tilde{X}_2(t) - X_2^{[a]}(t)) + (1-2s)\tilde{X}_3(t) \right) \\ &\quad \cdot \left( 1 - r\tilde{X}_1(t) \right), \\ D_2^{[a]}(t) &= (1-s) \left( 1 - \tilde{X}_2^{[a]}(t) + r\tilde{X}_1(t)\tilde{X}_2^{[a]}(t) \right) + \mu, \\ B_3^{[a]}(t) &= \left( \tilde{X}_0(t) + (1-s)\tilde{X}_1(t) + (1-s)\tilde{X}_2(t) + (1-2s)(\tilde{X}_3(t) - X_3^{[a]}(t)) \right) \\ &\quad \cdot \left( 1 - r\tilde{X}_0(t) \right), \end{aligned} \quad (3.24)$$

$$D_3^{[a]}(t) = (1-2s) \left( 1 - \tilde{X}_3^{[a]}(t) + r\tilde{X}_0(t)\tilde{X}_3^{[a]}(t) \right), \quad (3.25)$$

$$\begin{aligned} B_0^{[a]}(t) &= \left( \tilde{X}_0(t) - X_0^{[a]}(t) + (1-s)\tilde{X}_1(t) + (1-s)\tilde{X}_2(t) + (1-2s)\tilde{X}_3(t) \right) \\ &\quad \cdot \left( 1 - r\tilde{X}_3(t) \right), \end{aligned}$$

and

$$D_0^{[a]}(t) = \left(1 - \tilde{X}_0^{[a]}(t) + r\tilde{X}_3(t)\tilde{X}_0^{[a]}(t)\right) + 2\mu.$$

Note that for  $i = 0, 1, 2, 3$ , the process  $(X_i^{[a]}(t), t \geq a)$  increases by 1 at rate  $B_i^{[a]}(t)X_i^{[a]}(t)$  and decreases by 1 at rate  $D_i^{[a]}(t)X_i^{[a]}(t)$ . Also, for all  $t \geq a$  and  $i = 1, 2, 3$ , we can check that

$$B_i^{[a]}(t) - D_i^{[a]}(t) = G_i(t).$$

Lastly, we define  $G_0(t) = B_0^{[a]}(t) - D_0^{[a]}(t)$  for all  $t \geq a$ . It follows that

$$G_0(t) = -s\left(\tilde{X}_1(t) + \tilde{X}_2(t) + 2\tilde{X}_3(t)\right) - r\tilde{X}_3(t)\left(1 - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t)\right) - 2\mu. \tag{3.26}$$

We note that from (3.23),

$$G_{0r}^{(a,b)}(t) = G_0(t) + r\tilde{X}_1(t)\tilde{X}_2(t)1_{(a,b]}(t). \tag{3.27}$$

#### 4. Martingales and submartingales

In this section, we define martingales and submartingales that we use in the proofs of phase 1, 2, and 3. For  $i = 1, 2, 3$  and  $0 \leq a < b$ , we define  $Z_{im}^{(a,b]}(t) = 0$ , if  $0 \leq t < a$ , and we define

$$Z_{im}^{(a,b]}(t) = e^{-\int_a^t G_i(v)dv} X_{im}^{(a,b]}(t) - \int_a^t M_i^{(a,b]}(u) e^{-\int_a^u G_i(v)dv} du, \tag{4.1}$$

if  $0 \leq a < t$ . Also, for  $i = 0, 1, 2, 3$  and  $0 \leq a < b$ , we define  $Z_{ir}^{(a,b]}(t) = 0$ , if  $0 \leq t < a$ , and we define

$$Z_{ir}^{(a,b]}(t) = e^{-\int_a^t G_{ir}^{(a,b]}(v)dv} X_{ir}^{(a,b]}(t) - \int_a^t R_i^{(a,b]}(u) e^{-\int_a^u G_{ir}^{(a,b]}(v)dv} du, \tag{4.2}$$

$0 \leq a < t$ . One can check that for  $t \geq a$ ,

$$X_{im}^{(a,b]}(t) = \int_a^t M_i^{(a,b]}(u) e^{\int_u^t G_i(v)dv} du + Z_{im}^{(a,b]}(t) e^{\int_a^t G_i(v)dv}, \tag{4.3}$$

$$X_{ir}^{(a,b]}(t) = \int_a^t R_i^{(a,b]}(u) e^{\int_u^t G_{ir}^{(a,b]}(v)dv} du + Z_{ir}^{(a,b]}(t) e^{\int_a^t G_{ir}^{(a,b]}(v)dv}. \tag{4.4}$$

These expressions of  $X_{im}^{(a,b]}(t)$  and  $X_{ir}^{(a,b]}(t)$  are crucial to our proof. The first term on the right-hand side is the dominant term, while the second term represents a lower order term, which can be thought of as small fluctuation from the dominant term.

For the process  $((X_0(t), X_1(t), X_2(t), X_3(t)), t \geq 0)$ , we define  $(\mathcal{F}_t)_{t \geq 0}$  to be the natural filtration of this process.

**Proposition 4.1.** *For  $i = 1, 2, 3$ , the process  $(Z_{im}^{(a,b]}(t), t \geq a)$  is a mean-zero martingale, and for  $a \leq t$ ,*

$$\begin{aligned} &\text{Var}\left(Z_{im}^{(a,b]}(t)\right) \\ &= E\left[\int_a^t e^{-2\int_a^u G_i(v)dv} \left(M_i^{(a,b]}(u) + \left(B_{im}^{(a,b]}(u) + D_{im}^{(a,b]}(u)\right) X_{im}^{(a,b]}(u)\right) du\right]. \end{aligned}$$

Also, For  $i = 0, 1, 2, 3$  the process  $(Z_{ir}^{(a,b]}(t), t \geq a)$  is a mean-zero martingale, and for  $a \leq t$ ,

$$\begin{aligned} & \text{Var}\left(Z_{ir}^{(a,b]}(t)\right) \\ &= E\left[\int_a^t e^{-2\int_a^u G_{ir}^{(a,b]}(v)dv} \left(R_i^{(a,b]}(u) + \left(B_{ir}^{(a,b]}(u) + D_{ir}^{(a,b]}(u)\right)X_{ir}^{(a,b]}(u)\right)du\right]. \end{aligned}$$

Moreover, if  $T$  is a stopping time and  $T \geq a$ , then the process  $(Z_{im}^{(a,b]}(t \wedge T), t \geq a)$  is a mean-zero martingale for  $i = 1, 2, 3$ , and if  $a \leq t$ , then

$$\begin{aligned} & \text{Var}\left(Z_{im}^{(a,b]}(t \wedge T)\right) \\ &= E\left[\int_a^{t \wedge T} e^{-2\int_a^u G_i(v)dv} \left(M_i^{(a,b]}(u) + \left(B_{im}^{(a,b]}(u) + D_{im}^{(a,b]}(u)\right)X_{im}^{(a,b]}(u)\right)du\right]. \end{aligned}$$

Also, for  $i = 0, 1, 2, 3$ , the process  $(Z_{ir}^{(a,b]}(t \wedge T), t \geq a)$  is a mean-zero martingale, and if  $a \leq t$ , then

$$\begin{aligned} & \text{Var}\left(Z_{ir}^{(a,b]}(t \wedge T)\right) \\ &= E\left[\int_a^{t \wedge T} e^{-2\int_a^u G_{ir}^{(a,b]}(v)dv} \left(R_i^{(a,b]}(u) + \left(B_{ir}^{(a,b]}(u) + D_{ir}^{(a,b]}(u)\right)X_{ir}^{(a,b]}(u)\right)du\right]. \end{aligned}$$

*Proof:* The proof can be performed using the same technique showed in section 5.1 of [Schweinsberg \(2017\)](#).  $\square$

Since the process  $((X_0(t), X_1(t) - X_{1m}^{(a,b]}(t), X_{1m}^{(a,b]}(t), X_2(t), X_3(t)), t \geq 0)$  is a continuous-time Markov process, combining [Proposition 4.1](#) and Markov property yields the following result.

**Corollary 4.2.** *If  $T$  is a stopping time and  $T \geq a$ , then for  $i = 1, 2, 3$  and  $a \leq t$ ,*

$$\begin{aligned} & \text{Var}\left(Z_{im}^{(a,b]}(t \wedge T) \middle| \mathcal{F}_a\right) \\ &= E\left[\int_a^{t \wedge T} e^{-2\int_a^u G_i(v)dv} \left(M_i^{(a,b]}(u) + \left(B_{im}^{(a,b]}(u) + D_{im}^{(a,b]}(u)\right)X_{im}^{(a,b]}(u)\right)du \middle| \mathcal{F}_a\right]. \end{aligned}$$

Also, for  $i = 0, 1, 2, 3$  and  $a \leq t$ ,

$$\begin{aligned} & \text{Var}\left(Z_{ir}^{(a,b]}(t \wedge T) \middle| \mathcal{F}_a\right) \\ &= E\left[\int_a^{t \wedge T} e^{-2\int_a^u G_{ir}^{(a,b]}(v)dv} \left(R_i^{(a,b]}(u) + \left(B_{ir}^{(a,b]}(u) + D_{ir}^{(a,b]}(u)\right)X_{ir}^{(a,b]}(u)\right)du \middle| \mathcal{F}_a\right]. \end{aligned}$$

Now, for  $i = 0, 1, 2, 3$  and  $0 \leq a \leq t$ , we define

$$Z_i^{[a]}(t) = e^{-\int_a^t G_i(v)dv} X_i^{[a]}(t). \quad (4.5)$$

By similar arguments to the proofs of [Proposition 4.1](#) and [Corollary 4.2](#), we get the following result.

**Proposition 4.3.** *If  $T$  is a stopping time with  $T \geq a$ , then for  $i = 0, 1, 2, 3$ , the process  $(Z_i^{[a]}(t), t \geq a)$  is a martingale. Moreover, for all  $a \leq t$ ,*

$$\text{Var}\left(Z_i^{[a]}(t \wedge T) \middle| \mathcal{F}_a\right) = E\left[\int_a^{t \wedge T} e^{-2\int_a^u G_i(v)dv} \left(B_i^{[a]}(u) + D_i^{[a]}(u)\right)X_i^{[a]}(u)du \middle| \mathcal{F}_a\right].$$

Lastly, for  $i = 1, 2, 3$ ,  $0 \leq a < b$ , and  $a \leq t$ , we define

$$W_{im}^{(a,b]}(t) = e^{-\int_a^t G_i(v)dv} X_{im}^{(a,b]}(t), \tag{4.6}$$

and for  $i = 0, 1, 2, 3$ ,  $0 \leq a < b$ , and  $a \leq t$ , we define

$$W_{ir}^{(a,b]}(t) = e^{-\int_a^t G_{ir}^{(a,b]}(v)dv} X_{ir}^{(a,b]}(t).$$

By a standard argument, we can prove the following proposition.

**Proposition 4.4.** *If  $T$  is a stopping time and  $T \geq a$ , for  $i = 1, 2, 3$ , the process  $(W_{im}^{(a,b]}(t \wedge T), t \geq a)$  is a submartingale. Moreover, for  $a \leq t$ ,*

$$E\left[W_{im}^{(a,b]}(t \wedge T) \middle| \mathcal{F}_a\right] = E\left[\int_a^{t \wedge T} M_i^{(a,b]}(u) e^{-\int_a^u G_i(v)dv} du \middle| \mathcal{F}_a\right].$$

For  $i = 0, 1, 2, 3$  the process  $(W_{ir}^{(a,b]}(t \wedge T), t \geq a)$  is a submartingale, and for  $a \leq t$ ,

$$E\left[W_{ir}^{(a,b]}(t \wedge T) \middle| \mathcal{F}_a\right] = E\left[\int_a^{t \wedge T} R_i^{(a,b]}(u) e^{-\int_a^u G_{ir}^{(a,b]}(v)dv} du \middle| \mathcal{F}_a\right].$$

**5. Phase 1 and the proof of Proposition 2.2**

5.1. *Notations.* First, note that it is enough to show that Propositions 2.2, 2.3, 2.4, and 2.5 hold for all small values of  $\epsilon$  and  $\delta$ . We choose  $\epsilon$  and  $\delta$  as follows:

$$\epsilon \in \left(0, \frac{1}{16}\right), \tag{5.1}$$

and

$$\delta \in \left(0, \frac{1}{4}\right). \tag{5.2}$$

Next, in both the recombination dominating case and the mutation dominating case, we pick the following constants:

$$K > \frac{10}{\epsilon}, \tag{5.3}$$

$$C_1 > \ln\left(\frac{5K}{\epsilon}\right) \vee \ln\left(\frac{8}{\delta^2}\right), \tag{5.4}$$

$$C_{0,m} > 2 \ln\left(\frac{2K}{\epsilon}\right), \tag{5.5}$$

$$C_{0,m}^+ > C_{0,m} \vee \left(12e^{-C_1} + \ln\left(\frac{12K}{\epsilon(1-\delta^2)^2}\right)\right), \tag{5.6}$$

$$C_{0,r} > \ln\left(\frac{2K^2}{\epsilon}\right) \vee (C_1 + \ln 4), \tag{5.7}$$

$$\eta = 2Ke^{-C_1}. \tag{5.8}$$

These constants are chosen to make certain inequalities in the proof hold. We present them here for a record-keeping purpose. Also, since the proof involves many constants, we want to be clear on the order of picking these constants.

Now, we define several fixed times as follows:

$$t_{0,r} = \begin{cases} \frac{1}{s} \ln \left( \frac{s}{\mu\sqrt{Nr}} \right) - \frac{C_{0,r}}{s} & \text{in the recombination dominating case} \\ \frac{1}{s} \ln \left( \frac{s}{\mu\sqrt{Nr}} \right) - \frac{C_{0,r}}{s} & \text{in the mutation dominating case when } Nr \geq e \\ \frac{1}{s} \ln \left( \frac{s}{\mu} \right) - \frac{C_{0,r}}{s} & \text{in the mutation dominating case when } Nr < e, \end{cases} \tag{5.9}$$

and in both cases, we define

$$t_{0,m} = \frac{1}{s} \ln \left( \frac{s}{N\mu^2} \right) - \frac{C_{0,m}}{s}, \tag{5.10}$$

$$t_{0,m}^+ = \frac{1}{s} \ln \left( \frac{s}{N\mu^2} \right) + \frac{C_{0,m}^+}{s}, \tag{5.11}$$

$$t_1 = \frac{1}{s} \ln \left( \frac{s}{\mu} \right) - \frac{C_1}{s}. \tag{5.12}$$

From our parameter assumptions and Lemma 2.1, we have  $0 < t_{0,m} < t_{0,m}^+ < t_1$  and  $0 < t_{0,r} < t_1$ . The time  $t_1$  is a time that type 1 and type 2 become established in the population. The majority of type 3 individuals at time  $t_1$  descend from type 3 individuals that appear around time  $t_{0,r}$  in the recombination dominating case and around time  $t_{0,m}$  in the mutation dominating case.

Lastly, in both cases, we define the following stopping times:

$$T_1 = \inf \left\{ t \geq 0 : X_1(t) \geq \frac{KN\mu}{s} e^{st} \right\}, \tag{5.13}$$

$$T_2 = \inf \left\{ t \geq 0 : X_2(t) \geq \frac{KN\mu}{s} e^{st} \right\}, \tag{5.14}$$

$$T_3 = \inf \left\{ t \geq 0 : X_3(t) \geq \frac{N\mu}{s} e^{st} \right\}, \tag{5.15}$$

$$T_{(1)} = T_1 \wedge T_2 \wedge T_3. \tag{5.16}$$

As we mention before, we expect that

$$X_1(t) \approx X_2(t) \approx \frac{2N\mu}{s} e^{st}$$

during the time interval  $[0, t_1]$ . Also, the number of type 3 individuals during the first phase should be smaller in order when comparing with the number of type 1 individuals. Thus, the stopping time  $T_{(1)}$  should occur after time  $t_1$  with a probability close to 1.

5.2. *Upper bounds for expectations.* First of all, we will prove some inequalities regarding the growth rates of type 1, 2, and 3. They will be used quite often in this section. The essence of the following proposition is that the growth rates of type 1 and type 2 is approximately  $s$ , while the growth rate is about  $2s$  for type 3.

**Lemma 5.1.** *For sufficiently large  $N$  and  $t \in [0, t_1 \wedge T_{(1)})$ , the following statements hold:*

- (1)  $X_i(t) \leq \eta N$ , for  $i = 1, 2, 3$ .
- (2)  $G_1(t) \leq s$ ,  $G_2(t) \leq s$ , and  $G_3(t) \leq 2s$ .
- (3)  $G_1(t) \geq s - 4\eta s - r - \mu$ ,  $G_2(t) \geq s - 4\eta s - r - \mu$ , and  $G_3(t) \geq 2s - 4\eta s - r$ .

- (4) For  $0 < a < b$ , we have  $G_{1r}^{(a,b]}(t) \leq s + r1_{(a,b]}(t)$ ,  $G_{2r}^{(a,b]}(t) \leq s + r1_{(a,b]}(t)$ , and  $G_{3r}^{(a,b]}(t) \leq 2s + r1_{(a,b]}(t)$ .
- (5) For  $0 < a < b$ , we have  $G_{1r}^{(a,b]}(t) \geq s - 4\eta s - r - \mu$ ,  $G_{2r}^{(a,b]}(t) \geq s - 4\eta s - r - \mu$ , and  $G_{3r}^{(a,b]}(t) \geq 2s - 4\eta s - r$ .

*Proof:* By the definition of  $\eta$ ,  $t_1$  and  $T_{(1)}$  in (5.8), (5.12) and (5.16), for  $t \in [0, t_1 \wedge T_{(1)})$  and  $i = 1, 2, 3$ ,

$$X_i(t) < \frac{KN\mu}{s}e^{st} \leq \frac{KN\mu}{s}e^{st_1} = Ke^{-C_1}N < \eta N.$$

For the proof of the statement (2), since  $0 \leq \tilde{X}_1(t) + \tilde{X}_2(t) + \tilde{X}_3(t) \leq 1$  for all  $t \geq 0$ , and  $s \ll 1$ , it follows that  $0 < 1 - 2s \leq 1 - s\tilde{X}_1(t) - s\tilde{X}_2(t) - 2s\tilde{X}_3(t) \leq 1$  for all  $t \geq 0$ , when  $N$  sufficiently large. Thus, by the definition of  $G_1(t)$  in (3.18), for sufficiently large  $N$ , we have  $G_1(t) \leq s$  for all  $t \in [0, t_1 \wedge T_{(1)})$ . Also, by the statement (1), if  $t \in [0, t_1 \wedge T_{(1)})$ , then  $1 - \tilde{X}_1(t) - \tilde{X}_2(t) - 2\tilde{X}_3(t) \geq 1 - 4\eta$ . Again, by using the definition of  $G_1(t)$  in (3.18), we get the lower bound of  $G_1(t)$  in statement 3. Similar arguments can show both the upper and lower bounds for  $G_3(t)$ . Lastly, we can prove statements 4 and 5 by using (3.20), (3.21) and (3.22), along with statements 1, 2 and 3 of this lemma.  $\square$

Next, we will proceed to prove some results on the upper bounds for the expectations of  $X_{im}^{(a,b]}(t \wedge T_{(1)})$  and  $X_{ir}^{(a,b]}(t \wedge T_{(1)})$ . These bounds inform us the order of magnitude of the number of each type in the population. After we establish the bounds, We will apply them at the end of this subsection to show that  $T_{(1)} > t_1$  with high probability.

**Lemma 5.2.** *For sufficiently large  $N$ , for  $i = 1, 2$  and  $t \in [0, t_1]$ , we have*

$$E\left[e^{-s(t \wedge T_{(1)})} X_{im}(t \wedge T_{(1)})\right] \leq \frac{2N\mu}{s}, \tag{5.17}$$

and

$$E\left[X_{im}(t \wedge T_{(1)})\right] \leq \frac{2N\mu}{s}e^{st}.$$

*Proof:* The proof is similar to Lemma 5.1 in Schweinsberg (2017). We will prove only when  $i = 1$  since the argument is similar for  $i = 2$ . First, we show that for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and for  $t \in [0, t_1]$ ,

$$E\left[e^{-s(t \wedge T_{(1)})} X_{1m}^{(a,b]}(t \wedge T_{(1)})\right] \leq e^{(4\eta s + r + \mu)(b-a)} \cdot 2N\mu \int_a^b e^{-su} du. \tag{5.18}$$

If  $t \in [0, a)$ , this inequality is trivial since  $X_{1m}^{(a,b]}(t) = 0$  by the definition of  $X_{1m}^{(a,b]}(t)$ . Now, assume that  $t \in [a, t_1]$ . By Proposition 4.1 and the equation (4.1), we have  $E[Z_{1m}^{(a,b]}(t \wedge T_{(1)})] = 0$  and

$$E\left[e^{-\int_a^{t \wedge T_{(1)}} G_1(v)dv} X_{1m}^{(a,b]}(t \wedge T_{(1)})\right] = E\left[\int_a^{t \wedge T_{(1)}} M_1^{(a,b]}(u) e^{-\int_a^u G_1(v)dv} du\right]. \tag{5.19}$$

(Note that in the event  $T_{(1)} < a$ , we interpret the integral from  $a$  to  $t \wedge T_{(1)}$  as 0. Also, from the definition of  $X_{1m}^{(a,b]}(t)$ , in the event  $T_{(1)} < a$ , we have  $X_{1m}^{(a,b]}(t \wedge T_{(1)}) =$

0.) Now, using the upper bound for  $G_1(t)$  in Lemma 5.1, for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$  and  $t \in [a, t_1]$ ,

$$\begin{aligned} & E \left[ e^{-\int_a^{t \wedge T(1)} G_1(v) dv} X_{1m}^{(a,b]}(t \wedge T(1)) \right] \\ & \geq E \left[ e^{-\int_a^{t \wedge T(1)} s dv} X_{1m}^{(a,b]}(t \wedge T(1)) 1_{\{T(1) \geq a\}} \right] \\ & = e^{sa} E \left[ e^{-s(t \wedge T(1))} X_{1m}^{(a,b]}(t \wedge T(1)) \right]. \end{aligned} \tag{5.20}$$

Next, we use the lower bound for  $G_1(t)$  in Lemma 5.1. From (3.1), for sufficiently large  $N$ , when  $0 \leq a < b \leq t_1$  and  $t \in [a, t_1]$ ,

$$\begin{aligned} & E \left[ \int_a^{t \wedge T(1)} M_1^{(a,b]}(u) e^{-\int_a^u G_1(v) dv} du \right] \\ & = E \left[ \int_a^{t \wedge T(1)} \mu X_0(u) 1_{(a,b]}(u) e^{-\int_a^u G_1(v) dv} du \right] \\ & \leq \int_a^b 2N\mu e^{-(s-4\eta s-r-\mu)(u-a)} du \\ & \leq e^{(4\eta s+r+\mu)(b-a)} \cdot N\mu e^{sa} \int_a^b e^{-su} du. \end{aligned} \tag{5.21}$$

Hence, the inequality (5.18) follows from (5.19), (5.20), and (5.21).

Now, we will improve (5.18) by showing that the inequality holds even without the factor  $e^{(4\eta s+r+\mu)(b-a)}$ . For each  $n \in \mathbb{N}$ , let  $t'_j = (b-a)j/n + a$  for every  $j = 0, 1, \dots, n$ . It follows from (5.18) that

$$\begin{aligned} E \left[ e^{-s(t \wedge T(1))} X_{1m}^{(a,b]}(t \wedge T(1)) \right] & = E \left[ \sum_{j=0}^{n-1} e^{-s(t \wedge T(1))} X_{1m}^{(t'_j, t'_{j+1}]}(t \wedge T(1)) \right] \\ & \leq \sum_{j=0}^{n-1} e^{(4\eta s+r+\mu)(t'_{j+1}-t'_j)} \cdot 2N\mu \int_{t'_j}^{t'_{j+1}} e^{-su} du \\ & = e^{(4\eta s+r+\mu)(\frac{b-a}{n})} \cdot 2N\mu \int_a^b e^{-su} du \\ & \leq e^{(4\eta s+r+\mu)(\frac{b-a}{n})} \cdot \frac{N\mu}{s} e^{-sa}. \end{aligned}$$

By letting  $n \rightarrow \infty$ , we have that for sufficiently large  $N$ ,  $0 \leq a < b \leq t_1$ , and  $t \in [a, t_1]$ ,

$$E \left[ e^{-s(t \wedge T(1))} X_{1m}^{(a,b]}(t \wedge T(1)) \right] \leq \frac{2N\mu}{s} e^{-sa}. \tag{5.22}$$

The inequality (5.17) follows from the fact that  $X_{1m}(t \wedge T(1)) = X_{1m}^{(0,t]}(t \wedge T(1))$ . Also,

$$E \left[ X_{1m}(t \wedge T(1)) \right] \leq e^{st} E \left[ e^{-s(t \wedge T(1))} X_{1m}^{(0,t]}(t \wedge T(1)) \right] \leq \frac{2N\mu}{s} e^{st},$$

which completes the proof of this lemma. □

**Lemma 5.3.** For sufficiently large  $N$ , for  $i = 1, 2$ , and  $t \in [0, t_1]$ , we have

$$E \left[ e^{-s(t \wedge T(1))} X_{ir}(t \wedge T(1)) \right] \leq \left( \frac{2N\mu r}{s} \right) t,$$

and

$$E\left[X_{ir}(t \wedge T_{(1)})\right] \leq \left(\frac{2N\mu r}{s}\right)e^{st}.$$

*Proof:* Using the upper bound for  $G_{1r}^{(a,b]}(t)$  in Lemma 5.1, for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and  $t \in [a, t_1]$ ,

$$E\left[e^{-\int_a^{t \wedge T_{(1)}} G_{1r}^{(a,b]}(v)dv} X_{1r}^{(a,b]}(t \wedge T_{(1)})\right] \geq e^{sa-r(b-a)} E\left[e^{-s(t \wedge T_{(1)})} X_{1r}^{(a,b]}(t \wedge T_{(1)})\right].$$

Then, using the lower bound for  $G_{1r}^{(a,b]}(t)$  in Lemma 5.1, the upper bound for  $R_1^{(a,b]}(t)$  in (3.16), and the definition of  $T_3$  in (5.15), we have that for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and for  $t \in [a, t_1]$ ,

$$\begin{aligned} E\left[\int_a^{t \wedge T_{(1)}} R_1^{(a,b]}(u) e^{-\int_a^u G_{1r}^{(a,b]}(v)dv} du\right] &\leq \int_a^b 2Nr \cdot \frac{\mu}{s} e^{su} \cdot e^{-(s-4\eta s-r-\mu)(u-a)} du \\ &\leq e^{(4\eta s+r+\mu)(b-a)} \cdot \frac{2N\mu r}{s} e^{sa}(b-a). \end{aligned}$$

The rest of the proof follows as in Lemma 5.2. □

**Lemma 5.4.** For sufficiently large  $N$  and for  $t \in [0, t_1]$ , we have

$$\begin{aligned} E\left[e^{-s(t \wedge T_{(1)})} X_{3m}(t \wedge T_{(1)})\right] &\leq \frac{2KN\mu^2}{s^2} e^{st}, \\ E\left[X_{3m}^{(t_0, m, t_1]}(t \wedge T_{(1)})\right] &\leq \frac{2Ke^{C_{0,m}} N^2 \mu^4}{s^3} e^{2st}, \end{aligned}$$

and

$$E\left[X_{3m}^{(t_0^+, m, t_1]}(t \wedge T_{(1)})\right] \leq \frac{2Ke^{-C_{0,m}^+} N^2 \mu^4}{s^3} e^{2st},$$

*Proof:* Using the upper bound for  $G_3(t)$  in Lemma 5.1, for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and  $t \in [a, t_1]$ ,

$$E\left[e^{-\int_a^{t \wedge T_{(1)}} G_3(v)dv} X_{3m}^{(a,b]}(t \wedge T_{(1)})\right] \geq e^{-st+2sa} E\left[e^{-s(t \wedge T_{(1)})} X_{3m}^{(a,b]}(t \wedge T_{(1)})\right].$$

Now, we use the definition of  $M_3^{(a,b]}(t)$  in (3.11), the lower bound for  $G_3(t)$  in Lemma 5.1, and the definition of  $T_1$  and  $T_2$  in (5.13) and (5.14). It follows that for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and for  $t \in [a, t_1]$ ,

$$\begin{aligned} E\left[\int_a^{t \wedge T_{(1)}} M_3^{(a,b]}(u) e^{-\int_a^u G_3(v)dv} du\right] &\leq E\left[\int_a^{b \wedge T_{(1)}} \frac{2KN\mu^2}{s} e^{su} \cdot e^{-(2s-4\eta s-r)(u-a)} du\right] \\ &\leq e^{(4\eta s+r)(b-a)} \cdot \frac{2KN\mu^2}{s} e^{2sa} \int_a^b e^{-su} du. \end{aligned}$$

The rest of the proof follows the same argument in Lemma 5.2. □

**Lemma 5.5.** For sufficiently large  $N$  and  $0 \leq a < t_1$ , if  $t \in [0, t_1]$ , we have

$$E\left[e^{-s(t \wedge T_{(1)})} X_{3r}(t \wedge T_{(1)})\right] \leq \left(\frac{2K^2 N \mu^2 r}{s^2}\right) e^{st},$$

and if  $t \in [a, t_1]$ ,

$$E\left[X_{3r}^{(a,t_1]}(t \wedge T_{(1)})\right] \leq \left(\frac{2K^2 N \mu^2 r}{s^2}\right) e^{2st}(t - a).$$

*Proof:* Using the upper bound for  $G_{3r}^{(a,b]}(t)$  in Lemma 5.1, for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and for  $t \in [a, t_1]$ ,

$$E\left[e^{-\int_a^{t \wedge T_{(1)}} G_{3r}^{(a,b]}(v) dv} X_{3r}^{(a,b]}(t \wedge T_{(1)})\right] \geq e^{2sa - st - r(b-a)} E\left[e^{-s(t \wedge T_{(1)})} X_{3r}^{(a,b]}(t \wedge T_{(1)})\right].$$

Then, we use the lower bound for  $G_{3r}^{(a,b]}(t)$  in Lemma 5.1, the upper bound for  $R_3^{(a,b]}(t)$  in (3.16), and the definitions of  $T_1$  and  $T_2$  in (5.13) and (5.14), for sufficiently large  $N$ , for  $0 \leq a < b \leq t_1$ , and for  $t \in [a, t_1]$ ,

$$\begin{aligned} E\left[\int_a^{t \wedge T_{(1)}} R_3^{(a,b]}(u) e^{-\int_a^u G_{3r}^{(a,b]}(v) dv} du\right] &\leq \int_a^{t \wedge b} \frac{2K^2 N \mu^2 r}{s^2} e^{2su} \cdot e^{-(2s-4\eta s-r)(u-a)} du \\ &\leq e^{(4\eta s+r)(b-a)} \cdot \frac{2K^2 N \mu^2 r}{s^2} \cdot e^{2sa} \cdot \int_a^{t \wedge b} 1 du. \end{aligned}$$

The rest of the proof follows the argument in Lemma 5.2. □

Here, we define the event

$$A_1 = \{T_{(1)} > t_1\}, \tag{5.23}$$

and we are going to show that this event occurs with probability close to 1.

**Lemma 5.6.** *For sufficiently large  $N$ , we have  $P(A_1^c) \leq 2\epsilon$ .*

*Proof:* First of all, note that

$$P(A_1^c) = P(t_1 \wedge T_{(1)} = T_{(1)}) \leq \sum_{i=1}^3 P(t_1 \wedge T_{(1)} = T_i).$$

Now, consider the term  $P(t_1 \wedge T_{(1)} = T_i)$ , for  $i = 1, 2$ . By Markov's inequality,

$$\begin{aligned} P(t_1 \wedge T_{(1)} = T_i) &\leq P\left(X_i(t_1 \wedge T_{(1)}) \geq \frac{KN\mu}{s} e^{s(t_1 \wedge T_{(1)})}\right) \\ &\leq P\left(X_{im}(t_1 \wedge T_{(1)}) \geq \frac{KN\mu}{2s} e^{s(t_1 \wedge T_{(1)})}\right) \\ &\quad + P\left(X_{ir}(t_1 \wedge T_{(1)}) \geq \frac{KN\mu}{2s} e^{s(t_1 \wedge T_{(1)})}\right) \\ &\leq \frac{E[e^{-s(t_1 \wedge T_{(1)})} X_{im}(t_1 \wedge T_{(1)})]}{KN\mu/2s} + \frac{E[e^{-s(t_1 \wedge T_{(1)})} X_{ir}(t_1 \wedge T_{(1)})]}{KN\mu/2s}. \end{aligned}$$

Using Lemmas 5.2 and 5.3, the definition of  $t_1$  in (5.12), and (2.6), for sufficiently large  $N$ ,

$$P(t_1 \wedge T_{(1)} = T_i) \leq \frac{4}{K} + \frac{4rt_1}{K} \leq \frac{4}{K} + \frac{4}{K} \cdot \frac{r}{s} \ln\left(\frac{s}{\mu}\right) \leq 5K^{-1}.$$

By a similar argument, using Lemma 5.4 and 5.5, for sufficiently large  $N$ , we have

$$P(t_1 \wedge T_{(1)} = T_3) \leq 5Ke^{-C_1}.$$

Thus, from the way we choose  $K$  and  $C_1$  in (5.3) and (5.4), we have  $P(T_{(1)} \leq t_1) \leq 10K^{-1} + 5Ke^{-C_1} \leq 2\epsilon$ .  $\square$

5.3. *The variance bounds.* In this section, we will apply the expectation bounds from the previous sections to obtain variance bounds for our martingales, which were defined in section 4. Then, we use Doob’s maximal inequality to get bounds for martingales; this gives us control over the fluctuation term for the number of each type.

We define the following events:

$$A_2 = \left\{ \sup_{t \in [0, t_1]} \left| Z_{1m}^{(0, t_1]}(t \wedge T_{(1)}) \right| \leq \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu}{s^2}} \right\}. \tag{5.24}$$

$$A_3 = \left\{ \sup_{t \in [0, t_1]} \left| Z_{2m}^{(0, t_1]}(t \wedge T_{(1)}) \right| \leq \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu}{s^2}} \right\}.$$

$$A_4 = \left\{ \sup_{t \in [0, t_1]} \left| Z_{1r}^{(0, t_1]}(t \wedge T_{(1)}) \right| \leq \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu r}{s^3} \ln\left(\frac{s}{\mu}\right)} \right\}. \tag{5.25}$$

$$A_5 = \left\{ \sup_{t \in [0, t_1]} \left| Z_{2r}^{(0, t_1]}(t \wedge T_{(1)}) \right| \leq \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu r}{s^3} \ln\left(\frac{s}{\mu}\right)} \right\}.$$

$$A_6 = \left\{ \sup_{t \in [t_{0,m}^+, t_1]} \left| Z_{3m}^{(t_{0,m}^+, t_1]}(t \wedge T_{(1)}) \right| \leq \sqrt{\frac{12Ke^{C_{0,m}^+}}{\epsilon} \cdot \frac{1}{s^2}} \right\}. \tag{5.26}$$

$$A_7 = \left\{ \sup_{t \in [t_{0,r}, t_1]} \left| Z_{3r}^{(t_{0,r}, t_1]}(t \wedge T_{(1)}) \right| \leq \sqrt{\frac{8K^2e^{-2C_{0,r}}}{\epsilon} \cdot \frac{\ln_+(Nr)}{s^2}} \right\}. \tag{5.27}$$

For the event  $A_6$ , we focus on the martingale of type 3m starting from time  $t_{0,m}^+$ . In the mutation dominating case, the main contribution to the number of type 3 individuals at time  $t_1$  comes from type 3m ancestors that appear after time  $t_{0,m}$ . The reason that we use  $t_{0,m}^+$  instead of  $t_{0,m}$  is for technical purpose in the proof. For the event  $A_7$ , we focus on the martingale of type 3r starting from time  $t_{0,r}$  for a similar reason. In fact, we need to consider the event  $A_7$  only in the recombination dominating case.

**Lemma 5.7.** *The following statements hold:*

- (1) For sufficiently large  $N$  and  $i = 2, 3, 4, 5, 6$ , we have  $P(A_i^c) \leq \epsilon$ .
- (2) In the recombination dominating case, for sufficiently large  $N$ , we have  $P(A_7^c) \leq \epsilon$ .

*Proof:* As mention above, the main strategy is to establish variance bounds, and the results follow from the Doob’s maximal inequality. We will first show that  $P(A_2^c) \leq \epsilon$  when  $N$  is sufficiently large. From (3.4) and (3.6), using the facts that  $\mu \ll s \ll 1$  and  $r \ll s$ , for sufficiently large  $N$  and  $t \geq 0$ , we have  $B_{1m}^{(0, t_1]}(t) \leq 1$

and  $D_{1m}^{(0,t_1]}(t) \leq 1$ . By Proposition 4.1 and Lemma 5.1, for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{1m}^{(0,t_1]}(t_1 \wedge T_{(1)})\right) \\ &= E\left[\int_0^{t_1 \wedge T_{(1)}} e^{-2\int_0^u G_1(v)dv} \left(\mu X_0(u) + (B_{1m}^{(0,t_1]}(u) + D_{1m}^{(0,t_1]}(u))X_{1m}^{(0,t_1]}(u)\right) du\right] \\ &\leq E\left[\int_0^{t_1} e^{-2(s-4\eta s-r-\mu)u} (2N\mu + 2X_{1m}(u \wedge T_{(1)})) du\right] \\ &\leq e^{2(r+\mu)t_1} \int_0^{t_1} e^{-2s(1-4\eta)u} (2N\mu + 2E[X_{1m}(u \wedge T_{(1)})]) du \end{aligned}$$

It follows from Lemma 5.2 that

$$\begin{aligned} \text{Var}\left(Z_{1m}^{(0,t_1]}(t_1 \wedge T_{(1)})\right) &\leq e^{2(r+\mu)t_1} \int_0^{t_1} e^{-2s(1-4\eta)u} \left(2N\mu + \frac{4N\mu}{s} e^{su}\right) du \\ &\leq e^{2(r+\mu)t_1} \cdot \frac{2N\mu}{s} \cdot \left(\frac{1}{2(1-4\eta)} + \frac{2}{s(1-8\eta)}\right). \end{aligned} \tag{5.28}$$

From the definition of  $t_1$  in (5.12), along with (2.6), and the facts that  $\mu \ll s$  and  $r \ll s$ , we have that

$$(r + \mu)t_1 = \frac{r}{s} \ln\left(\frac{s}{\mu}\right) + \frac{\mu}{s} \ln\left(\frac{s}{\mu}\right) - \frac{C_1(r + \mu)}{s} \ll 1. \tag{5.29}$$

By the way we choose  $\epsilon, C_1$  and  $\eta$  in (5.1), (5.4) and (5.8),

$$\eta = 2Ke^{-C_1} < 2\epsilon/5 < \epsilon \leq 1/16. \tag{5.30}$$

By (5.28), (5.29) and (5.30), for sufficiently large  $N$ ,

$$\text{Var}\left(Z_{1m}^{(0,t_1]}(t_1 \wedge T_{(1)})\right) \leq 2\left(\frac{2N\mu}{s}\right)\left(\frac{3}{s(1-8\eta)}\right) = \left(\frac{12}{1-8\eta}\right)\left(\frac{N\mu}{s^2}\right) \leq \frac{24N\mu}{s^2}.$$

Hence,  $P(A_2^c) \leq \epsilon$  follows from Doob’s maximal inequality. By the same argument,  $P(A_3^c) \leq \epsilon$ .

Next, we prove that  $P(A_4^c) \leq \epsilon$ . From Lemma 5.1 and inequality (3.16), for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{1r}^{(0,t_1]}(t_1 \wedge T_{(1)})\right) \\ &\leq E\left[\int_0^{t_1} e^{-2(s-4\eta s-r-\mu)u} (2Nr\tilde{X}_3(u \wedge T_{(1)}) + 2X_{1r}(u \wedge T_{(1)})) du\right] \\ &\leq e^{2(r+\mu)t_1} \int_0^{t_1} e^{-2s(1-4\eta)u} \left(2Nr \cdot \frac{\mu}{s} e^{su} + 2E[X_{1r}(u \wedge T_{(1)})]\right) du. \end{aligned}$$

From Lemma 5.3, for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{1r}^{(0,t_1]}(t_1 \wedge T_{(1)})\right) \\ &\leq e^{2(r+\mu)t_1} \int_0^{t_1} e^{-2s(1-4\eta)u} \left(\frac{2N\mu r}{s} e^{su} + \frac{4N\mu r}{s^2} \ln\left(\frac{s}{\mu}\right) e^{su}\right) du \\ &\leq e^{2(r+\mu)t_1} \frac{2N\mu r}{s^2(1-8\eta)} \left(1 + \frac{2}{s} \ln\left(\frac{s}{\mu}\right)\right) \\ &\leq \frac{24N\mu r}{s^3} \ln\left(\frac{s}{\mu}\right). \end{aligned}$$

By the same argument,  $P(A_5^c) \leq \epsilon$ .

Now, we give a proof that  $P(A_6^c) \leq \epsilon$ . Using Proposition 4.1, Lemma 5.1, and the definitions of  $T_1$  and  $T_2$  in (5.13) and (5.14), for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{3m}^{(t_0^+, t_1^+)}(t_1 \wedge T_{(1)})\right) \\ & \leq E \left[ \int_{t_0^+}^{t_1 \wedge T_{(1)}} e^{-2 \int_{t_0^+}^u (2s-4\eta s) dv} \left( \mu \cdot \frac{2KN\mu}{s} e^{su} + 2X_{3m}^{(t_0^+, t_1^+)}(u \wedge T_{(1)}) \right) du \right] \\ & \leq \int_{t_0^+}^{t_1} e^{-2(2s-4\eta s)(u-t_0^+)} \left( \mu \cdot \frac{2KN\mu}{s} e^{su} + 2E[X_{3m}^{(t_0^+, t_1^+)}(u \wedge T_{(1)})] \right) du. \end{aligned}$$

By Lemma 5.4 and the definition of  $t_{0,m}^+$  in (5.11), for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{3m}^{(t_0^+, t_1^+)}(t_1 \wedge T_{(1)})\right) \\ & \leq e^{2rt_1} \cdot e^{2(2s-4\eta s)t_{0,m}^+} \int_{t_{0,m}^+}^{t_1} e^{-2(2s-4\eta s)u} \left( \frac{2KN\mu^2}{s} e^{su} + \frac{4Ke^{-C_{0,m}^+} N^2 \mu^4}{s^3} e^{2su} \right) du \\ & \leq e^{2rt_1} \cdot e^{2(2s-4\eta s)t_{0,m}^+} \cdot \frac{2KN\mu^2}{s} \left( \frac{e^{-(3s-8\eta s)t_{0,m}^+}}{s(3-8\eta)} + \frac{2e^{-C_{0,m}^+} N\mu^2}{s^2} \cdot \frac{e^{-(2s-8\eta s)t_{0,m}^+}}{s(2-8\eta)} \right) \\ & = e^{2rt_1} \cdot \frac{2K}{s^2} \left( \frac{e^{C_{0,m}^+} s}{3-8\eta} + \frac{2e^{C_{0,m}^+}}{2-8\eta} \right) \\ & \leq \frac{12Ke^{C_{0,m}^+}}{s^2}. \end{aligned}$$

Lastly, we prove statement (2) of this lemma. From Proposition 4.1, Lemma 5.1, inequality (3.17), and the definition of  $T_1$  and  $T_2$  in (5.13) and (5.14), for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{3r}^{(t_0, r, t_1)}(t_1 \wedge T_{(1)})\right) \\ & \leq E \left[ \int_{t_0, r}^{t_1 \wedge T_{(1)}} e^{-2(2s-4\eta s-r)(u-t_0, r)} \left( 2Nr\tilde{X}_1(u)\tilde{X}_2(u) + 2X_{3r}^{(t_0, r, t_1)}(u \wedge T_{(1)}) \right) du \right] \\ & \leq \int_{t_0, r}^{t_1} e^{-2(2s-4\eta s-r)(u-t_0, r)} \left( \frac{2K^2 N\mu^2 r}{s^2} e^{2su} + 2E[X_{3r}^{(t_0, r, t_1)}(u \wedge T_{(1)})] \right) du. \end{aligned}$$

By Lemma 5.5, for sufficiently large  $N$ ,

$$\begin{aligned} & \text{Var}\left(Z_{3r}^{(t_0, r, t_1)}(t_1 \wedge T_{(1)})\right) \\ & \leq e^{2rt_1} \cdot e^{2(2s-4\eta s)t_{0, r}} \cdot \int_{t_{0, r}}^{t_1} e^{-2(2s-4\eta s)u} \left( \frac{2K^2 N\mu^2 r}{s^2} e^{2su} + \frac{4K^2 N\mu^2 r}{s^2} e^{2su}(u-t_{0, r}) \right) du \\ & \leq e^{2rt_1} \cdot e^{2(2s-4\eta s)t_{0, r}} \cdot \frac{2K^2 N\mu^2 r}{s^2} \left( 1 + 2(t_1 - t_{0, r}) \right) \cdot \frac{e^{-(2s-8\eta s)t_{0, r}}}{s(2-8\eta)}. \end{aligned}$$

From the definitions of  $t_{0,r}$  and  $t_1$  in (5.9) and (5.12),

$$\begin{aligned} \text{Var}\left(Z_{3r}^{(t_{0,r},t_1]}(t_1 \wedge T_{(1)})\right) &\leq e^{2rt_1} \cdot \frac{2K^2 e^{-2C_{0,r}}}{s(2-8\eta)} \left(1 + \frac{2}{s} \ln(Nr) + \frac{2(C_{0,r} - C_1)}{s}\right) \\ &\leq \frac{8K^2 e^{-2C_{0,r}} \ln(Nr)}{s^2}. \end{aligned}$$

Then, the result follows from the Doob’s maximal inequality. □

5.4. *Results on type 3 individuals.* In this subsection, we prove several results on type 3 individuals. First, we define the following events:

$$A_8 = \left\{ X_{3m}^{(0,t_{0,m}]}(t_1 \wedge T_{(1)}) = 0 \right\}. \tag{5.31}$$

$$A_9 = \left\{ X_{3r}^{(0,t_{0,r}]}(t_1 \wedge T_{(1)}) = 0 \right\}. \tag{5.32}$$

$$A_{10} = \left\{ X_{3m}^{(t_{0,m},t_1]}(t_1 \wedge T_{(1)}) \leq \left( \frac{2K e^{-2C_1 + C_{0,m}}}{\epsilon} \right) \frac{N^2 \mu^2}{s} \right\}. \tag{5.33}$$

$$A_{11} = \left\{ X_{3r}^{(t_{0,r},t_1]}(t_1 \wedge T_{(1)}) \leq \left( \frac{K^2 e^{-2C_1} (2(C_{0,r} - C_1) + 1)}{\epsilon} \right) \frac{(1 \vee Nr \ln_+(Nr))}{s} \right\}. \tag{5.34}$$

We will show that the events  $A_8$  and  $A_9$  occur with high probability; with probability close to 1, there are no type 3m (or 3r) individuals at time  $t_1$  who descend from type 3m (or 3r) ancestors that appear before time  $t_{0,m}$  (or  $t_{0,r}$ ). The proof consists of two main ideas.

- (1) With probability close to 1, the number of type 3m (or 3r) ancestors that appear before time  $t_{0,m}$  (or  $t_{0,r}$ ) is small. This part is just an application of Markov’s inequality and will be shown in Lemma 5.8.
- (2) With probability close to 1, each of these early ancestors will not have any alive descendant by time  $t_1$ . For this part, we will need some couplings with birth-death processes. This will be proved in Lemma 5.9 and 5.10.

At the end of this subsection, we will show that the events  $A_{10}$  and  $A_{11}$  also occur with high probability. These two events give bounds on the orders of the numbers of type 3m and type 3r individuals at time  $t_1$ .

**Lemma 5.8.** *We define  $m(t)$  and  $\rho(t)$  to be the number of type 3m ancestors and 3r ancestors, respectively, that appear in the time interval  $(0, t]$ . For sufficiently large  $N$ , the following statements hold.*

- (1)  $P\left(m(t_{0,m} \wedge T_{(1)}) \geq \frac{e^{-C_{0,m}/2}}{s}\right) \leq \epsilon.$
- (2)  $P\left(\rho(t_{0,r} \wedge T_{(1)}) \geq \frac{e^{-C_{0,r}+1}}{s}\right) \leq \epsilon.$

*Proof:* The process  $(m(t), t \geq 0)$  is a pure birth process with total birth rate  $M_3^{(0,t]}(t)$  as defined in (3.11). Then, there is a mean-zero martingale  $(W'(t), t \geq 0)$  such that for all  $t \geq 0$ ,

$$m(t) = W'(t) + \int_0^t M_3^{(0,u]}(u) du.$$

By Doob's stopping theorem,  $(W'(t \wedge T_{(1)}), t \geq 0)$  is a mean-zero martingale. Using the definitions of  $T_1$  and  $T_2$  in (5.13) and (5.14),

$$\begin{aligned} E[m(t_{0,m} \wedge T_{(1)})] &= E\left[\int_0^{t_{0,m} \wedge T_{(1)}} \mu(X_1(u) + X_2(u))du\right] \\ &\leq \int_0^{t_{0,m}} \frac{2KN\mu^2}{s} e^{su} du \\ &\leq \frac{2Ke^{-C_{0,m}}}{s}. \end{aligned}$$

So, by Markov's inequality and by the definition of  $C_{0,m}$  in (5.5),

$$P\left(m(t_{0,m} \wedge T_{(1)}) \geq \frac{e^{-C_{0,m}/2}}{s}\right) \leq \frac{E[m(t_{0,m} \wedge T_{(1)})]}{e^{-C_{0,m}/2}/s} \leq 2Ke^{-C_{0,m}/2} \leq \epsilon.$$

Now, consider the process  $(\rho(t), t \geq 0)$ . By similar argument, we get

$$\begin{aligned} E[\rho(t_{0,r} \wedge T_{(1)})] &= E\left[\int_0^{t_{0,r} \wedge T_{(1)}} R_3^{(0,u)}(u)du\right] \\ &\leq E\left[\int_0^{t_{0,r} \wedge T_{(1)}} 2Nr\tilde{X}_1(u)\tilde{X}_2(u)du\right] \\ &\leq \frac{2K^2N\mu^2r}{s^3} \cdot e^{2st_{0,r}}. \end{aligned}$$

From the definition of  $t_{0,r}$  in (5.9), in the recombination dominating case, and in the mutation dominating case with  $Nr \geq e$ ,

$$\frac{2K^2N\mu^2r}{s^3} \cdot e^{2st_{0,r}} = \frac{2K^2e^{-2C_{0,r}}}{s},$$

while in the mutation dominating case with  $Nr < e$ , we have

$$\frac{2K^2N\mu^2r}{s^3} \cdot e^{2st_{0,r}} = \frac{2K^2e^{-2C_{0,r}}Nr}{s} \leq \frac{2K^2e^{-2C_{0,r}+1}}{s}.$$

Hence,

$$E[\rho(t_{0,r} \wedge T_{(1)})] \leq \frac{2K^2e^{-2C_{0,r}+1}}{s}.$$

Thus, by Markov's inequality and the definition of  $C_{0,r}$  in (5.7),

$$P\left(\rho(t_{0,r} \wedge T_{(1)}) \geq \frac{e^{-C_{0,r}+1}}{s}\right) \leq \frac{E[\rho(t_{0,r} \wedge T_{(1)})]}{e^{-C_{0,r}+1}/s} \leq 2K^2e^{-C_{0,r}} \leq \epsilon,$$

which completes the proof. □

**Lemma 5.9.** For  $i \in \mathbb{N}$ , we define  $\tau_{i,m}$  to be the time that the  $i^{\text{th}}$  type  $3m$  ancestor appears, and we set  $\tau_{i,m} = \infty$  if the  $i^{\text{th}}$  type  $3m$  ancestor never appears. Let  $Y_{i,m}(t)$  be the number of descendants of the  $i^{\text{th}}$  type  $3m$  ancestor alive at time  $t$ . Then, for sufficiently large  $N$ ,

$$P\left(\{Y_{i,m}(t_1) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\right) \leq 3s,$$

for all  $i \in \mathbb{N}$ .

*Proof:* First, we define  $\tilde{Y}_{i,m}(t) = Y_{i,m}(t)/(2N)$  for all  $t \geq 0$  and  $i \in \mathbb{N}$ . By following the same reason that led us to the rates in (3.9) and (3.10), we have that on the event  $\tau_{i,m} \leq t_{0,m} \wedge T_{(1)}$ , the process  $(Y_{i,m}(t + \tau_{i,m}), t \geq 0)$  is a birth-death process with  $Y_{i,m}(\tau_{i,m}) = 1$  where each individual gives birth at rate

$$b(t) = [\tilde{X}_0(t + \tau_{i,m}) + (1 - s)(\tilde{X}_1(t + \tau_{i,m}) + \tilde{X}_2(t + \tau_{i,m})) \\ + (1 - 2s)(\tilde{X}_3(t + \tau_{i,m}) - \tilde{Y}_{i,m}(t + \tau_{i,m}))](1 - r\tilde{X}_0(t + \tau_{i,m})),$$

and dies at rate

$$d(t) = (1 - 2s)(1 - \tilde{Y}_{i,m}(t + \tau_{i,m}) + r\tilde{X}_0(t + \tau_{i,m})\tilde{Y}_{i,m}(t + \tau_{i,m})).$$

Note that for  $t \geq 0$ ,

$$b(t) \leq \tilde{X}_0(t + \tau_{i,m}) + \tilde{X}_1(t + \tau_{i,m}) + \tilde{X}_2(t + \tau_{i,m}) + (\tilde{X}_3(t + \tau_{i,m}) - \tilde{Y}_{i,m}(t + \tau_{i,m})) \\ = 1 - \tilde{Y}_{i,m}(t + \tau_{i,m}),$$

and

$$d(t) \geq (1 - 2s)(1 - \tilde{Y}_{i,m}(t + \tau_{i,m})).$$

Next, we apply a time change on the process  $(Y_{i,m}(t + \tau_{i,m}), t \geq 0)$  to obtain a new birth-death process with constant bounds for birth and death rates. For  $t \geq 0$ , define  $\lambda(t) = \int_{\tau_{i,m}}^{t+\tau_{i,m}} (1 - \tilde{Y}_{i,m}(v))dv$ . We also define  $Y_{i,m}^*(t) = Y_{i,m}(\lambda^{-1}(t) + \tau_{i,m})$  for  $t \in [0, \lambda((t_1 \wedge T_{(1)}) - \tau_{i,m})]$ . The process  $(Y_{i,m}^*(t), 0 \leq t < \lambda((t_1 \wedge T_{(1)}) - \tau_{i,m}))$  is a birth-death process with  $Y_{i,m}^*(0) = 1$  where each individual gives birth at rate

$$b^*(t) = b(\lambda^{-1}(t)) \cdot (\lambda^{-1})'(t) = \frac{b(\lambda^{-1}(t))}{1 - \tilde{Y}_{i,m}(\lambda^{-1}(t) + \tau_{i,m})} \leq 1,$$

and dies at rate

$$d^*(t) = d(\lambda^{-1}(t)) \cdot (\lambda^{-1})'(t) = \frac{d(\lambda^{-1}(t))}{1 - \tilde{Y}_{i,m}(\lambda^{-1}(t) + \tau_{i,m})} \geq 1 - 2s.$$

Let  $(Y^\#(t), t \geq 0)$  be a birth-death process with  $Y^\#(0) = 1$  where each individual gives birth at rate 1 and dies at rate  $1 - 2s$ . It is possible to couple the process  $(Y^\#(t), t \geq 0)$  with the population process, such that

- (1) on the event  $t_1 < T_{(1)}$ , for any time  $t$ , if  $Y_{i,m}^*(t) > 0$ , then  $Y^\#(t) > 0$ , and
- (2) the process  $(Y^\#(t), t \geq 0)$  is independent of  $\mathcal{F}_{\tau_{i,m}}$ .

The coupling can be done by generating birth and death events for the process with a larger birth rate and lower death rate, and then we thin them up properly to obtain birth and death events for the process with a lower birth rate but a larger death rate. It follows from this coupling that

$$P\left(\{Y_{i,m}(t_1) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\right) \\ \leq P\left(\{Y_{i,m}(t_1 - t_{0,m} + \tau_{i,m}) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\right) \\ = P\left(\{Y_{i,m}^*(\lambda(t_1 - t_{0,m})) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\right) \\ \leq P(Y^\#(\lambda(t_1 - t_{0,m})) > 0 \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}).$$

By Lemma 5.1 and (5.30), on the event  $t_1 < T_{(1)}$ , we have  $Y_{i,m}(t) \leq X_3(t) \leq \eta N \leq \frac{N}{2}$  for all  $t \in [0, t_1]$ . Hence,

$$\lambda(t_1 - t_{0,m}) = \int_{\tau_{i,m}}^{t_1 - t_{0,m} + \tau_{i,m}} 1 - \tilde{Y}_{i,m}(v) dv \geq \int_{\tau_{i,m}}^{t_1 - t_{0,m} + \tau_{i,m}} \frac{1}{2} dv \geq \frac{t_1 - t_{0,m}}{2}. \tag{5.35}$$

Therefore,

$$P\left(\{Y_{i,m}(t_1) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\right) \leq P\left(Y^\# \left(\frac{t_1 - t_{0,m}}{2}\right) > 0\right). \tag{5.36}$$

Lastly, from the generating function of birth and death process (in section 5 of Chapter III of Athreya and Ney, 1972), for  $t \geq 0$ ,

$$P(Y^\#(t) > 0) = \frac{1 - (1 - 2s)}{1 - (1 - 2s)e^{-(1-(1-2s))t}} \leq \frac{2s}{1 - e^{-2st}}. \tag{5.37}$$

Since  $1 \ll N\mu$ , for sufficiently large  $N$ ,

$$P\left(Y^\# \left(\frac{t_1 - t_{0,m}}{2}\right) > 0\right) \leq \frac{2s}{1 - e^{-s(t_1 - t_{0,m})}} = \frac{2s}{1 - \frac{1}{N\mu} e^{C_1 - C_{0,m}}} \leq 3s. \tag{5.38}$$

The proof is completed by (5.36) and (5.38). □

**Lemma 5.10.** For  $i \in \mathbb{N}$ , we define  $\tau_{i,r}$  to be the time that the  $i^{\text{th}}$  type 3r ancestor appears, and we set  $\tau_{i,r} = \infty$  if the  $i^{\text{th}}$  type 3r ancestor never appears. Let  $Y_{i,r}(t)$  be the number of descendants of the  $i^{\text{th}}$  type 3r ancestor alive at time  $t$ . Then, for sufficiently large  $N$ ,

$$P\left(\{Y_{i,r}(t_1) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,r} \leq t_{0,r} \wedge T_{(1)}\right) \leq 4s,$$

for all  $i \in \mathbb{N}$ .

*Proof:* The proof is similar to that of Lemma 5.9. First, define  $\tilde{Y}_{i,r}(t) = Y_{i,r}(t)/N$  for all  $t \geq 0$  and  $i \in \mathbb{N}$ . We have that on the event  $\tau_{i,r} \leq t_{0,r} \wedge T_{(1)}$ , the process  $(Y_{i,r}(t + \tau_{i,r}), t \geq 0)$  is a birth-death process with  $Y_{i,r}(\tau_{i,r}) = 1$  where each individual gives birth at rate

$$b(t) = [\tilde{X}_0(t + \tau_{i,r}) + (1 - s)(\tilde{X}_1(t + \tau_{i,r}) + \tilde{X}_2(t + \tau_{i,r})) + (1 - 2s)(\tilde{X}_3(t + \tau_{i,r}) - \tilde{Y}_{i,r}(t + \tau_{i,r}))](1 - r\tilde{X}_0(t + \tau_{i,r})),$$

and dies at rate

$$d(t) = (1 - 2s)(1 - \tilde{Y}_{i,r}(t + \tau_{i,r}) + r\tilde{X}_0(t + \tau_{i,r})\tilde{Y}_{i,r}(t + \tau_{i,r}) - r\tilde{X}_1(t + \tau_{i,r})\tilde{X}_2(t + \tau_{i,r})).$$

Note that when  $t \geq 0$ , we have  $b(t) \leq 1 - \tilde{Y}_{i,r}(t + \tau_{i,r})$ .

Next, we define a time change; for  $t \geq 0$ , let  $\lambda(t) = \int_{\tau_{i,r}}^{t + \tau_{i,r}} (1 - \tilde{Y}_{i,r}(v)) dv$ . Then, we define  $Y_{i,r}^*(t) = Y_{i,r}(\lambda^{-1}(t) + \tau_{i,r})$  for  $t \in [0, \lambda((t_1 \wedge T_{(1)}) - \tau_{i,r})]$ . The process  $(Y_{i,r}^*(t), 0 \leq t < \lambda((t_1 \wedge T_{(1)}) - \tau_{i,r}))$  is a birth-death process with  $Y_{i,r}^*(0) = 1$  where each individual gives birth at rate

$$b^*(t) = \frac{b(\lambda^{-1}(t))}{1 - \tilde{Y}_{i,r}(\lambda^{-1}(t) + \tau_{i,r})} \leq 1,$$

and dies at rate

$$d^*(t) = \frac{d(\lambda^{-1}(t))}{1 - \tilde{Y}_{i,r}(\lambda^{-1}(t) + \tau_{i,r})} \geq (1 - 2s) \left( 1 - \frac{r\tilde{X}_1(\lambda^{-1}(t) + \tau_{i,r})\tilde{X}_2(\lambda^{-1}(t) + \tau_{i,r})}{1 - \tilde{Y}_{i,r}(\lambda^{-1}(t) + \tau_{i,r})} \right).$$

Since the function  $\lambda$  is strictly increasing on the interval  $[0, (t_1 \wedge T_{(1)}) - \tau_{i,r})$ , we have that if  $t \in [0, \lambda((t_1 \wedge T_{(1)}) - \tau_{i,r}))$ , then  $\lambda^{-1}(t) + \tau_{i,r}(t) \leq t_1 \wedge T_{(1)}$ . Hence, from Lemma 5.1, for every  $t \in [0, \lambda((t_1 \wedge T_{(1)}) - \tau_{i,r}))$  and  $j = 1, 2$  and  $3$ , we have  $\tilde{X}_j(\lambda^{-1}(t) + \tau_{i,r}) \leq \eta$  and  $\tilde{Y}_{i,r}(\lambda^{-1}(t) + \tau_{i,r}) \leq \tilde{X}_3(\lambda^{-1}(t) + \tau_{i,r}) \leq \eta$ . Now, because  $r \ll s$ , for sufficiently large  $N$ , for  $t \in [0, \lambda((t_1 \wedge T_{(1)}) - \tau_{i,r}))$ ,

$$d^*(t) \geq (1 - 2s) \left( 1 - \left( \frac{\eta^2}{1 - \eta} \right) r \right) \geq (1 - 2s)(1 - s) > 1 - 3s.$$

Let  $(Y^\#(t), t \geq 0)$  be a birth-death process with  $Y^\#(0) = 1$  where each individual gives birth at rate 1 and dies at rate  $1 - 3s$ . Again, it is possible to couple the process  $(Y^\#(t), t \geq 0)$  with the population process, such that

- (1) on the event  $t_1 < T_{(1)}$ , for any time  $t$ , if  $Y_{i,m}^*(t) > 0$ , then  $Y^\#(t) > 0$ , and
- (2) the process  $(Y^\#(t), t \geq 0)$  is independent of  $\mathcal{F}_{\tau_{i,r}}$ .

On the event  $t_1 < T_{(1)}$ , using (5.30), we have  $Y_{i,r}(t) \leq X_3(t) \leq \eta N \leq \frac{N}{3}$  for all  $t \in [0, t_1]$ . Following the same reasoning in (5.35), we get

$$\lambda(t_1 - t_{0,r}) \geq \frac{t_1 - t_{0,r}}{3}.$$

By similar argument as in Lemma 5.9, we have

$$P\left(\{Y_{i,r}(t_1) > 0\} \cap \{t_1 < T_{(1)}\} \middle| \tau_{i,r} \leq t_{0,r} \wedge T_{(1)}\right) \leq P\left(Y^\#\left(\frac{t_1 - t_{0,r}}{3}\right) > 0\right).$$

Here, we finish this prove by showing that for sufficiently large  $N$ ,

$$P\left(Y^\#\left(\frac{t_1 - t_{0,r}}{3}\right) > 0\right) \leq 4s.$$

By the same argument that we used to get (5.37), for  $t \geq 0$ ,

$$P(Y^\#(t) > 0) = \frac{1 - (1 - 3s)}{1 - (1 - 3s)e^{-(1 - (1 - 3s))t}} \leq \frac{3s}{1 - e^{-3st}}. \tag{5.39}$$

From (5.39) and the definition of  $C_{0,r}$  in (5.7), in the recombination dominating case and the mutation dominating case with  $Nr \geq e$ , we have that for sufficiently large  $N$ ,

$$\begin{aligned} P\left(Y^\#\left(\frac{t_1 - t_{0,r}}{3}\right) > 0\right) &\leq \frac{3s}{1 - e^{-s(t_1 - t_{0,r})}} \\ &= \frac{3s}{1 - \frac{1}{\sqrt{Nr}}e^{-(C_{0,r} - C_1)}} \\ &< \frac{3s}{1 - e^{-(C_{0,r} - C_1)}} \\ &\leq 4s. \end{aligned}$$

In the mutation dominating case with  $Nr \leq e$ , we also have

$$P\left(Y^\# \left(\frac{t_1 - t_{0,r}}{3}\right) > 0\right) \leq \frac{3s}{1 - e^{-s(t_1 - t_{0,r})}} = \frac{3s}{1 - e^{-(C_{0,r} - C_1)}} \leq 4s.$$

Thus, we prove our result.  $\square$

At this point, we are ready to show that the events  $A_8$  and  $A_9$  occur with probability close to 1.

**Lemma 5.11.** *For sufficiently large  $N$ , we have that  $P(A_8^c) \leq 4\epsilon$  and  $P(A_9^c) \leq 4\epsilon$ .*

*Proof:* Recall the definitions of  $A_8$  and  $A_9$  in (5.31) and (5.32). We will only show that  $P(A_8^c) \leq 4\epsilon$ . The same reasoning can be applied to prove that  $P(A_9^c) \leq 4\epsilon$ .

Let  $J = \lfloor e^{-C_{0,m}/2}/s \rfloor$ . By Lemma 5.9, we have that for sufficiently large  $N$ ,

$$\begin{aligned} &P(\{X_{3m}^{(0,t_0,m]}(t_1) > 0\} \cap \{m(t_{0,m} \wedge T_{(1)}) < e^{-C_{0,m}/2}/s\} \cap \{t_1 < T_{(1)}\}) \\ &\leq \sum_{i=1}^J P\left(\{Y_{i,m}(t_1) > 0\} \cap \{\tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\} \cap \{t_1 < T_{(1)}\}\right) \\ &\leq \sum_{i=1}^J P\left(\{Y_{i,m}(t_1) > 0\} \cap \{t_1 < T_{(1)}\} \mid \tau_{i,m} \leq t_{0,m} \wedge T_{(1)}\right) \\ &\leq 3sJ \\ &\leq 3e^{-C_{0,m}/2}. \end{aligned}$$

Hence, by Lemmas 5.6 and 5.8, along with the way we choose constants  $\epsilon, K$  and  $C_{0,m}$  in (5.1), (5.3) and (5.5), for sufficiently large  $N$ ,

$$\begin{aligned} &P(\{X_{3m}^{(0,t_0,m]}(t_1 \wedge T_{(1)}) > 0\}) \\ &\leq P(\{X_{3m}^{(0,t_0,m]}(t_1 \wedge T_{(1)}) > 0\} \cap \{m(t_{0,m} \wedge T_{(1)}) < e^{-C_{0,m}/2}/s\} \cap \{t_1 < T_{(1)}\}) \\ &\quad + P(m(t_{0,m} \wedge T_{(1)}) \geq e^{-C_{0,m}/2}/s) + P(T_{(1)} \leq t_1) \\ &\leq 3e^{-C_{0,m}/2} + 3\epsilon \\ &\leq 4\epsilon. \end{aligned}$$

So, this proves that  $P(A_8^c) \leq 4\epsilon$ .  $\square$

In the last part of this subsection, we show that events  $A_{10}$  and  $A_{11}$  occur with probability close to 1.

**Lemma 5.12.** *For sufficiently large  $N$ , we have  $P(A_{10}^c) \leq \epsilon$ , and  $P(A_{11}^c) \leq \epsilon$ .*

*Proof:* Recall the definition of  $A_{10}$  in (5.33). From Lemma 5.4 and the definition of  $t_1$  in (5.12), for sufficiently large  $N$ ,

$$E\left[X_{3m}^{(t_0,m,t_1]}(t_1 \wedge T_{(1)})\right] \leq \frac{2Ke^{C_{0,m}}N^2\mu^4}{s^3}e^{2st_1} = \frac{2Ke^{-2C_1+C_{0,m}}N^2\mu^2}{s}.$$

So, we get that  $P(A_{10}^c) \leq \epsilon$  from the Markov's inequality.

Now, recall the definition of  $A_{11}$  in (5.34). We will first consider the recombination dominating case and the mutation dominating case with  $Nr \geq e$ . In the

recombination dominating case,  $Nr \gg 1$ . From Lemma 5.5 and the definition of  $t_{0,r}$  in (5.9), for sufficiently large  $N$ ,

$$\begin{aligned} E \left[ X_{3r}^{(t_{0,r}, t_1]}(t_1 \wedge T_{(1)}) \right] &\leq \frac{2K^2 N \mu^2 r}{s^2} e^{2st_1} (t_1 - t_{0,r}) \\ &= 2K^2 e^{-2C_1} Nr \left( \frac{\ln(Nr)}{2s} + \frac{C_{0,r} - C_1}{s} \right) \\ &\leq \frac{K^2 e^{-2C_1} (2(C_{0,r} - C_1) + 1) Nr \ln(Nr)}{s}. \end{aligned}$$

In the mutation dominating case with  $Nr < e$ , from Lemma 5.5 and the definition of  $t_{0,r}$  in (5.9), for sufficiently large  $N$ ,

$$\begin{aligned} E \left[ X_{3r}^{(t_{0,r}, t_1]}(t_1 \wedge T_{(1)}) \right] &\leq \frac{K^2 N \mu^2 r}{s^2} e^{2st_1} (t_1 - t_{0,r}) \\ &= K^2 e^{-2C_1} Nr \left( \frac{C_{0,r} - C_1}{s} \right) \\ &\leq \frac{K^2 e^{-2C_1} (2(C_{0,r} - C_1) + 1)}{2s}. \end{aligned}$$

Thus, in both cases, for sufficiently large  $N$ ,

$$E \left[ X_{3r}^{(t_{0,r}, t_1]}(t_1 \wedge T_{(1)}) \right] \leq \frac{K^2 e^{-2C_1+1} (2(C_{0,r} - C_1) + 1) (1 \vee Nr \ln_+(Nr))}{2s}.$$

Therefore,  $P(A_{11}^c) \leq \epsilon$  by the Markov’s inequality. □

5.5. *The proof of Proposition 2.2.* First of all, we define

$$A_{(1)} = \begin{cases} \bigcap_{1 \leq i \leq 11, i \neq 6} A_i & \text{in the recombination dominating case} \\ \bigcap_{1 \leq i \leq 11, i \neq 7} A_i & \text{in the mutation dominating case.} \end{cases}$$

We will give both upper and lower bounds for the numbers of type 1 and 2 individuals on the event  $A_{(1)}$  before we prove Proposition 2.2.

**Lemma 5.13.** *The following statements hold.*

- (1) *On the event  $A_{(1)}$ , for  $i = 1, 2$ , for sufficiently large  $N$  and for  $t \in [0, t_1]$ ,*

$$X_i(t) \leq (1 + \delta^2) \frac{2N\mu}{s} e^{st}.$$

- (2) *In the recombination dominating case, on the event  $A_{(1)}$ , for  $i = 1, 2$  and for sufficiently large  $N$ ,*

$$X_i(t) \geq (1 - \delta^2) \frac{2N\mu}{s} e^{st}$$

*for all  $t \in [t_{0,r}, t_1]$ .*

- (3) *In the mutation dominating case, on the event  $A_{(1)}$ , for  $i = 1, 2$  and for sufficiently large  $N$ ,*

$$X_i(t) \geq (1 - \delta^2) \frac{2N\mu}{s} e^{st}$$

*for all  $t \in [t_{0,m}, t_1]$ .*

*Proof:* Throughout this proof, we assume that we are on the event  $A_{(1)}$ . From (4.3), we have that for all  $t \in (0, t_1]$ ,

$$X_{1m}(t) = X_{1m}^{(0,t_1]}(t) = \int_0^t M_1^{(0,t_1]}(u) e^{\int_u^t G_1(v) dv} du + Z_{1m}^{(0,t_1]}(t) e^{\int_0^t G_1(v) dv}. \quad (5.40)$$

From Lemma 5.1, definitions of  $A_1$  and  $A_2$  in (5.23) and (5.24), and the fact that  $1 \ll N\mu$ , for sufficiently large  $N$ ,

$$\begin{aligned} X_{1m}(t) &\leq \int_0^t 2N\mu e^{\int_u^t s dv} du + \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu}{s^2}} \cdot e^{\int_0^t s dv} \\ &\leq \frac{2N\mu}{s} \cdot e^{st} + \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu}{s^2}} \cdot e^{st} \\ &\leq \left(1 + \frac{\delta^2}{2}\right) \frac{2N\mu}{s} e^{st} \end{aligned}$$

for all  $t \in (0, t_1]$ . Next, from (4.4), we have that for all  $t \in (0, t_1]$ ,

$$X_{1r}(t) = X_{1r}^{(0,t_1]}(t) = \int_0^t R_1^{(0,t_1]}(u) e^{\int_u^t G_{1r}^{(0,t_1]}(v) dv} du + Z_{1r}^{(0,t_1]}(t) e^{\int_0^t G_{1r}^{(0,t_1]}(v) dv}.$$

From (3.16), Lemma 5.1, and definitions of  $A_1$  and  $A_4$  in (5.23) and (5.25), for sufficiently large  $N$  and  $t \in (0, t_1]$ ,

$$X_{1r}(t) \leq \int_0^t Nr \tilde{X}_0(u) \tilde{X}_3(u) e^{\int_u^t (s+r) dv} du + \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu r}{s^3} \ln\left(\frac{s}{\mu}\right)} e^{\int_0^t (s+r) dv}.$$

Using the definition of  $T_3$  in (5.15), we have that for sufficiently large  $N$ ,

$$\begin{aligned} X_{1r}(t) &\leq 2Nr \int_0^t \frac{\mu}{s} e^{su} \cdot e^{s(t-u)+rt} du + \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu r}{s^3} \ln\left(\frac{s}{\mu}\right)} e^{st+rt} \\ &= e^{rt} \left( \frac{2N\mu r}{s} e^{st} t + \sqrt{\frac{24}{\epsilon} \cdot \frac{N\mu r}{s^3} \ln\left(\frac{s}{\mu}\right)} e^{st} \right) \\ &\leq \frac{2N\mu}{s} e^{st} \cdot e^{rt_1} \left( rt_1 + \sqrt{\frac{6}{\epsilon} \cdot \frac{1}{N\mu} \cdot \frac{r}{s} \ln\left(\frac{s}{\mu}\right)} \right) \\ &\leq \frac{\delta^2}{2} \cdot \frac{2N\mu}{s} e^{st} \end{aligned}$$

for all  $t \in (0, t_1]$ . Therefore, for sufficiently large  $N$ , for all  $t \in [0, t_1]$ , we have

$$X_1(t) = X_{1m}(t) + X_{1r}(t) \leq (1 + \delta^2) \frac{2N\mu}{s} e^{st},$$

Note that by a similar argument, we can also prove the upper bound for  $X_2(t)$ .

For the lower bound for  $X_1(t)$  in the recombination dominating case, it is enough to find a lower bound of  $X_{1m}^{(0,t_1]}(t)$ , since  $X_1(t) \geq X_{1m}^{(0,t_1]}(t)$  for all  $t \geq 0$ . We first need to consider the term  $\int_u^t G_1(v) dv$ . By using (3.18), part 1 of this lemma, and the definition of  $T_3$  in (5.15), when  $N$  is sufficiently large, if  $0 \leq u < t \leq t_1$ ,

then

$$\begin{aligned} \int_u^t G_1(v)dv &\geq \int_u^t (s - s\tilde{X}_1(v) - s\tilde{X}_2(v) - 2s\tilde{X}_3(v) - r - \mu)dv \\ &\geq \int_u^t \left( s - s \cdot (1 + \delta^2) \frac{\mu}{s} e^{su} - s \cdot (1 + \delta^2) \frac{\mu}{s} e^{su} - 2s \cdot \frac{\mu}{s} e^{su} - r - \mu \right) dv \\ &\geq s(t - u) - \frac{(4 + 2\delta^2)\mu}{s} e^{st_1} - (r + \mu)t_1. \end{aligned}$$

Now, using the fact that  $\delta < 1$ , the definition of  $t_1$  in (5.12), along with (5.29), we have that when  $N$  is sufficiently large, when  $0 \leq u < t \leq t_1$ ,

$$\int_u^t G_1(v)dv \geq s(t - u) - \frac{6\mu}{s} e^{st_1} - (r + \mu)t_1 \geq s(t - u) - 7e^{-C_1}. \quad (5.41)$$

Also, using statement (1) of this lemma, the definition of  $T_3$  in (5.15), and the fact that  $\delta < 1$ , for sufficiently large  $N$ , and  $u \in [0, t_1]$ ,

$$\begin{aligned} X_0(u) &= 2N - X_1(u) - X_2(u) - X_3(u) \\ &\geq 2N - 2(1 + \delta^2) \frac{2N\mu}{s} e^{su} - \frac{N\mu}{s} e^{su} \\ &\geq 2N - \frac{9N\mu}{s} e^{su}. \end{aligned} \quad (5.42)$$

Thus, from (5.40), (5.41), and (5.42), along with the definition of  $A_4$  in (5.25), for sufficiently large  $N$ , when  $t \in [t_{0,r}, t_1]$ ,

$$\begin{aligned} X_{1m}^{(0,t_1]}(t) &\geq \int_0^t \mu \left( 2N - \frac{9N\mu}{s} e^{su} \right) e^{s(t-u)-7e^{-C_1}} du - \sqrt{\frac{24}{\epsilon}} \cdot \frac{N\mu}{s^2} e^{st} \\ &= \frac{2N\mu}{s^2} e^{st} \cdot \left[ e^{-7e^{-C_1}} \left( 1 - e^{-st} - \frac{9}{2}\mu t \right) - \sqrt{\frac{6}{\epsilon}} \cdot \frac{1}{2N\mu} \right] \\ &\geq \frac{2N\mu}{s^2} e^{st} \cdot \left[ (1 - 7e^{-C_1}) \left( 1 - e^{-st_{0,r}} - \frac{9}{2}\mu t_1 \right) - \sqrt{\frac{6}{\epsilon}} \cdot \frac{1}{N\mu} \right]. \end{aligned} \quad (5.43)$$

In the recombination dominating case,  $N\mu^2 \ll s$  and  $r \ll s$ . By using the definition of  $t_{0,r}$  in (5.9), we have that

$$st_{0,r} = \frac{1}{2} \ln \left( \frac{s^2}{N\mu^2 r} \right) - C_{0,r} \gg 1.$$

Thus, from (5.43), (5.29), and the way we choose  $C_1$  as in (5.4), for sufficiently large  $N$ , and all  $t \in [t_{0,r}, t_1]$ ,

$$X_1(t) \geq (1 - 8e^{-C_1}) \frac{2N\mu}{s} e^{st} \geq (1 - \delta^2) \frac{2N\mu}{s} e^{st}.$$

The proof for the mutation dominating case is almost the same by replacing  $t_{0,r}$  by  $t_{0,m}$  and using the fact that  $st_{0,m} = \ln \left( \frac{s}{N\mu^2} \right) - C_{0,m} \gg 1$ .  $\square$

Now, we give a proof of Proposition 2. The main strategy is using (4.3) and (4.4) to show that the dominant terms have higher orders than (or equal to) the fluctuation terms on event  $A_{(1)}$ .

*Proof:* By Lemmas 5.6, 5.7, 5.11, and 5.12, for sufficiently large  $N$ , we have that  $P(A_{(1)}) \geq 1 - 17\epsilon$ . From now on, we will assume that we are in the event  $A_{(1)}$ . Statement (1) follows from Lemma 5.13 by inserting  $t = t_1$ .

Now consider  $X_3(t_1) = X_{3m}(t_1) + X_{3r}(t_1)$ . From the definitions of  $A_8, A_9, A_{10}$  and  $A_{11}$ , in (5.31), (5.32), (5.33) and (5.34), it follows that

$$X_{3m}(t_1) = X_{3m}^{(0,t_0,m]}(t_1) + X_{3m}^{(t_0,m,t_1]}(t_1) \leq \left( \frac{2Ke^{-2C_1+C_{0,m}}}{\epsilon} \right) \frac{N^2\mu^2}{s},$$

and

$$\begin{aligned} X_{3r}(t_1) &= X_{3r}^{(0,t_0,r]}(t_1) + X_{3r}^{(t_0,r,t_1]}(t_1) \\ &\leq \left( \frac{K^2e^{-2C_1}(2(C_{0,r} - C_1) + 1)}{\epsilon} \right) \frac{(1 \vee Nr \ln_+(Nr))}{s}. \end{aligned}$$

In the recombination dominating case,  $N^2\mu^2 \ll Nr \ln(Nr)$ . Thus, there is positive constant  $K_{1r}^+$ , for sufficiently large  $N$ ,

$$X_3(t_1) \leq \frac{K_{1r}^+ Nr \ln(Nr)}{s}.$$

In the mutation dominating case,  $Nr \ln_+(Nr) \leq CN^2\mu^2$ . Thus, there is positive constant  $K_{1m}^+$ , for sufficiently large  $N$ ,

$$X_3(t_1) \leq \frac{K_{1m}^+ N^2\mu^2}{s}.$$

Now, we prove the lower bound of  $X_3(t_1)$ . First, consider the recombination dominating case. In this case, the significant contribution to  $X_3(t_1)$  comes from the number of type 3r individuals descend from ancestors that originate after time  $t_{0,r}$ . Thus, it is enough to consider  $X_{3m}^{(t_0,r,t_1]}(t_1)$ .

We need to consider the term  $\int_u^{t_1} G_{3r}^{(t_0,r,t_1]}(v)dv$ . Similar to the way we get (5.41), by using (3.22) instead of (3.20), if  $t_{0,r} \leq u \leq t_1$ , then

$$\begin{aligned} &\int_u^{t_1} G_{3r}^{(t_0,r,t_1]}(v)dv \\ &\geq \int_u^{t_1} (2s - s\tilde{X}_1(v) - s\tilde{X}_2(v) - 2s\tilde{X}_3(v) - r)dv \\ &\geq \int_u^{t_1} \left( 2s - s \cdot (1 + \delta^2) \frac{\mu}{s} e^{su} - s \cdot (1 + \delta^2) \frac{\mu}{s} e^{su} - 2s \cdot \frac{\mu}{2s} e^{su} - r \right) dv \\ &\geq 2s(t_1 - u) - 5e^{-C_1} - rt_1. \end{aligned}$$

Since  $rt_1 \ll 1$  by (5.29), when  $N$  is sufficiently large, for  $t_{0,r} \leq u \leq t_1$ ,

$$\int_u^{t_1} G_{3r}^{(t_0,r,t_1]}(v)dv \geq 2s(t_1 - u) - 6e^{-C_1}. \tag{5.44}$$

By (3.14) and Lemma 5.13, for sufficiently large  $N$ , when  $t \in [t_{0,r}, t_1]$ ,

$$\begin{aligned} R_3^{(t_{0,r}, t_1]}(t) &\geq X_0(t) \cdot r \tilde{X}_1(t) \tilde{X}_2(t) \\ &= (2N - X_1(t) - X_2(t) - X_3(t)) \cdot r \tilde{X}_1(t) \tilde{X}_2(t) \\ &\geq \left(N - 2(1 + \delta^2) \frac{N\mu}{s} e^{st} - \frac{N\mu}{s} e^{st}\right) \left((1 - \delta^2)^2 \frac{\mu^2 r}{s^2} e^{2st}\right) \\ &= \left(1 - \frac{9e^{-C_1}}{2}\right) (1 - \delta^2)^2 \cdot \frac{2N\mu^2 r}{s^2} e^{2st}. \end{aligned}$$

Using (4.4), Lemma 5.1, and the definitions of  $A_7$  in (5.27), for sufficiently large  $N$ ,

$$\begin{aligned} X_{3r}^{(t_{0,r}, t_1]}(t_1) &\geq \int_{t_{0,r}}^{t_1} \left(1 - \frac{9e^{-C_1}}{2}\right) (1 - \delta^2)^2 \cdot \frac{2N\mu^2 r}{s^2} e^{2su} \cdot e^{2s(t_1-u) - 6e^{-C_1}} du \\ &\quad - \sqrt{\frac{8K^2 e^{-2C_{0,r}}}{\epsilon} \cdot \frac{\ln(Nr)}{s^2}} \cdot e^{\int_{t_{0,r}}^{t_1} (2s+r)dv} \\ &= e^{-6e^{-C_1}} \left(1 - \frac{9e^{-C_1}}{2}\right) (1 - \delta^2)^2 \cdot \frac{2N\mu^2 r}{s^2} e^{2st_1} (t_1 - t_{0,r}) \\ &\quad - \sqrt{\frac{8K^2 e^{-2C_{0,r}}}{\epsilon} \cdot \frac{\ln(Nr)}{s^2}} \cdot e^{(2s+r)(t_1 - t_{0,r})} \\ &= \frac{Nr \ln(Nr)}{s} \cdot \left[ e^{-6e^{-C_1}} \left(1 - \frac{9e^{-C_1}}{2}\right) (1 - \delta^2)^2 e^{-2C_1} \left(1 + \frac{2(C_{0,r} - C_1)}{\ln(Nr)}\right) \right. \\ &\quad \left. - \frac{\sqrt{8K} e^{-2C_1 + C_{0,r} + rt_1}}{\sqrt{\epsilon \ln(Nr)}} \right]. \end{aligned}$$

Since  $rt_1 \ll 1$  and  $1 \ll Nr$ , for sufficiently large  $N$ , there is a positive constant  $K_{1r}^-$  such that

$$X_3(t_1) \geq \frac{K_{1r}^- Nr \ln(Nr)}{s}.$$

Then, we choose the positive constant

$$K_{1r}^- = \frac{e^{-7e^{-C_1}} (1 - 5e^{-C_1}) (1 - \delta^2)^2 e^{-2C_1}}{3}.$$

Lastly, consider the mutation dominating case. Here, we will find a lower bound of  $X_{3m}^{(t_{0,m}^+, t_1]}(t_1)$  instead of  $X_{3r}^{(t_{0,r}, t_1]}(t_1)$ . By the argument used to obtain (5.44), for sufficiently large  $N$ , if  $t_{0,m} \leq u \leq t_1$ , then  $\int_u^{t_1} G_3(v)dv \geq 2s(t_1 - u) - 6e^{-C_1}$ . From (4.3), Lemma 5.1, Lemma 5.13, and the definition of  $A_6$  in (5.26), for sufficiently

large  $N$ ,

$$\begin{aligned}
 X_{3m}^{(t_0^+, t_1]}(t_1) &= \int_{t_0^+}^{t_1} \mu(X_1(u) + X_2(u)) e^{\int_u^{t_1} G_3(v) dv} du + Z_{3m}^{(t_0^+, t_1]}(t_1) e^{\int_{t_0^+}^{t_1} G_3(v) dv} \\
 &\geq \int_{t_0^+}^{t_1} 2(1 - \delta^2) \cdot \frac{2N\mu^2}{s} e^{su} \cdot e^{2s(t_1-u) - 6e^{-C_1}} du - \sqrt{\frac{12Ke^{C_{0,m}^+}}{\epsilon}} \cdot \frac{1}{s^2} \cdot e^{\int_{t_0^+}^{t_1} 2s dv} \\
 &= 4(1 - \delta^2) e^{-6e^{-C_1}} \cdot \frac{N\mu^2}{s^2} e^{2st_1} (e^{-st_0^+} - e^{-st_1}) - \sqrt{\frac{12Ke^{C_{0,m}^+}}{\epsilon}} \cdot \frac{1}{s^2} \cdot e^{2s(t_1 - t_0^+)} \\
 &= \frac{N^2\mu^2}{s} \left( 4(1 - \delta^2) e^{-6e^{-C_1} - 2C_1 - C_{0,m}^+} \left( 1 - e^{C_1 + C_{0,m}^+} \cdot \frac{1}{N\mu} \right) \right. \\
 &\quad \left. - \sqrt{\frac{12Ke^{C_{0,m}^+}}{\epsilon}} \cdot e^{-2C_1 - 2C_{0,m}^+} \right) \\
 &\geq \frac{N^2\mu^2}{s} \cdot e^{-2C_1 - C_{0,m}^+} \left( (1 - \delta^2) e^{-6e^{-C_1}} - \sqrt{\frac{12Ke^{-C_{0,m}^+}}{\epsilon}} \right).
 \end{aligned}$$

Note that we defined  $C_{0,m}^+$  in (5.6) precisely to make

$$(1 - \delta^2) e^{-6e^{-C_1}} - \sqrt{\frac{12Ke^{-C_{0,m}^+}}{\epsilon}} > 0.$$

This completes the proof. □

### 6. Phase 2 and the proof of Proposition 2.3

Firstly, we show that with high probability, the numbers of type 1 individuals and type 2 individual grow approximately logistically in subsection 6.1. The proof will use a theorem in Darling and Norris (2008). Then, in subsection 6.2, we prove the results on type 3 by splitting type 3 individuals into two groups: one for those that descend from type 3 individuals at time  $t_1$  and another for individuals that descend from a type 3 ancestor that appears in the time interval  $(t_1, t_2]$ . The reason that we split type 3 into two groups is that the majority of type 3 individuals at time  $t_2$  should have descended from those at time  $t_1$ , and the type 3 ancestors that appear in the time interval  $(t_1, t_2]$  do not make significant contribution to type 3 population at time  $t_2$ . The results are again proved by expectation and variance technique used in the proof of phase 1. Lastly, by combining the results from these two subsections, we can prove Proposition 2.3.

6.1. *Comparing the Markov chain with logistic functions.* We will compare our population process with the solution to a differential equation. The components of this solution will be logistic functions. Theorem 6.1 below is a special case of Theorem 4.1 in Darling and Norris (2008). The following is the setup for Theorem 6.1. Let  $(\mathbf{X}(t), t \geq 0)$  be a continuous-time Markov chain with finite state space  $S \subset \mathbb{R}^3$ .

Let  $q(\xi, \xi')$  be the jump rate from the state  $\xi$  to the state  $\xi'$ . For each state  $\xi \in S$ , define the function  $\alpha : S \rightarrow \mathbb{R}$  by

$$\alpha(\xi) = \sum_{\xi' \neq \xi} |\xi' - \xi|^2 q(\xi, \xi'), \tag{6.1}$$

where  $|\cdot|$  is the Euclidean norm. We also define the function  $\beta : S \rightarrow \mathbb{R}^3$  by

$$\beta(\xi) = \sum_{\xi' \neq \xi} (\xi' - \xi) q(\xi, \xi') \tag{6.2}$$

for every  $\xi \in S$ . It follows that

$$\mathbf{X}(t) = \mathbf{X}(0) + M(t) + \int_0^t \beta(\mathbf{X}(s)) ds, \quad \text{for } t \geq 0,$$

for some martingale  $(M(t), t \geq 0)$ .

Let  $b : [0, 1]^3 \rightarrow \mathbb{R}^3$  be a Lipschitz function with Lipschitz constant  $K$ . Let  $x : [0, \infty) \rightarrow \mathbb{R}^3$  be the function that satisfies

$$x(t) = x(0) + \int_0^t b(x(s)) ds, \quad \text{for } t \geq 0.$$

The goal is to compare  $\mathbf{X}(t)$  with  $x(t)$ .

Fix  $T > 0$ ,  $\epsilon_0 > 0$ ,  $L > 0$ , and let  $\Delta = \epsilon_0 e^{-KT}/3$ . Define the events

$$\begin{aligned} \Omega_0 &= \{|\mathbf{X}(0) - x(0)| \leq \Delta\}, \\ \Omega_1 &= \left\{ \int_0^T |\beta(\mathbf{X}(t)) - b(\mathbf{X}(t))| dt \leq \Delta \right\}, \\ \Omega_2 &= \left\{ \int_0^T \alpha(\mathbf{X}(t)) dt \leq LT \right\}. \end{aligned}$$

**Theorem 6.1.** *Under all the assumptions above,*

$$P\left(\sup_{0 \leq t \leq T} |\mathbf{X}(t) - x(t)| > \epsilon_0\right) \leq \frac{4LT}{\Delta^2} + P(\Omega_0^c \cup \Omega_1^c \cup \Omega_2^c).$$

Now, we will apply this to our process  $((X_0(t), X_1(t), X_2(t), X_3(t)), t \geq 0)$ . First, for  $t \geq 0$ , we define

$$\mathbf{X}(t) = (\tilde{X}_1(t), \tilde{X}_2(t), \tilde{X}_3(t)), \tag{6.3}$$

and  $S = \{(\xi_1, \xi_2, \xi_3) \in \{0, \frac{1}{2N}, \dots, \frac{2N-1}{2N}, 1\}^3 : \xi_1 + \xi_2 + \xi_3 \leq 1\}$ . We are thinking of  $\xi_1, \xi_2$  and  $\xi_3$  as the fractions of type 1, 2 and 3 individuals in the population. For better understanding in the following formulas, we will define  $\xi_0 = 1 - \xi_1 - \xi_2 - \xi_3$ , which represents the fraction of type 0 individuals in the population. Now, for each  $\xi = (\xi_1, \xi_2, \xi_3) \in S$ , we define

$$\begin{aligned} f_0(\xi) &= (1 - r)\xi_0 + r(\xi_0 + \xi_1)(\xi_0 + \xi_2), \\ f_1(\xi) &= (1 - r)\xi_1 + r(\xi_1 + \xi_3)(\xi_0 + \xi_1), \\ f_2(\xi) &= (1 - r)\xi_2 + r(\xi_0 + \xi_2)(\xi_2 + \xi_3), \\ f_3(\xi) &= (1 - r)\xi_3 + r(\xi_1 + \xi_3)(\xi_2 + \xi_3). \end{aligned}$$

Note that for each  $i = 0, 1, 2, 3$ , the quantity  $f_i(\xi)$  represents the probability that a new individual born is of type  $i$ . Next, for  $\xi = (\xi_1, \xi_2, \xi_3)$  and  $\xi' = (\xi'_1, \xi'_2, \xi'_3)$  in

$S$ , the transition rate  $q(\xi, \xi')$  is given by

$$q(\xi, \xi') = \begin{cases} 2N\xi_0f_1(\xi) + 2N\xi_0\mu, & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1 + \frac{1}{2N}, \xi_2, \xi_3) \\ 2N\xi_0f_2(\xi) + 2N\xi_0\mu, & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1, \xi_2 + \frac{1}{2N}, \xi_3) \\ 2N\xi_0f_3(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1, \xi_2, \xi_3 + \frac{1}{2N}) \\ 2N(1-s)\xi_1f_0(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1 - \frac{1}{2N}, \xi_2, \xi_3) \\ 2N(1-s)\xi_1f_2(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1 - \frac{1}{2N}, \xi_2 + \frac{1}{2N}, \xi_3) \\ 2N(1-s)\xi_1f_3(\xi) + 2N\xi_1\mu, & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1 - \frac{1}{2N}, \xi_2, \xi_3 + \frac{1}{2N}) \\ 2N(1-s)\xi_2f_0(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1, \xi_2 - \frac{1}{2N}, \xi_3) \\ 2N(1-s)\xi_2f_1(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1 + \frac{1}{2N}, \xi_2 - \frac{1}{2N}, \xi_3) \\ 2N(1-s)\xi_2f_3(\xi) + 2N\xi_2\mu, & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1, \xi_2 - \frac{1}{2N}, \xi_3 + \frac{1}{2N}) \\ 2N(1-2s)\xi_3f_0(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1, \xi_2, \xi_3 - \frac{1}{2N}) \\ 2N(1-2s)\xi_3f_1(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1 + \frac{1}{2N}, \xi_2, \xi_3 - \frac{1}{2N}) \\ 2N(1-2s)\xi_3f_2(\xi), & \text{if } (\xi'_1, \xi'_2, \xi'_3) = (\xi_1, \xi_2 + \frac{1}{2N}, \xi_3 - \frac{1}{2N}) \\ 0, & \text{otherwise.} \end{cases} \tag{6.4}$$

The reasons behind the formulas for these rates are similar to the ones we used to obtain the birth and death rates in section 3. For instance, the first rate is the rate that the number of type 0 individuals decreases by 1, and the number of type 1 individuals increases by 1. There are two ways for this to occur. First, a type 0 individual mutates to type 1, and the mutation occurs at a total rate of  $2N\xi_0\mu$ . Second, a type 0 individual dies and is replaced by a type 1 individual. The total rate that a type 0 individual dies is  $2N\xi_0$ , and the probability that the replacement is of type 1 is  $f_1(\xi)$ .

We define the functions  $\alpha$  and  $\beta$  as in (6.1) and (6.2). For  $\xi, \xi' \in S$  such that  $q(\xi, \xi') \neq 0$ , we have  $|\xi - \xi'|^2 \leq 1/(2N^2)$  since it is equal to  $1/(4N^2)$  or  $1/(2N^2)$ . Because  $\mu \ll s \ll 1$ , for sufficiently large  $N$ , we have  $q(\xi, \xi') \leq 4N$  for all  $\xi, \xi' \in S$ . By the definition of  $\alpha$  in (6.1), for sufficiently large  $N$ ,

$$\alpha(\xi) \leq \frac{24}{N}. \tag{6.5}$$

For each  $\xi \in S$ , we define

$$\gamma_s(\xi) = (\xi_0\xi_3 - \xi_1\xi_2)(1 - s\xi_1 - s\xi_2 - 2s\xi_3).$$

A tedious calculation gives

$$\beta(\xi) = s \begin{pmatrix} (1 - \xi_1 - \xi_2 - 2\xi_3)\xi_1 \\ (1 - \xi_1 - \xi_2 - 2\xi_3)\xi_2 \\ (2 - \xi_1 - \xi_2 - 2\xi_3)\xi_3 \end{pmatrix} + r\gamma_s(\xi) \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \mu \begin{pmatrix} \xi_0 - \xi_1 \\ \xi_0 - \xi_2 \\ \xi_1 + \xi_2 \end{pmatrix}. \tag{6.6}$$

Note that for  $i = 1, 2, 3$ , the  $i^{th}$  row of  $2N\beta(\xi)$  is exactly the rate at which the number of type  $i$  individuals increases by 1 subtracted by the rate at which the number of type  $i$  individuals decreases by 1.

Here, we define the functions  $b : [0, 1]^3 \rightarrow \mathbb{R}^3$  and  $\tilde{b} : [0, 1]^3 \rightarrow \mathbb{R}^3$  by

$$b(x_1, x_2, x_3) = s((1 - x_1 - x_2 - 2x_3)x_1, (1 - x_1 - x_2 - 2x_3)x_2, (2 - x_1 - x_2 - 2x_3)x_3), \tag{6.7}$$

and

$$\tilde{b}(x_1, x_2, x_3) = b(x_1, x_2, x_3)/s. \tag{6.8}$$

Since all first partial derivatives of  $\tilde{b}$  are bounded, the function  $\tilde{b}$  is Lipschitz. Hence,  $b$  is also Lipschitz with Lipschitz constant  $ks$ , where  $k > 0$  and  $k$  does not depend on  $N$ .

Next, we define a random variable  $B$  such that on the event that  $\tilde{X}_1(t_1) + \tilde{X}_2(t_1) > 0$ , we have

$$B = (\tilde{X}_1(t_1) + \tilde{X}_2(t_1))^{-1} - 1. \tag{6.9}$$

The value of  $B$  on the event that  $\tilde{X}_1(t_1) + \tilde{X}_2(t_1) = 0$  is not of interest, as we will work only on the event  $A_{(1)}$  when  $N$  is sufficiently large. Note that from Proposition 2.2,  $\tilde{X}_1(t_1) + \tilde{X}_2(t_1) > 0$  on the event  $A_{(1)}$ . Next, for  $t \geq t_1$ , we define

$$f(t) = \frac{1}{1 + Be^{-s(t-t_1)}}, \tag{6.10}$$

and we let

$$\begin{aligned} x(t) &= (x_1(t), x_2(t), x_3(t)) \\ &= \left( \left( \frac{\tilde{X}_1(t_1)}{\tilde{X}_1(t_1) + \tilde{X}_2(t_1)} \right) f(t), \left( \frac{\tilde{X}_2(t_1)}{\tilde{X}_1(t_1) + \tilde{X}_2(t_1)} \right) f(t), 0 \right). \end{aligned} \tag{6.11}$$

One could check that  $x'(t) = b(x(t))$  when  $t \geq t_1$ .

At last, we pick the constant

$$C_2 = -C_1 + \ln \left( \frac{e^{C_1}}{2(1 + \delta^2)} - 1 \right) + \ln \left( \frac{1}{\delta^2} - 1 \right), \tag{6.12}$$

and define

$$t_2 = \frac{1}{s} \ln \left( \frac{s}{\mu} \right) + \frac{C_2}{s}. \tag{6.13}$$

Note that  $C_2 > 0$ , which can be checked by considering the way we pick  $\delta$  and  $C_1$  in (5.2) and (5.4).

Now, we will use Theorem 6.1 to show that with probability almost 1, both  $X_1(t)$  and  $X_2(t)$  are close to  $2Nx_1(t)$  and  $2Nx_2(t)$  for  $t \in [t_1, t_2]$ . We define the event

$$A_{12} = \left\{ \sup_{t \in [t_1, t_2]} |X_i(t) - 2Nx_i(t)| \leq \left( \frac{\delta^4}{2} \right) N \text{ for } i = 1, 2 \right\}. \tag{6.14}$$

**Lemma 6.2.** *For sufficiently large  $N$ , we have  $P(A_{12}^c | \mathcal{F}_{t_1}) \leq \epsilon$  on the event  $A_{(1)}$ .*

*Proof:* Let  $\Delta = \delta^4 e^{-k(C_2+C_1)}/12$ . It is enough to prove that for sufficiently large  $N$ , on the event  $A_{(1)}$ ,

$$P \left( \sup_{t \in [t_1, t_2]} |\mathbf{X}(t) - x(t)| > \frac{\delta^4}{4} \mid \mathcal{F}_{t_1} \right) \leq \epsilon.$$

By (6.6) and (6.7), we have

$$\begin{aligned} &\beta(\mathbf{X}(t)) - b(\mathbf{X}(t)) \\ &= r\gamma_s((\tilde{X}_1(t), \tilde{X}_2(t), \tilde{X}_3(t))) \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \mu \begin{pmatrix} 1 - 2\tilde{X}_1(t) - \tilde{X}_2(t) - \tilde{X}_3(t) \\ 1 - \tilde{X}_1(t) - 2\tilde{X}_2(t) - \tilde{X}_3(t) \\ \tilde{X}_1(t) + \tilde{X}_2(t) \end{pmatrix}. \end{aligned}$$

Because  $\tilde{X}_i(t) \in [0, 1]$  for all  $i = 1, 2, 3$ , and  $t \geq 0$ , we have

$$|\beta(\mathbf{X}(t)) - b(\mathbf{X}(t))| \leq Dr + D'\mu$$

for some positive constants  $D$  and  $D'$ . Thus,

$$\int_{t_1}^{t_2} |\beta(\mathbf{X}(t)) - b(\mathbf{X}(t))| dt \leq (Dr + D'\mu)(t_2 - t_1) = (C_2 + C_1) \left( D \left( \frac{r}{s} \right) + D' \left( \frac{\mu}{s} \right) \right). \tag{6.15}$$

We note that for sufficiently large  $N$ ,

$$\left( \frac{96(C_2 + C_1)}{\Delta^2} \right) \left( \frac{1}{Ns} \right) \leq \epsilon, \tag{6.16}$$

and

$$(C_2 + C_1) \left( D \left( \frac{r}{s} \right) + D' \left( \frac{\mu}{s} \right) \right) \leq \Delta. \tag{6.17}$$

Also, in the recombination dominating case, for sufficiently large  $N$ ,

$$\frac{K_{1r}^+ r \ln(Nr)}{s} \leq \Delta, \tag{6.18}$$

and in the mutation dominating case, for sufficiently large  $N$ ,

$$\frac{K_{1m}^+ N \mu^2}{s} \leq \Delta, \tag{6.19}$$

where  $K_{1r}^+$  and  $K_{1m}^+$  are positive constants from Proposition 2.

From now on in this proof, we assume that  $N$  is large enough so that (6.16), (6.17) and (6.18) hold in the recombination dominating case and (6.16), (6.17) and (6.19) hold in the mutation dominating case.

Now, let us consider the process  $(\mathbf{X}(t), t \geq 0)$ . By Markov property of the process, if we condition on  $\mathcal{F}_{t_1}$ , the process after time  $t_1$  behaves as if we start the whole process again with  $\mathbf{X}(t_1)$  as the initial condition. Now, let us fix the value of  $\mathbf{X}(t_1) = (\xi_1, \xi_2, \xi_3)$  and consider the process starting at time  $t_1$  with this initial condition. Note that by starting the process from this fixed start point, the function  $f$  and  $x$  defined in (6.10) and (6.11) are no longer random, which allows us to use Theorem 6.1.

We define  $T = t_2 - t_1$ , and note that  $\Delta = \delta^4 e^{-k(C_2+C_1)}/12 = (\delta^4/4) \cdot e^{-(ks)T}/3$ , which is in the form required to use Theorem 6.1. We let  $L = 24/N$  and define the events

$$\begin{aligned} \Omega_0 &= \{|\mathbf{X}(t_1) - x(t_1)| \leq \Delta\}, \\ \Omega_1 &= \left\{ \int_{t_1}^{t_2} |\beta(\mathbf{X}(t)) - b(\mathbf{X}(t))| dt \leq \Delta \right\}, \\ \Omega_2 &= \left\{ \int_{t_1}^{t_2} \alpha(\mathbf{X}(t)) dt \leq LT \right\}. \end{aligned}$$

First, we consider  $\Omega_0$ . Note that on the event  $A(1)$ , we have  $\tilde{X}_1(t_1) = x_1(t_1)$  and  $\tilde{X}_2(t_1) = x_2(t_1)$ . Also, by Proposition 2, (6.18), and (6.19), we have  $X_3(t_1) \leq \Delta$ . Hence, in both the recombination dominating case and the mutation dominating case,  $|\mathbf{X}(t_1) - x(t_1)| = \tilde{X}_3(t_1)\Delta$ .

Next, because of (6.15) and (6.17), we have that  $\Omega_1^c = \emptyset$ . Lastly, by (6.5), it follows that

$$\int_{t_1}^{t_2} \alpha(\mathbf{X}(t))dt \leq \left(\frac{24}{N}\right)(t_2 - t_1) = LT.$$

So,  $\Omega_2^c = \emptyset$ .

Therefore, by Theorem 6.1 and (6.16), we have that

$$\begin{aligned} P\left(\sup_{t \in [t_1, t_2]} |\mathbf{X}(t) - x(t)| > \frac{\delta^4}{2} \mid \mathbf{X}(t_1) = (\xi_1, \xi_2, \xi_3)\right) \\ \leq \frac{4LT}{\Delta^2} + 0 \\ = \left(\frac{192(C_2 + C_1)}{\Delta^2}\right)\left(\frac{1}{Ns}\right) \\ \leq \epsilon. \end{aligned}$$

Note that the upper bound does not depend on the value of  $(\xi_1, \xi_2, \xi_3)$ . Thus, on the event  $A_{(1)}$ ,

$$P\left(\sup_{t \in [t_1, t_2]} |\mathbf{X}(t) - x(t)| > \frac{\delta^4}{2} \mid \mathcal{F}_{t_1}\right) \leq \epsilon.$$

We completes the proof. □

6.2. *Results on type 3 individuals.* We will now show that for sufficiently large  $N$ , with probability close to 1,  $X_3(t_2)$  has the same order as  $X_3(t_1)$ . The proof contains two parts. In the first part, we will show that  $X_3^{[t_1]}(t_2)$ , which was defined to be the number of type 3 individuals at time  $t$  that descend from the type 3 individuals at time  $t_1$ , has order  $(Nr \ln(Nr))/s$  in the recombination dominating case, and  $(N^2\mu^2)/s$  in the mutation dominating case. In the second part, we show that  $X_{3m}^{(t_1, t_2]}(t_2)$  and  $X_{3r}^{(t_1, t_2]}(t_2)$  have smaller orders.

Here, we choose the following constants:

$$K_{0m} = 2e^{2(C_2+C_1)}K_{1m}^+, \tag{6.20}$$

$$K_{0r} = 2e^{2(C_2+C_1)}K_{1r}^+, \tag{6.21}$$

$$K'_1 = 2e^{2(C_2+C_1)}(C_2 + C_1), \tag{6.22}$$

$$K'_2 = 2e^{3(C_2+C_1)}(C_2 + C_1). \tag{6.23}$$

We also define the following events.

$$A_{13} = \left\{X_{3m}^{(t_1, t_2]}(t_2) < \frac{K'_1}{\epsilon_1} \cdot \frac{N\mu}{s}\right\}, \tag{6.24}$$

$$A_{14} = \left\{X_{3r}^{(t_1, t_2]}(t_2) < \frac{K'_2}{\epsilon_1} \cdot \frac{Nr}{s}\right\}. \tag{6.25}$$

In the recombination dominating case, we define

$$A_{15} = \left\{\left|Z_3^{[t_1]}(t_2) - X_3(t_1)\right| < \sqrt{\frac{K_{0r}}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}}\right\}, \tag{6.26}$$

while in the mutation dominating case, we define

$$A_{15} = \left\{ \left| Z_3^{[t_1]}(t_2) - X_3(t_1) \right| < \sqrt{\frac{K_{0m}}{\epsilon}} \cdot \frac{N\mu}{s} \right\}. \tag{6.27}$$

We will first show that the event  $A_{15}$  occurs with high probability. The proof is shown in Lemma 6.4, which will need the following lemma in the argument.

**Lemma 6.3.** *For sufficiently large  $N$ , if  $t \geq t_1$ , then*

$$E \left[ X_3^{[t_1]}(t) \mid \mathcal{F}_{t_1} \right] \leq e^{2s(t-t_1)} X_3(t_1).$$

*Proof:* From (4.5), Proposition 4.3, and the fact that  $G_3(t) \leq 2s$  for all  $t \geq 0$ , we have that for  $t \geq t_1$ ,

$$\begin{aligned} X_3(t_1) &= Z_3^{[t_1]}(t_1) \\ &= E \left[ e^{-\int_{t_1}^t G_3(v)dv} X_3^{[t_1]}(t) \mid \mathcal{F}_{t_1} \right] \\ &\geq e^{-2s(t-t_1)} E \left[ X_3^{[t_1]}(t) \mid \mathcal{F}_{t_1} \right], \end{aligned}$$

which proves this lemma. □

**Lemma 6.4.** *For sufficiently large  $N$ , on the event  $A_{(1)}$ , we have  $P(A_{15}^c \mid \mathcal{F}_{t_1}) \leq \epsilon$ .*

*Proof:* First, consider the recombination dominating case. From (3.24) and (3.25), for all  $t \geq 0$ , we have that  $B_3^{[t_1]}(t) \leq 1$  and  $D_3^{[t_1]}(t) \leq 1$ . Also, from (3.19) and the fact that  $s \ll 1$ , for sufficiently large  $N$ , if  $t \geq 0$ , then

$$G_3(t) \geq -r. \tag{6.28}$$

By Proposition 4.3 and Lemma 6.3, for sufficiently large  $N$ ,

$$\begin{aligned} \text{Var} \left( Z_3^{[t_1]}(t_2) \mid \mathcal{F}_{t_1} \right) &\leq E \left[ \int_{t_1}^{t_2} e^{2r(u-t_1)} \cdot 2X_3^{[t_1]}(u)du \mid \mathcal{F}_{t_1} \right] \\ &\leq 2e^{2r(t_2-t_1)} \int_{t_1}^{t_2} e^{2s(u-t_1)} X_3(t_1) du \\ &= e^{2r(t_2-t_1)} \left( \frac{e^{2s(t_2-t_1)} - 1}{s} \right) X_3(t_1). \end{aligned}$$

By Proposition 2.2 and the definitions of  $t_1$  and  $t_2$  in (5.12) and (6.13), for sufficiently large  $N$ , on the event  $A_{(1)}$ ,

$$\text{Var} \left( Z_3^{[t_1]}(t_2) \mid \mathcal{F}_{t_1} \right) \leq e^{2(C_2+C_1) \cdot \frac{r}{s}} \left( \frac{e^{2(C_2+C_1)} K_{1r}^+ N r \ln(Nr)}{s^2} \right) \tag{6.29}$$

$$\begin{aligned} &\leq \frac{2e^{2(C_2+C_1)} K_{1r}^+ N r \ln(Nr)}{s^2} \\ &= \frac{K_{0r}^+ N r \ln(Nr)}{s^2}. \end{aligned} \tag{6.30}$$

The result follows from Chebyshev’s inequality.

For the mutation dominating case, the proof is almost precisely the same. The only difference is the inequality (6.29) which Proposition 2.2 gives that

$$\text{Var} \left( Z_3^{[t_1]}(t_2) \mid \mathcal{F}_{t_1} \right) \leq e^{2(C_2+C_1) \cdot \frac{r}{s}} \left( \frac{e^{2(C_2+C_1)} K_{1m}^+ N^2 \mu^2}{s^2} \right) \leq \frac{K_{0m}^+ N^2 \mu^2}{s^2}.$$

This completes the proof. □

Next, we prove that the events  $A_{13}$  and  $A_{14}$  occur with probabilities close to 1.

**Lemma 6.5.** *For sufficiently large  $N$ ,  $P(A_{13}^c | \mathcal{F}_{t_1}) \leq \epsilon$  and  $P(A_{14}^c | \mathcal{F}_{t_1}) \leq \epsilon$ .*

*Proof:* First of all, there is a mean-zero martingale  $(W_m(t), t \geq t_1)$  such that for all  $t \geq t_1$ ,

$$X_{3m}^{(t_1, t_2]}(t) = W_m(t) + \int_{t_1}^t \left( M_3^{(t_1, t_2]}(u) + G_3(u) X_{3m}^{(t_1, t_2]}(u) \right) du.$$

Also,  $G_3(u) \leq 2s$  for all  $u \geq t_1$ . Thus, from Lemma 5.1, for sufficiently large  $N$ , if  $t \in [t_1, t_2]$ , then

$$\begin{aligned} E \left[ X_{3m}^{(t_1, t_2]}(t) \mid \mathcal{F}_{t_1} \right] &= E \left[ \int_{t_1}^t \left( M_3^{(t_1, t_2]}(u) + G_3(u) X_{3m}^{(t_1, t_2]}(u) \right) du \mid \mathcal{F}_{t_1} \right] \\ &\leq 2N\mu(t_2 - t_1) + \int_{t_1}^t 2sE \left[ X_{3m}^{(t_1, t_2]}(u) \mid \mathcal{F}_{t_1} \right] du. \end{aligned}$$

Hence, from Gronwall's inequality,

$$E \left[ X_{3m}^{(t_1, t_2]}(t_2) \mid \mathcal{F}_{t_1} \right] \leq 2N\mu(t_2 - t_1)e^{2s(t_2 - t_1)} = \frac{K'_1 N \mu}{s}.$$

From Markov's inequality,  $P(A_{13}^c | \mathcal{F}_{t_1}) \leq \epsilon$ .

Similarly, there is a mean-zero martingale  $(W_r(t), t \geq t_1)$  such that

$$X_{3r}^{(t_1, t_2]}(t) = W_r(t) + \int_{t_1}^t \left( R_3^{(t_1, t_2]}(u) + G_{3r}^{(t_1, t_2]}(u) X_{3r}^{(t_1, t_2]}(u) \right) du$$

for all  $t \geq t_1$ . From (3.14), Lemma 5.1 and  $r \ll s$ , for sufficiently large  $N$  and for  $t \in [t_1, t_2]$ ,

$$\begin{aligned} E \left[ X_{3r}^{(t_1, t_2]}(t) \mid \mathcal{F}_{t_1} \right] &= E \left[ \int_{t_1}^t \left( R_3^{(t_1, t_2]}(u) + G_{3r}^{(t_1, t_2]}(u) X_{3r}^{(t_1, t_2]}(u) \right) du \mid \mathcal{F}_{t_1} \right] \\ &\leq E \left[ \int_{t_1}^t \left( 2Nr \tilde{X}_1(u) \tilde{X}_2(u) + (2s + r) X_{3r}^{(t_1, t_2]}(u) \right) du \mid \mathcal{F}_{t_1} \right] \\ &\leq 2Nr(t_2 - t_1) + \int_{t_1}^t 3sE \left[ X_{3r}^{(t_1, t_2]}(u) \mid \mathcal{F}_{t_1} \right] du. \end{aligned}$$

Thus, from Gronwall's inequality, we have

$$E \left[ X_{3r}^{(t_1, t_2]}(t_2) \mid \mathcal{F}_{t_1} \right] \leq Nr(t_2 - t_1)e^{3s(t_2 - t_1)} = \frac{K'_2 Nr}{s}.$$

So,  $P(A_{14}^c | \mathcal{F}_{t_1}) \leq \epsilon$  by Markov's inequality. □

### 6.3. The proof of Proposition 2.3.

*Proof:* First, we define

$$A_{(2)} = A_{(1)} \cap \left( \bigcap_{i=12}^{15} A_i \right).$$

From Proposition 6.2, Lemma 6.4, and Lemma 6.5, for sufficiently large  $N$ , on event  $A_{(1)}$

$$P\left(\bigcap_{i=12}^{15} A_i \mid \mathcal{F}_{t_1}\right) \geq 1 - 4\epsilon.$$

Thus, from Proposition 2.2, we have

$$P(A_{(2)}) = P\left(A_{(1)} \cap \left(\bigcap_{i=12}^{15} A_i\right)\right) \geq (1 - 4\epsilon) - P(A_{(1)}^c) \geq 1 - 21\epsilon.$$

From now on, we will work on event  $A_{(2)}$ . One can check that on event  $A_{(1)}$ , for sufficiently large  $N$ , and for  $i = 1, 2$ ,

$$\frac{1 - \delta^2}{2} \leq \frac{\tilde{X}_i(t_1)}{\tilde{X}_1(t_1) + \tilde{X}_2(t_1)} \leq \frac{1 + \delta^2}{2} \tag{6.31}$$

By the definition of the event  $A_{12}$  in (6.14), the definition of the function  $x$  in (6.11), and (6.31), for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$\begin{aligned} X_1(t_2)2N &\leq x_1(t_2) + \left(\frac{\delta^4}{2}\right)N \\ &= \left(\frac{\tilde{X}_1(t_1)}{\tilde{X}_1(t_1) + \tilde{X}_2(t_1)}\right)f(t_2) \cdot 2N + \left(\frac{\delta^4}{2}\right)N \\ &\leq \left(\frac{1 + \delta^2}{2}\right)(1 - \delta^2) \cdot 2N + \left(\frac{\delta^4}{2}\right)N \\ &= \left(1 - \frac{\delta^4}{2}\right)N, \end{aligned}$$

and

$$\begin{aligned} X_1(t_2) &\geq \left(\frac{1 - \delta^2}{2}\right)\left(\frac{1 - \delta^2}{1 + \delta^2}\right) \cdot 2N - \left(\frac{\delta^4}{2}\right)N \\ &\geq (1 - 3\delta^2)N. \end{aligned}$$

Both the upper and lower bounds for  $X_2(t_2)$  follow from the same argument.

Now, we proceed to prove statement (2). Assume that we are in the recombination dominating case. Recall the definition of  $Z_3^{[t_1]}(t)$  in (4.5) By the definition of  $A_{15}$  in (6.26), the inequality (6.28) and Proposition 2.2, on the event  $A_{(2)}$ ,

$$\begin{aligned} X_3^{[t_1]}(t_2) &\geq \left(X_3(t_1) - \sqrt{\frac{K_{0r}}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}}\right)e^{-r(t_2-t_1)} \\ &\geq \left(\frac{K_{1r}^- Nr \ln(Nr)}{s} - \sqrt{\frac{K_{0r}}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}}\right)e^{-(C_2+C_1) \cdot \frac{r}{s}} \\ &= \frac{Nr \ln(Nr)}{s} \cdot e^{-(C_2+C_1) \cdot \frac{r}{s}} \left(K_{1r}^- - \sqrt{\frac{K_{0r}}{\epsilon} \cdot \frac{1}{Nr \ln(Nr)}}\right). \end{aligned}$$

Because  $1 \ll Nr$  and  $r \ll s$ , there must be a positive constant  $K_{2r}^-$  such that for sufficiently large  $N$ ,

$$X_3^{[t_1]}(t_2) \geq \frac{K_{2r}^- Nr \ln(Nr)}{s}.$$

Hence, the result follows because  $X_3(t_2) \geq X_3^{[t_1]}(t_2)$

By the definitions of  $A_{13}, A_{14}$ , and  $A_{15}$  in (6.24), (6.25), and (6.26), and by Proposition 2.2, we have that for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$\begin{aligned} X_3(t_2) &= X_3^{[t_1]}(t_2) + X_{3m}^{(t_1, t_2]}(t_2) + X_{3r}^{(t_1, t_2]}(t_2) \\ &\leq \left( \frac{K_{1r}^+ Nr \ln(Nr)}{s} + \sqrt{\frac{K_{0r}}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}} \right) e^{2(C_2+C_1)} + \frac{K'_1}{\epsilon} \cdot \frac{N\mu}{s} + \frac{K'_2}{\epsilon} \cdot \frac{Nr}{s} \\ &= \frac{Nr \ln(Nr)}{s} \cdot \left[ \left( K_{1r}^+ + \sqrt{\frac{K_{0r}}{\epsilon} \cdot \frac{1}{Nr \ln(Nr)}} \right) e^{2(C_2+C_1)} + \frac{K'_1}{\epsilon} \cdot \frac{\mu}{r \ln(Nr)} \right. \\ &\quad \left. + \frac{K'_2}{\epsilon} \cdot \frac{1}{\ln(Nr)} \right]. \end{aligned}$$

Then, we define the constant

$$K_{2r}^+ = 2K_{1r}^+ e^{2(C_2+C_1)}. \tag{6.32}$$

Because  $1 \ll N\mu$  and  $\mu \ll r$ , for sufficiently large  $N$ ,

$$X_3(t_2) \leq \frac{K_{2r}^+ Nr \ln(Nr)}{s}.$$

Lastly, consider the mutation dominating case, where we will prove statement 3. First, by using (6.27) instead of (6.26), for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$\begin{aligned} X_3^{[t_1]}(t_2) &\geq \left( \frac{K_{1m}^- N^2 \mu^2}{s} - \sqrt{\frac{K_{0m}}{\epsilon} \cdot \frac{N\mu}{s}} \right) e^{-(C_2+C_1) \cdot \frac{r}{s}} \\ &= \frac{N^2 \mu^2}{s} \cdot e^{-(C_2+C_1) \cdot \frac{r}{s}} \left( K_{1m}^- - \sqrt{\frac{K_{0m}}{\epsilon} \cdot \frac{1}{N\mu}} \right). \end{aligned}$$

Since  $1 \ll N\mu$ , there must be a positive constant  $K_{2m}^-$  such that for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$X_3^{[t_1]}(t_2) \geq \frac{K_{2m}^- N^2 \mu^2}{s}.$$

Similar to the proof of part 2, for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$\begin{aligned} X_3(t_2) &\leq \left( \frac{K_{1m}^+ N^2 \mu^2}{s} + \sqrt{\frac{K_{0m}}{\epsilon} \cdot \frac{N\mu}{s}} \right) e^{2s(t_2-t_1)} + \frac{K'_1}{\epsilon} \cdot \frac{N\mu}{s} + \frac{K'_2}{\epsilon} \cdot \frac{Nr}{s} \\ &= \frac{N^2 \mu^2}{s} \cdot \left( \left( K_{1m}^+ + \sqrt{\frac{K_{0m}}{\epsilon} \cdot \frac{1}{N\mu}} \right) e^{2(C_2+C_1)} + \frac{K'_1}{\epsilon} \cdot \frac{1}{N\mu} + \frac{K'_2}{\epsilon} \cdot \frac{r}{N\mu^2} \right). \end{aligned}$$

Here, we pick the constant

$$K_{2m}^+ = 2K_{1m}^+ e^{2(C_2+C_1)}. \tag{6.33}$$

The upper bound for  $X_3(t_2)$  follows from the facts that  $1 \ll N\mu$  and  $r \ll N\mu^2$ .  $\square$

**7. Phase 3 and the proof of Proposition 2.4**

In this phase, the number of type 3 individuals experiences an exponential growth of rate about  $s$ . Note that the growth rate is approximately  $s$  because the majority of the population already has one beneficial allele. We will employ the technique of using expectation and variance bounds to prove results on type 3. However, unlike the previous two phases, which we kept tracking type 1 and type 2 individuals, we will track the number of type 0 individuals instead. Of course, the number of type 0 individuals is going to be small since type 0 is the least fit in the population, but we want to understand the size of type 0 population.

We begin by defining the constant

$$C_3 = \begin{cases} -C_2 + 3 + \ln\left(\frac{K_{2r}^+}{\delta^2}\right) & \text{in the recombination dominating case} \\ -C_2 + 3 + \ln\left(\frac{K_{2m}^+}{\delta^2}\right) & \text{in the mutation dominating case,} \end{cases} \tag{7.1}$$

where the constants  $K_{2r}^+$  and  $K_{2m}^+$  are defined in the equations (6.32) and (6.33). By tedious calculation, one can check that  $C_3$  is positive. Next, we define the time

$$t_3 = \begin{cases} \frac{1}{s} \ln\left(\frac{s^2}{\mu r \ln(Nr)}\right) - \frac{C_3}{s} & \text{in the recombination dominating case} \\ \frac{1}{s} \ln\left(\frac{s^2}{N\mu^3}\right) - \frac{C_3}{s} & \text{in the mutation dominating case.} \end{cases} \tag{7.2}$$

The time  $t_3$  is the time that the number of type 3 population reaches the order  $N$ . Now, we define the following stopping times:

$$T_4 = \inf\{t \geq t_2 : X_1(t) + X_2(t) \leq (2 - 6\delta)N\}, \tag{7.3}$$

$$T_5 = \inf\{t \geq t_2 : X_0(t) \geq 2\delta N e^{-s(1-3\delta)(t-t_2)}\}, \tag{7.4}$$

$$T_6 = \inf\left\{t \geq t_2 : s \int_{t_2}^t \tilde{X}_3(v)dv \geq 1\right\}, \tag{7.5}$$

$$T_{(3)} = T_4 \wedge T_5 \wedge T_6. \tag{7.6}$$

During the time interval  $[t_2, t_3]$ , most individuals in the population have one beneficial mutation. We define the stopping time  $T_4$  to stop the population process if the proportion of individuals with one beneficial mutation decrease below a certain constant threshold. Similarly, type 0 population is small during this time interval, and its size should experience exponential decay at a rate almost  $s$ . We stop the process as time  $T_5$  if the number of type 0 individuals exceeds the exponential decaying threshold. The factor  $1 - 3\delta$  in the exponent is for a technical purpose in the proof. The last stopping time  $T_6$  is also for a technical purpose; when we prove the expectation and variance bounds the term  $\int_{t_2}^t \tilde{X}_3(v)dv$  will frequently appear in the exponent.

Certainly, we are aiming to show that  $t_3 < T_{(3)}$  with probability close to 1, similar to what was done in phase 1. Hence, in both the mutation dominating case

and the recombination dominating case, we define the following events:

$$A_{16} = \{T_4 > t_3 \wedge T_{(3)}\}, \tag{7.7}$$

$$A_{17} = \{X_0^{[t_2]}(t \wedge T_{(3)}) < \delta N e^{-s(1-3\delta)(t \wedge T_{(3)} - t_2)}, \text{ for all } t \geq t_2\}, \tag{7.8}$$

$$A_{18} = \left\{ s \int_{t_2}^{t_3 \wedge T_{(3)}} \tilde{X}_3(v) dv < 1 \right\} = \{T_6 > t_3 \wedge T_{(3)}\}. \tag{7.9}$$

Lastly, we define events that are related to the expectation and the variance bounds, similar to what we did in phase 1. However, the sizes of the type 0 and type 3 populations depend on the sizes at time  $t_2$ , which depend on the cases. Thus, the events are defined differently between the recombination dominating case and the mutation dominating case. In the recombination dominating case, we define the following events:

$$A_{19} = \left\{ X_{0r}^{(t_2, t_3)}(t \wedge T_{(3)}) \leq \frac{2e^{3-(C_3+C_2)}}{\epsilon} \cdot \frac{N}{\ln(Nr)} \cdot e^{-s(1-3\delta)(t \wedge T_{(3)} - t_2)}, \right. \\ \left. \text{for all } t \in [t_2, t_3] \right\} \tag{7.10}$$

$$A_{20} = \left\{ X_{3m}^{(t_2, t_3)}(t_3 \wedge T_{(3)}) < \left( \frac{2\delta^2}{\epsilon K_{2r}^+} \right) \cdot \frac{N\mu}{r \ln(Nr)} \right\} \tag{7.11}$$

$$A_{21} = \left\{ X_{3r}^{(t_2, t_3)}(t_3 \wedge T_{(3)}) \leq \left( \frac{2\delta^2}{\epsilon K_{2r}^+} \right) \cdot \frac{N}{\ln(Nr)} \right\} \tag{7.12}$$

$$A_{22} = \left\{ \sup_{t \in [t_2, t_3]} \left| Z_3^{[t_2]}(t \wedge T_{(3)}) - X_3(t_2) \right| < \sqrt{\frac{2e^4 K_{2r}^+}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}} \right\} \tag{7.13}$$

In contrast, in the mutation dominating case, we define

$$A_{19} = \left\{ X_{0r}^{(t_2, t_3)}(t \wedge T_{(3)}) \leq \frac{2e^{3-(C_3+C_2)}}{\epsilon} \cdot \frac{r}{\mu^2} \cdot e^{-s(1-3\delta)(t \wedge T_{(3)} - t_2)}, \right. \\ \left. \text{for all } t \in [t_2, t_3] \right\},$$

$$A_{20} = \left\{ X_{3m}^{(t_2, t_3)}(t_3 \wedge T_{(3)}) < \left( \frac{\delta^2}{\epsilon K_{2m}^+} \right) \cdot \frac{1}{\mu} \right\}, \tag{7.14}$$

$$A_{21} = \left\{ X_{3r}^{(t_2, t_3)}(t_3 \wedge T_{(3)}) \leq \left( \frac{\delta^2}{\epsilon K_{2m}^+} \right) \cdot \frac{r}{\mu^2} \right\}, \tag{7.15}$$

$$A_{22} = \left\{ \sup_{t \in [t_2, t_3]} \left| Z_3^{[t_2]}(t \wedge T_{(3)}) - X_3(t_2) \right| < \sqrt{\frac{2e^4 K_{2m}^+}{\epsilon} \cdot \frac{N\mu}{s}} \right\}.$$

**7.1. Results on type 0 individuals.** In this subsection, we split type 0 individuals into two groups: one for those that descend from type 0 individuals at time  $t_2$  and the other for those that descend from ancestors that appear in the time interval  $(t_2, t_3]$ . We will show that with probability almost 1, the events  $A_{17}$  and  $A_{19}$  occur, in which the numbers of individuals in these two groups decay at least exponentially fast. The proofs can be done by using the Doob's maximal inequality and the fact that  $(W_{0r}^{(t_2, t_3)}(t), t \geq t_2)$  is a submartingale.

First, we give bounds on the growth rates of type 0 and type 3 populations. The essence of the following lemma is that the growth rate of type 0 is about  $-s$ , and the growth rate of type 3 is about  $s$ . Also, note that the growth rate of type 0 is negative.

**Lemma 7.1.** *The following statements hold.*

- (1) If  $t \in [t_2, T_4)$ , then  $G_0(t) \leq -s(1 - 3\delta)$ .
- (2) If  $t \in [t_2, T_4)$ , then  $-s(1 + \tilde{X}_3(t)) - r - 2\mu \leq G_{0r}^{(t_2, t_3]}(t) \leq -s(1 - 3\delta) + r$ .
- (3) If  $t \in [t_2, T_5)$ , then  $s(1 - \tilde{X}_3(t)) - r \leq G_3(t) \leq s(1 + \delta e^{-s(1-3\delta)(t-t_2)})$ .
- (4) If  $t \in [t_2, T_5)$ , then  $s(1 - \tilde{X}_3(t)) - r \leq G_{3r}^{(t_2, t_3]}(t) \leq s(1 + \delta e^{-s(1-3\delta)(t-t_2)}) + r$ .

*Proof:* By the definition of  $T_4$  in (7.3), if  $t \in [t_2, T_4)$ , then  $\tilde{X}_1(t) + \tilde{X}_2(t) > 1 - 3\delta$ . Then, from (3.26), we have that  $G_0(t) \leq -s(\tilde{X}_1(t) + \tilde{X}_2(t)) < -s(1 - 3\delta)$ . From (3.27), if  $t \in [t_2, T_4)$ , then  $G_{0r}^{(t_2, t_3]}(t) \leq -s(1 - 3\delta) + r$ . Also, from the fact that  $\tilde{X}_1(u) + \tilde{X}_2(u) + \tilde{X}_3(u) \leq 1$ , we get  $G_{0r}^{(t_2, t_3]}(t) \geq -s(1 + \tilde{X}_3(t)) - r - 2\mu$ .

Now, from the definition of  $T_5$  in (7.4), if  $t \in [t_2, T_5)$ , then the equation (3.19) implies statement (3). Statement (4) follows directly from statement (3) and (3.22). □

We will now proceed to prove the result on the number of type 0 individuals.

**Lemma 7.2.** *For sufficiently large  $N$ , on the event  $A_{(2)}$ , we have  $P(A_{17}^c | \mathcal{F}_{t_2}) \leq 6\delta$ .*

*Proof:* First, from part 2 of Proposition 2.3, on the event  $A_{(2)}$ , we have that  $X_0(t_2) \leq 2N - X_1(t_2) - X_2(t_2) \leq 6\delta^2 N$ . From Proposition 4.3, the process  $(Z_0^{[t_2]}(t \wedge T_{(3)}), t \geq t_2)$  is a nonnegative martingale. Hence, by Lemma 7.1 and Doob’s maximal inequality, for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$\begin{aligned} P(A_{17}^c | \mathcal{F}_{t_2}) &\leq P\left(\sup_{t \geq t_2} Z_0^{[t_2]}(t \wedge T_{(3)}) \geq \delta N \mid \mathcal{F}_{t_2}\right) \\ &\leq \frac{E\left[Z_0^{[t_2]}(t_2) \mid \mathcal{F}_{t_2}\right]}{\delta N / 2} \\ &= \frac{X_0(t_2)}{\delta N} \\ &\leq 6\delta, \end{aligned}$$

which proves this lemma. □

**Lemma 7.3.** *For sufficiently large  $N$ , we have  $P(A_{19}^c | \mathcal{F}_{t_2}) \leq \epsilon$ .*

*Proof:* We first prove this result in the recombination dominating case. By Proposition 4.4, the process  $(W_{0r}^{(t_2, t_3]}(t \wedge T_{(3)}), t \geq t_2)$  is a submartingale. Also, note that from the definitions of  $t_2$  and  $t_3$  in (6.13) and (7.2), we have that

$$t_3 - t_2 = \frac{1}{s} \ln\left(\frac{s}{r \ln(Nr)}\right) - \frac{C_3 + C_2}{s}. \tag{7.16}$$

From Proposition 4.4, Lemma 7.1 part 2, and (3.15), we have

$$E\left[W_{0r}^{(t_2, t_3]}(t_3 \wedge T_{(3)}) \mid \mathcal{F}_{t_2}\right] \leq E\left[\int_{t_2}^{t_3 \wedge T_{(3)}} 2Nr e^{\int_{t_2}^u s(1 + \tilde{X}_3(v)) + r + 2\mu dv} du \mid \mathcal{F}_{t_2}\right]$$

By the definition of  $T_6$  in (7.5),

$$\begin{aligned} E\left[W_{0r}^{(t_2, t_3]}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] &\leq 2e^{1+(r+2\mu)(t_3-t_2)} Nr \int_{t_2}^{t_3} e^{s(u-t_2)} du \\ &\leq 2e^{1+(r+2\mu)(t_3-t_2)} \cdot e^{-(C_3+C_2)} \cdot \frac{N}{\ln(Nr)}. \end{aligned} \quad (7.17)$$

Because  $1 \ll Nr$  and  $r \ll s$ , for sufficiently large  $N$ ,

$$\frac{r}{s} \ln\left(\frac{s}{r \ln(Nr)}\right) \leq \frac{r}{s} \ln\left(\frac{s}{r}\right) \ll 1.$$

It follows that

$$r(t_3 - t_2) \ll 1. \quad (7.18)$$

Also, since  $\mu \ll r$ , we have

$$\mu(t_3 - t_2) \ll 1. \quad (7.19)$$

Hence, from (7.17), for sufficiently large  $N$ , we have

$$E\left[W_{0r}^{(t_2, t_3]}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq 2e^{2+(-C_3+C_2)} \cdot \frac{N}{\ln(Nr)}.$$

Thus, from (7.18), Lemma 7.1 part 2 and Doob's maximal inequality, for sufficiently large  $N$ ,

$$\begin{aligned} &P(A_{19}^c | \mathcal{F}_{t_2}) \\ &\leq P\left(\sup_{t \in [t_2, t_3]} X_{0r}^{(t_2, t_3]}(t \wedge T_{(3)}) e^{s(1-3\delta)(t \wedge T_{(3)} - t_2) - r(t_3 - t_2)} \geq \frac{2e^{2-(C_3+C_2)}}{\epsilon} \cdot \frac{N}{\ln(Nr)} \middle| \mathcal{F}_{t_2}\right) \\ &\leq P\left(\sup_{t \in [t_2, t_3]} W_{0r}^{(t_2, t_3]}(t \wedge T_{(3)}) \geq \frac{e^{2-(C_3+C_2)}}{\epsilon} \cdot \frac{N}{\ln(Nr)} \middle| \mathcal{F}_{t_2}\right) \\ &\leq \epsilon. \end{aligned}$$

Now, for the mutation dominating case, we observe that from the definitions of  $t_2$  and  $t_3$  in (6.13) and (7.2), we have

$$t_3 - t_2 = \frac{1}{s} \ln\left(\frac{s}{N\mu^2}\right) + \frac{C_3 - C_2}{s}. \quad (7.20)$$

One can check that (7.18) and (7.19) also hold in this case. By following the same argument as in the recombination dominating case, we obtain that for sufficiently large  $N$ ,

$$E\left[W_{0r}^{(t_2, t_3]}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq 2e^{2-(C_3+C_2)} \cdot \frac{r}{\mu^2},$$

which implies that  $P(A_{19}^c | \mathcal{F}_{t_2}) \leq \epsilon$ .  $\square$

**7.2. Results on type 3 individuals.** In this subsection, we show results on type 3 using the expectation and variance technique that we employ in the proof of phase 1. Since the number of type 3 individuals at time  $t_2$  are of different orders in the mutation dominating case and the recombination dominating case, the proofs will be split between these two cases.

**Lemma 7.4.** *For sufficiently large  $N$ , we have that for  $t \in [t_2, t_3]$ ,*

$$E\left[X_{3m}^{(t_2, t_3]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq \frac{e^3 N \mu}{s} \cdot e^{s(t-t_2)},$$

and  $P(A_{20}^c | \mathcal{F}_{t_2}) \leq \epsilon$ .

*Proof:* The proof is similar to that of Lemma 5.4. First, by Proposition 4.1, recall that the process  $(Z_{3m}^{(t_2, t_3]}(t \wedge T_{(3)}), t \geq t_2)$  is a mean-zero martingale. From (3.11), Lemma 7.1 part 3, and the definition of  $T_6$  in (7.5), for every  $t \in [t_2, t_3]$ ,

$$\begin{aligned} E\left[\int_{t_2}^{t \wedge T_{(3)}} M_3^{(t_2, t_3]}(u) e^{-\int_{t_2}^u G_3(v) dv} du \middle| \mathcal{F}_{t_2}\right] &\leq e^{1+r(t_3-t_2)} \cdot 2N\mu \cdot \int_{t_2}^t e^{-s(u-t_2)} du \\ &\leq 2e^{1+r(t_3-t_2)} \cdot \frac{N\mu}{s}. \end{aligned} \quad (7.21)$$

From (7.18), for sufficiently large  $N$  and for all  $t \in [t_2, t_3]$ ,

$$E\left[\int_{t_2}^{t \wedge T_{(3)}} M_3^{(t_2, t_3]}(u) e^{-\int_{t_2}^u G_3(v) dv} du \middle| \mathcal{F}_{t_2}\right] \leq \frac{2e^2 N \mu}{s}.$$

Also, by Lemma 7.1 part 3, for all  $t \geq t_2$ ,

$$\begin{aligned} E\left[e^{-\int_{t_2}^{t \wedge T_{(3)}} G_3(v) dv} X_{3m}^{(t_2, t_3]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \\ \geq e^{-\int_{t_2}^t s(1+\delta e^{-s(1-3\delta)(v-t_2)}) dv} \cdot E\left[X_{3m}^{(t_2, t_3]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \\ \geq e^{-s(t-t_2) - \frac{\delta}{1-3\delta}} \cdot E\left[X_{3m}^{(t_2, t_3, r]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right]. \end{aligned}$$

Using  $\delta < \frac{1}{4}$ , we get the expectation bound. By using Markov's inequality, we have that  $P(A_{20}^c | \mathcal{F}_{t_2}) \leq \epsilon$ .  $\square$

**Lemma 7.5.** *For sufficiently large  $N$ , we have that for  $t \in [t_2, t_3]$ ,*

$$E\left[X_{3r}^{(t_2, t_3]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq \frac{e^3 N r}{s} \cdot e^{s(t-t_2)},$$

and  $P(A_{21}^c | \mathcal{F}_{t_2}) \leq \epsilon$ .

*Proof:* The proof is almost exactly the same as that of Lemma 7.4. Note that in this proof, we have to consider the term  $R_3^{(t_2, t_3]}(u)$  instead of  $M_3^{(t_2, t_3]}(u)$ . From (3.17), we have that  $R_3^{(t_2, t_3]}(u) \leq 2Nr$  for all  $u \geq t_2$ . The rest of the prove is following the argument in Lemma 7.4.  $\square$

Next, we will show that the events  $A_{16}$ ,  $A_{18}$ , and  $A_{22}$  occur with high probability. To prove this, we will need an upper bound for the term  $E[X_3^{[t_2]}(t \wedge T_{(3)}) | \mathcal{F}_{t_2}]$ , which is proved in the following lemma.

**Lemma 7.6.** *For sufficiently large  $N$ , if  $t \geq t_2$ , then the following statements hold on the event  $A_{(2)}$ .*

(1) *In the recombination dominating case,*

$$E\left[X_3^{[t_2]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq \frac{eK_{2r}^+ N r \ln(Nr)}{s} \cdot e^{s(t-t_2)}.$$

(2) *In the mutation dominating case,*

$$E\left[X_3^{[t_2]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq \frac{eK_{2m}^+ N^2 \mu^2}{s} \cdot e^{s(t-t_2)}.$$

*Proof:* From Proposition 4.3, we know that  $(Z_3^{[t_2]}(t \wedge T_{(3)}), t \geq t_2)$  is a martingale. So, from (4.5), Lemma 7.1 part 3, and the fact that  $\delta < \frac{1}{4}$ , for all  $t \geq t_2$ ,

$$\begin{aligned} E\left[Z_3^{[t_2]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] &\geq E\left[X_3^{[t_2]}(t \wedge T_{(3)}) e^{-\int_{t_2}^{t \wedge T_{(3)}} s(1+\delta e^{-s(1-3\delta)(v-t_2)}) dv} \middle| \mathcal{F}_{t_2}\right] \\ &\geq e^{-s(t-t_2)-1} E\left[X_3^{[t_2]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right]. \end{aligned}$$

Therefore, for all  $t \geq t_2$ ,

$$E\left[X_3^{[t_2]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \leq e^{s(t-t_2)+1} E\left[Z_3^{[t_2]}(t \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] = e^{s(t-t_2)+1} X_3(t_2).$$

Thus, the result follows from the upper bound for  $X_3(t_2)$  on the event  $A_{(2)}$  in Proposition 2.3.  $\square$

**Lemma 7.7.** *For sufficiently large  $N$ , on the event  $A_{(2)}$ , we have  $P(A_{16}^c | \mathcal{F}_{t_2}) \leq \delta$ .*

*Proof:* First, by the definition of  $T_5$  in (7.4),  $X_0(t_3 \wedge T_{(3)}) \leq 2\delta N + 1 < 3\delta N$  when  $N$  is sufficiently large. It follows from this inequality and Markov's inequality that for sufficiently large  $N$ , on the event  $A_{(2)}$ ,

$$\begin{aligned} P(T_4 = t_3 \wedge T_{(3)} | \mathcal{F}_{t_2}) &\leq P(X_0(t_3 \wedge T_{(3)}) \geq 3\delta N | \mathcal{F}_{t_2}) + P(X_3(t_3 \wedge T_{(3)}) \geq 3\delta N | \mathcal{F}_{t_2}) \\ &\leq E\left[X_3(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \cdot \frac{1}{3\delta N}. \end{aligned}$$

By Lemma 7.4 and Lemma 7.5, we have

$$\begin{aligned} P(T_4 = t_3 \wedge T_{(3)} | \mathcal{F}_{t_2}) &\leq \left( E\left[X_3^{[t_2]}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] + E\left[X_{3m}^{(t_2, t_3)}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \right. \\ &\quad \left. + E\left[X_{3r}^{(t_2, t_3)}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] \right) \cdot \frac{1}{3\delta N} \\ &\leq \left( E\left[X_3^{[t_2]}(t_3 \wedge T_{(3)}) \middle| \mathcal{F}_{t_2}\right] + \frac{2e^3 N \mu}{s} \cdot e^{s(t_3-t_2)} + \frac{2e^3 N r}{s} \cdot e^{s(t_3-t_2)} \right) \cdot \frac{1}{3\delta N} \end{aligned} \tag{7.22}$$

At this point, the calculation splits between the two cases. In the recombination dominating case, by (7.22), (7.16) and Lemma 7.6,

$$\begin{aligned} P(T_4 = t_3 \wedge T_{(3)} | \mathcal{F}_{t_2}) &\leq \left( \frac{eK_{2r}^+ N r \ln(Nr)}{s} \cdot e^{s(t_3-t_2)} + \frac{e^3 N \mu}{s} \cdot e^{s(t_3-t_2)} + \frac{e^3 N r}{s} \cdot e^{s(t_3-t_2)} \right) \cdot \frac{1}{3\delta N} \\ &= \frac{e^{-2\delta}}{3} + \frac{2e^{3-(C_3+C_2)}}{3\delta} \cdot \left( \frac{\mu}{r \ln(Nr)} + \frac{1}{\ln(Nr)} \right). \end{aligned}$$

The result follows from  $1 \ll Nr$  and  $\mu \ll r \ln(Nr)$ ,

The proof for the mutation dominating case is almost the same, except at (7.22), where Lemma 7.6 gives

$$\begin{aligned} &P(T_4 = t_3 \wedge T_{(3)} | \mathcal{F}_{t_2}) \\ &\leq \left( \frac{eK_{2m}^+ N^2 \mu^2}{s} \cdot e^{s(t_3-t_2)} + \frac{2e^3 N \mu}{s} \cdot e^{s(t_3-t_2)} + \frac{2e^3 N r}{s} \cdot e^{s(t_3-t_2)} \right) \cdot \frac{1}{3\delta N} \\ &= \frac{e^{-2}\delta}{3} + \frac{2e^{3-(C_3+C_2)}}{3\delta} \cdot \left( \frac{1}{N\mu} + \frac{r}{N\mu^2} \right). \end{aligned}$$

Then, the result follows from the facts that  $1 \ll N\mu$  and  $r \ll N\mu^2$ . □

**Lemma 7.8.** *For sufficiently large  $N$ , on the event  $A_{(2)}$ , we have  $P(A_{18}^c | \mathcal{F}_{t_2}) \leq \delta^2$ .*

*Proof:* In the recombination dominating case, from Lemmas 7.4, 7.5 and 7.6, we have

$$\begin{aligned} &E \left[ s \int_{t_2}^{t_3 \wedge T_{(3)}} \tilde{X}_3(v) dv \mid \mathcal{F}_{t_2} \right] \\ &\leq \int_{t_2}^{t_3} s \left( E \left[ \tilde{X}_3^{[t_2]}(v \wedge T_{(3)}) \mid \mathcal{F}_{t_2} \right] + E \left[ \tilde{X}_{3m}^{(t_2, t_3]}(v \wedge T_{(3)}) \mid \mathcal{F}_{t_2} \right] \right. \\ &\quad \left. + E \left[ \tilde{X}_{3r}^{(t_2, t_3]}(v \wedge T_{(3)}) \mid \mathcal{F}_{t_2} \right] \right) dv \\ &\leq \int_{t_2}^{t_3} \left( \frac{eK_{2r}^+}{2} r \ln(Nr) \cdot e^{s(v-t_2)} + e^3 \cdot \mu \cdot e^{s(v-t_2)} + e^3 \cdot r \cdot e^{s(v-t_2)} \right) dv. \quad (7.23) \end{aligned}$$

Because  $\mu \ll r$  and  $1 \ll Nr$ , along with the definition of  $C_3$  in (7.1), for sufficiently large  $N$ , on the event  $A_{(2)}$ , we have

$$\begin{aligned} E \left[ s \int_{t_2}^{t_3 \wedge T_{(3)}} \tilde{X}_3(v) dv \mid \mathcal{F}_{t_2} \right] &\leq \int_{t_2}^{t_3} e^3 K_{2r}^+ r \ln(Nr) \cdot e^{s(v-t_2)} dv \\ &\leq \frac{e^3 K_{2r}^+ r \ln(Nr)}{s} \cdot e^{s(t_3-t_2)} \\ &= \delta^2. \end{aligned}$$

Thus, by Markov's inequality, we have  $P(A_{18}^c) \leq \delta^2$ . For the mutation dominating case, the term  $K_{2r}^+ r \ln(Nr)$  in (7.23) is replaced by  $K_{2m}^+ N\mu^2$ . Then, the result follows by a similar argument. □

**Lemma 7.9.** *For sufficiently large  $N$ , on the event  $A_{(2)}$ , we have  $P(A_{22}^c | \mathcal{F}_{t_2}) \leq \epsilon$ .*

*Proof:* We first consider the recombination dominating case. From Proposition 4.3, part 3 of Lemma 7.1, Lemma 7.6, and (7.18), for sufficiently large  $N$ , on the event

$A_{(2)}$ , we have that

$$\begin{aligned} & \text{Var}\left(Z_3^{[t_2]}(t_3 \wedge T_{(3)}) \mid \mathcal{F}_{t_2}\right) \\ & \leq E\left[\int_{t_2}^{t_3 \wedge T_{(3)}} e^{-2\int_{t_2}^u (s(1-\tilde{X}_3(v))-r)dv} \cdot 2X_3^{[t_2]}(u \wedge T_{(3)})du \mid \mathcal{F}_{t_2}\right] \\ & \leq \int_{t_2}^{t_3} e^{-2s(u-t_2)+2+2r(t_3-t_2)} \cdot 2E\left[X_3^{[t_2]}(u \wedge T_{(3)}) \mid \mathcal{F}_{t_2}\right] du \\ & \leq 2e^{2+2r(t_3-t_2)} \cdot \int_{t_2}^{t_3} e^{-2s(u-t_2)} \cdot \frac{eK_{2r}^+Nr \ln(Nr)}{s} \cdot e^{s(u-t_2)} du \\ & \leq \frac{2e^4 K_{2r}^+ Nr \ln(Nr)}{s^2}. \end{aligned}$$

The result follows from Doob’s maximal inequality. By a similar argument, we can prove the result in the mutation dominating case.  $\square$

7.3. *The proof of Proposition 2.4.* In the previous section, we show that each of the events  $A_{16}$  to  $A_{21}$  conditioned on  $\mathcal{F}_{t_2}$  occurs with probability close to 1 on the event  $A_{(2)}$ . Here, in both the mutation dominating case and the recombination dominating case, we define

$$A_{(3)} = A_{(2)} \cap \left(\bigcap_{i=16}^{22} A_i\right).$$

In the next step, before we eventually prove Proposition 2.4, we show that  $T_{(3)} > t_3$  on event  $A_{(3)}$ .

**Lemma 7.10.** *For sufficiently large  $N$ , on the event  $A_{(3)}$ , we have that  $T_{(3)} > t_3$ .*

*Proof:* In this proof, we are working on the event  $A_{(3)}$ . Since  $A_{(3)} \subseteq A_{16} \cap A_{18}$ , we have  $T_4 > t_3 \wedge T_{(3)}$  and  $T_6 > t_3 \wedge T_{(3)}$ . It is left to show that  $T_5 > t_3 \wedge T_{(3)}$ .

In the recombination dominating case, by the definitions of the events  $A_{17}$  and  $A_{19}$  in (7.8) and (7.10), if  $t \in [t_2, t_3]$ , then

$$\begin{aligned} X_0(t \wedge T_{(3)}) &= X_0^{[t_2]}(t \wedge T_{(3)}) + X_{0r}^{(t_2, t_3]}(t \wedge T_{(3)}) \\ &\leq \left(\delta + \frac{2e^{3-(C_3+C_2)}}{\epsilon} \cdot \frac{1}{\ln(Nr)}\right) \cdot Ne^{-s(1-3\delta)(t \wedge T_{(3)}-t_2)}. \end{aligned}$$

Since  $1 \ll Nr$ , for sufficiently large  $N$ , we have that

$$X_0(t \wedge T_{(3)}) < \delta Ne^{-s(1-3\delta)(t \wedge T_{(3)}-t_2)}$$

for all  $t \in (t_2, t_3]$ . Therefore,  $T_6 > t_3 \wedge T_{(3)}$ . Similarly, in the mutation dominating case, for all  $t \in [t_2, t_3]$ ,

$$X_0(t \wedge T_{(3)}) \leq \left(\delta + \frac{2e^{3-(C_3+C_2)}}{\epsilon} \cdot \frac{r}{N\mu^2}\right) \cdot Ne^{-s(1-3\delta)(t \wedge T_{(3)}-t_2)}.$$

The result follows from  $r \ll N\mu^2$ .  $\square$

At last, we prove Proposition 2.4.

*Proof:* From Lemmas 7.2, 7.3, 7.4, 7.5, 7.8, 7.9, and 7.7, for sufficiently large  $N$ , on the event  $A_{(2)}$ , we have  $P(\bigcap_{i=16}^{22} A_i \mid \mathcal{F}_{t_2}) \geq 1 - 4\epsilon - 7\delta - \delta^2$ . Thus, by Proposition 2.3, we have  $P(A_{(3)}) \geq 1 - 25\epsilon - 7\delta - \delta^2$  for sufficiently large  $N$ .

Next, assume that we are on the event  $A_{(3)}$  and recall that  $T_{(3)} > t_3$  when  $N$  is sufficiently large. So, by the definition of  $T_5$  in (7.4),

$$X_0(t_3) < 2\delta N e^{-s(1-3\delta)(t_3-t_2)}.$$

Then, by using the definition of  $t_3$  in (7.2), we prove the first part of the proposition.

For the proof of the second part of this proposition, we define

$$K_3 = \begin{cases} \frac{K_{2r}^- e^{-(C_3+C_2)-2}}{2} & \text{in the recombination dominating case} \\ \frac{K_{2m}^- e^{-(C_3+C_2)-2}}{2} & \text{in the mutation dominating case.} \end{cases} \tag{7.24}$$

We will first consider the recombination dominating case. From (4.5), the definition of the event  $A_{22}$  in (7.13), and Proposition 2.3, for sufficiently large  $N$ ,

$$\begin{aligned} X_3^{[t_2]}(t_3) &\geq \left( X_3^{[t_2]}(t_2) - \sqrt{\frac{2e^4 K_{2r}^+}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}} \right) e^{\int_{t_2}^{t_3} G_3(v) dv} \\ &\geq \left( \frac{K_{2r}^- Nr \ln(Nr)}{s} - \sqrt{\frac{2e^4 K_{2r}^+}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}} \right) e^{\int_{t_2}^{t_3} G_3(v) dv} \\ &\geq \frac{K_{2r}^-}{2} \cdot \frac{Nr \ln(Nr)}{s} \cdot e^{\int_{t_2}^{t_3} G_3(v) dv}. \end{aligned} \tag{7.25}$$

Hence, from Lemma 7.1, the definition of  $T_5$  in (7.4), inequality (7.18), and the definition of  $K_3$  in (7.24), for sufficiently large  $N$ , we have that

$$\begin{aligned} X_3(t_3) &\geq \frac{K_{2r}^-}{2} \cdot \frac{Nr \ln(Nr)}{s} \cdot e^{\int_{t_2}^{t_3} (s(1-\tilde{X}_3(v)) - r) dv} \\ &\geq \frac{K_{2r}^-}{2} \cdot \frac{Nr \ln(Nr)}{s} \cdot e^{s(t_3-t_2)-2} \\ &= K_3 N. \end{aligned}$$

For the upper bound for  $X_3(t_3)$ , from (4.5), the definition of the event  $A_{22}$  in (7.13), Proposition 2.3, the fact that  $\delta < \frac{1}{4}$ , and the definitions of  $C_3$  in (7.1), we have

$$\begin{aligned} X_3^{[t_2]}(t_3) &\leq \left( \frac{K_{2r}^+ Nr \ln(Nr)}{s} + \sqrt{\frac{2e^4 K_{2r}^+}{\epsilon} \cdot \frac{Nr \ln(Nr)}{s^2}} \right) e^{s(t_3-t_2)+1} \\ &= \left( e^{-2\delta^2} + e^{-(C_3+C_2)+1} \cdot \sqrt{\frac{2e^4 K_{2r}^+}{\epsilon} \cdot \frac{1}{Nr \ln(Nr)}} \right) N. \end{aligned} \tag{7.26}$$

Since  $1 \ll Nr$ , for sufficiently large  $N$ , we have  $X_3^{[t_2]}(t_3) \leq \frac{\delta^2 N}{3}$ . It follows from the definitions of the events  $A_{20}$  and  $A_{21}$  as defined in (7.11) and (7.12) that for sufficiently large  $N$ , we have  $X_{3m}^{(t_2, t_3]}(t_3) \leq \frac{\delta^2 N}{3}$  and  $X_{3r}^{(t_2, t_3]}(t_3) \leq \frac{\delta^2 N}{3}$ . Therefore,  $X_3(t_3) \leq \delta^2 N$  for sufficiently large  $N$ .

We will now consider the mutation dominating case. Due to the differences in the definition of  $A_{22}$  and the lower bound of  $X_3^{[t_2]}(t_3)$  from Proposition 2.3, instead of having inequality (7.25), we will have

$$\begin{aligned} X_3(t_3) &\geq \left( \frac{K_{2m}^- N^2 \mu^2}{s} - \sqrt{\frac{2e^4 K_{2m}^+}{\epsilon}} \cdot \frac{N\mu}{s} \right) e^{\int_{t_2}^{t_3} G_3(v) dv} \\ &= \frac{K_{2m}^-}{2} \cdot \frac{N^2 \mu^2}{s} \cdot e^{\int_{t_2}^{t_3} G_3(v) dv}. \end{aligned}$$

By following the argument in the previous case, we have that  $X_3(t_3) \geq K_3 N$ . For the upper bound for  $X_3(t_2)$ , instead of having inequality (7.26), we will have

$$\begin{aligned} X_3^{[t_2]}(t_3) &\leq \left( \frac{K_{2m}^+ N^2 \mu^2}{s} + \sqrt{\frac{8e^4 K_{2m}^+}{\epsilon}} \cdot \frac{N\mu}{s} \right) e^{s(t_3-t_2) + \frac{\delta}{1-3\delta}} \\ &\leq \left( e^{-2\delta^2} + e^{(C_3-C_2)+1} \cdot \sqrt{\frac{8e^4 K_{2m}^+}{\epsilon}} \cdot \frac{1}{N\mu} \right) N, \end{aligned}$$

and the upper bound of  $X_3(t_3)$  follows by a similar argument in the previous case. □

**8. Phase 4 and the proof of Proposition 2.5**

In this phase, the number of type 3 individuals experience logistic growth. This fact can be proved by using Theorem 6.1 as we did in phase 2. Then, we proceed to prove Proposition 2.5, which implies that at time  $t_4$ , most individuals have become type 3 while the number of individuals with only one beneficial allele remains a small fraction of the population.

To begin the proof, first recall the definitions of  $\mathbf{X}(t)$ ,  $q$ ,  $\alpha$ ,  $\beta$ ,  $b$  and  $\tilde{b}$  as in (6.3), (6.4), (6.1), (6.6), (6.7) and (6.8), respectively. Next, we define a random variable  $B^*$  such that on the event that  $\tilde{X}_3(t_3) > 0$ , we have

$$B^* = \frac{1}{\tilde{X}_3(t_3)} - 1. \tag{8.1}$$

The definition of  $B^*$  when  $\tilde{X}_3(t_3) = 0$  is not of interest, as we will work only on the event  $A_{(3)}$ , on which from Proposition 2.4, we know that  $\tilde{X}_3(t_3) > 0$ . Next, for  $t \geq t_3$ , we define

$$f^*(t) = \frac{1}{1 + B^* e^{-s(t-t_3)}}. \tag{8.2}$$

and define

$$\begin{aligned} x^*(t) &= (x_1^*(t), x_2^*(t), x_3^*(t)) \\ &= f^*(t) \left( \left( \frac{1 - \tilde{X}_2(t_3) - \tilde{X}_3(t_3)}{\tilde{X}_3(t_3)} \right) e^{-s(t-t_3)}, \left( \frac{\tilde{X}_2(t_3)}{\tilde{X}_3(t_3)} \right) e^{-s(t-t_3)}, 1 \right). \end{aligned} \tag{8.3}$$

One can check that  $(x^*)'(t) = b(x^*(t))$  when  $t \geq t_3$ .

Lastly, we define

$$c_4 = -C_3 + \ln \left( \left( \frac{1}{\delta^2} - 1 \right) \left( \frac{1}{K_3} - 1 \right) \right), \tag{8.4}$$

and

$$t_4 = \begin{cases} \frac{1}{s} \ln \left( \frac{s^2}{\mu r \ln(Nr)} \right) + \frac{c_4}{s} & \text{in the recombination dominating case} \\ \frac{1}{s} \ln \left( \frac{s^2}{N\mu^3} \right) + \frac{c_4}{s} & \text{in the mutation dominating case,} \end{cases} \tag{8.5}$$

where  $K_3$  is a positive constant that was defined in (7.24). Here, we do not specify whether  $c_4$  is positive or negative; this will not affect the proof.

**Lemma 8.1.** *For sufficiently large  $N$ , on the event  $A_{(3)}$ , we have*

$$P \left( \sup_{t \in [t_3, t_4]} |X_i(t) - 2Nx_i^*(t)| \leq \frac{K_3^2 N}{2\delta^2} \text{ for } i = 1, 2, 3 \mid \mathcal{F}_{t_3} \right) \geq 1 - \epsilon.$$

*Proof:* The proof is almost exactly the same as the proof of Lemma 6.2. Recall from section 6 that  $k$  is a constant not depending on  $N$  such that  $ks$  is a Lipschitz constant of the function  $b$ . We define  $\Delta^* = K_3^2 e^{-k(c_4 + C_3)} / (12\delta^2)$ , and  $L = 48/N$ . We also define

$$\begin{aligned} \Omega_0^* &= \{ |\mathbf{X}(t_3) - x^*(t_3)| \leq \Delta^* \} \\ \Omega_1^* &= \left\{ \int_{t_3}^{t_4} |\beta(\mathbf{X}(t)) - b(\mathbf{X}(t))| dt \leq \Delta^* \right\} \\ \Omega_2^* &= \left\{ \int_{t_3}^{t_4} \alpha(\mathbf{X}(t)) dt \leq L(t_4 - t_3) \right\}. \end{aligned}$$

The result follows by the same argument in Lemma 6.2. □

Now, we proceed to prove Proposition 2.5.

*Proof of Proposition 2.5:* First, we define the event

$$A_{(4)} = A_{(3)} \cap \left\{ \sup_{t \in [t_3, t_4]} |X_i(t) - 2Nx_i^*(t)| \leq \frac{K_3^2 N}{2\delta^2} \text{ for } i = 1, 2, 3 \right\} \tag{8.6}$$

By Propositions 2.4 and Lemma 8.1, for sufficiently large  $N$ , we have

$$P(A_{(4)}) \geq 1 - \epsilon - P(A_{(3)}^c) \geq 1 - 26\epsilon - 7\delta - \delta^2.$$

From this point, we will work on the event  $A_{(4)}$ . From the definition of  $B^*$  in (8.1) and Proposition 2.4,

$$\frac{1}{\delta^2} - 1 \leq B^* \leq \frac{1}{K_3} - 1. \tag{8.7}$$

By the definitions of  $f^*(t)$ ,  $t_3$ ,  $t_4$ ,  $C_3$  and  $C_4$  in (8.2), (7.2), (8.5), (7.1) and (8.4), along with the inequality (8.7), we obtain that

$$\begin{aligned} f^*(t_4) &= \frac{1}{1 + B^* e^{-(C_4 - C_3)}} \\ &= \frac{1}{1 + B^* \left(\frac{1}{\delta^2} - 1\right)^{-1} \left(\frac{1}{K_3} - 1\right)^{-1}} \\ &\leq \frac{1}{1 + \left(\frac{1}{K_3} - 1\right)^{-1}} \\ &= 1 - K_3, \end{aligned} \tag{8.8}$$

and

$$f^*(t_4) = \frac{1}{1 + B^* \left(\frac{1}{\delta^2} - 1\right)^{-1} \left(\frac{1}{K_3} - 1\right)^{-1}} \geq \frac{1}{1 + \left(\frac{1}{\delta^2} - 1\right)^{-1}} = 1 - \delta^2. \tag{8.9}$$

Since  $K_3 \leq \delta^2$ , by the definition of  $x_3^*(t)$  in (8.3), along with (8.8) and (8.9),

$$X_3(t_4) \leq 2f^*(t_4)N + \frac{K_3^2 N}{2\delta^2} \leq \left(2 - 2K_3 + \frac{K_3^2}{2\delta^2}\right)N \leq \left(2 - \frac{3K_3}{2}\right)N,$$

and

$$X_3(t_4) \geq \left(2 - 2\delta^2 - \frac{K_3^2}{2\delta^2}\right)N \geq \left(2 - \frac{5\delta^2}{2}\right)N.$$

Lastly, we also have that

$$X_1(t_4) + X_2(t_4) \geq 2N(x_1^*(t_4) + x_2^*(t_4)) - \frac{K_3^2 N}{\delta^2} = 2(1 - f^*(t_4))N - \frac{K_3^2 N}{\delta^2} \geq K_3 N.$$

This completes the proof of the proposition.  $\square$

## 9. Phase 5 and the proof of Theorem 1.1

We begin by defining

$$t_{5+} = t_4 + \frac{1}{1 - 2\delta^2} \cdot \frac{1}{s} \ln(Ns), \tag{9.1}$$

$$t_{5-} = t_4 + (1 - \delta) \cdot \frac{1}{s} \ln(Ns), \tag{9.2}$$

$$T_7 = \inf\{t \geq t_4 : X_3(t) = 2N\}, \tag{9.3}$$

$$T_8 = \inf\{t \geq t_4 : X_3(t) \leq 2N - \lfloor 4\delta^2 N \rfloor\},$$

At time  $t_4$ , type 3 has already become the majority of the population and is waiting to fixate. The fixation will occur between the time  $t_{5-}$  and the time  $t_{5+}$ . We will first prove Lemma 9.1, which shows that type 3 almost certainly fixates before the number of type 3 hits the threshold level  $2N - \lfloor 4\delta^2 N \rfloor$ . In Lemma 9.2, we show that the fixation of type 3 almost certainly occurs before time  $t_{5+}$ . Then, in Lemma 9.3, we show that fixation of type 3 happens after time  $t_{5-}$  with high probability. The technique used in the proofs involves coupling with a branching process; this is similar to the proof of Lemma 5.9. Lastly, we combine the results from these three lemmas in Lemma 9.4, which directly implies Theorem 1.

**Lemma 9.1.** *For sufficiently large  $N$ ,  $P(T_7 < T_8 | \mathcal{F}_{t_4}) \geq 1 - \epsilon$  on the event  $A_{(4)}$ .*

*Proof:* We consider the process  $(2N - X_3(t), t \geq t_4)$ . For  $t \geq 0$ , let  $B(t)$  and  $D(t)$  be the rates that this process increases and decreases by 1 at time  $t$ . This process increases by 1 at rate

$$B(t) = (1 - 2s)X_3(t)(1 - (1 - r)\tilde{X}_3(t) - r(\tilde{X}_1(t) + \tilde{X}_3(t))(\tilde{X}_2(t) + \tilde{X}_3(t))),$$

and decreases by 1 at rate

$$D(t) = (X_0(t) + (1 - s)X_1(t) + (1 - s)X_2(t)) \cdot [(1 - r)\tilde{X}_3(t) + r(\tilde{X}_1(t) + \tilde{X}_3(t))(\tilde{X}_2(t) + \tilde{X}_3(t))] + \mu(X_1(t) + X_2(t)).$$

Then, for all  $t \geq t_4$ , we have

$$\begin{aligned} B(t) &= (1 - 2s)X_3(t)(1 - \tilde{X}_3(t) + r(\tilde{X}_0(t)\tilde{X}_3(t) - \tilde{X}_1(t)\tilde{X}_2(t))) \\ &\leq (1 - 2s)(1 + r)X_3(t)(1 - \tilde{X}_3(t)) \\ &\leq (1 - 2s + r)X_3(t)(1 - \tilde{X}_3(t)), \end{aligned}$$

and

$$\begin{aligned} D(t) &\geq (1 - s)(X_0(t) + X_1(t) + X_2(t)) \cdot (1 - r)\tilde{X}_3(t) \\ &= (1 - s)(1 - r)X_3(t)(1 - \tilde{X}_3(t)) \\ &\geq (1 - s - r)X_3(t)(1 - \tilde{X}_3(t)). \end{aligned}$$

Hence, we can think of the process  $(2N - X_3(t), t \in [t_4, T_7])$  as a birth-death process in which each individual gives birth at rate bounded above by  $(1 - 2s + r)\tilde{X}_3(t)$  and dies at rate bounded below by  $(1 - s - r)\tilde{X}_3(t)$ .

Let  $(Y(t), t \geq t_4)$  be a birth-death process in which each individual gives birth at rate  $b(t) = (1 - 2s + r)\tilde{X}_3(t)$  and dies at rate  $d(t) = (1 - s - r)\tilde{X}_3(t)$ , and we set  $Y(t_4) = 2N - X_3(t_4)$ . It is possible to couple the process  $(Y(t), t \geq t_4)$  with the process  $(2N - X_3(t), t \geq t_4)$  such that for any time  $t \geq t_4$ , we have  $Y(t) \geq 2N - X_3(t)$ . This implies that if the process  $Y$  reaches 0 before  $\lfloor 4\delta^2 N \rfloor$ , then the process  $2N - X_3$  will also reach 0 before  $\lfloor 4\delta^2 N \rfloor$ , which means that  $T_7 < T_8$ .

Here, since we are only interested in the probability that the process  $Y$  reaches 0 before  $\lfloor 2\delta^2 N \rfloor$ , we will consider the induced discrete-time jump process of  $Y$  in the time interval  $[t_4, T_7 \wedge T_8)$ . It is an asymmetric random walk process that jumps up by 1 with probability

$$\frac{b(t)}{b(t) + d(t)} = \frac{1 - 2s + r}{2 - 3s},$$

and jumps down by 1 with probability

$$\frac{d(t)}{b(t) + d(t)} = \frac{1 - s - r}{2 - 3s}.$$

On the event  $A_{(4)}$ , we have from Proposition 2.5 that  $2N - X_3(t_4) \leq 5\delta^2 N/2$ . Let  $q = (1 - s - r)/(1 - 2s + r)$ . Since  $r \ll s$ , for sufficiently large  $N$ , we have  $q \geq (1 - 1.1s)/(1 - 1.9s) > 1$ . For sufficiently large  $N$ , on the event  $A_{(4)}$ , conditioning on the event  $2N - X_3(t_4) = k$ , the probability that this asymmetric random walk

reaches 0 before  $\lfloor 4\delta^2 N \rfloor$  is

$$\begin{aligned} 1 - \frac{q^k - 1}{q^{\lfloor 2\delta^2 N \rfloor} - 1} &\geq 1 - q^{k-2\delta^2 N} \\ &\geq 1 - q^{(\frac{5\delta^2}{2} - 4\delta^2)N} \\ &= 1 - q^{-3\delta^2 N/2} \\ &\geq 1 - \left( \frac{(1 - 1.9s)^{1/s}}{(1 - 1.1s)^{1/s}} \right)^{3\delta^2 Ns/2}, \end{aligned}$$

and note that this upper bound is no longer depends on  $k$ . Since  $s \ll 1$ , when  $N \rightarrow \infty$ , we have

$$\frac{(1 - 1.9s)^{1/s}}{(1 - 1.1s)^{1/s}} \rightarrow \frac{e^{-1.9}}{e^{-1.1}} = e^{-0.8}.$$

Also, because  $Ns \gg 1$ , it follows that when  $N \rightarrow \infty$ , we have

$$\left( \frac{(1 - 1.9s)^{1/s}}{(1 - 1.1s)^{1/s}} \right)^{3\delta^2 Ns/4} \rightarrow 0.$$

Thus, on the event  $A_{(4)}$ , for sufficiently large  $N$ , the probability that the asymmetric random walk reaches 0 before  $\lfloor 4\delta^2 N \rfloor$  is bounded below by  $1 - \epsilon$ . Therefore, through the coupling, for sufficiently large  $N$ , we have  $P(T_7 < T_8 | \mathcal{F}_{t_4}) \geq 1 - \epsilon$  on the event  $A_{(4)}$ .  $\square$

**Lemma 9.2.** *For sufficiently large  $N$ ,  $P(T_7 \leq t_{5+} | \mathcal{F}_{t_4}) \geq 1 - \epsilon - \delta$  on the event  $A_{(4)}$ .*

*Proof:* We use the process  $(Y(t), t \geq t_4)$  defined in the proof of the previous lemma. From the previous lemma, for sufficiently large  $N$ , on the event  $A_{(4)}$ ,

$$\begin{aligned} P(T_7 \leq t_{5+} | \mathcal{F}_{t_4}) &\geq P(T_7 < T_8 | \mathcal{F}_{t_4}) - P(t_{5+} < T_7 \wedge T_8 | \mathcal{F}_{t_4}) \\ &\geq 1 - \epsilon - P(t_{5+} < T_7 \wedge T_8 | \mathcal{F}_{t_4}). \end{aligned}$$

So, we only need to show that for sufficiently large  $N$ , on the event  $A_{(4)}$ , we have  $P(t_{5+} < T_7 \wedge T_8 | \mathcal{F}_{t_4}) \leq \delta$ .

Now, we perform a time change. We define  $\lambda(t) = \int_0^t \tilde{X}_3(t_4 + v) dv$  for  $t \in [0, (T_7 \wedge T_8) - t_4]$ . Then, we define  $Y^*(t) = Y(\lambda^{-1}(t))$  for  $t \in [0, \lambda((T_7 \wedge T_8) - t_4)]$ . The process  $(Y^*(t), t \in [0, \lambda((T_7 \wedge T_8) - t_4)])$  is a birth-death process satisfying  $Y^*(0) = 2N - X_3(t_4)$  and each individual gives birth at rate  $1 - 2s + r$  and dies at rate  $1 - s - r$ .

For sufficiently large  $N$ , on the event that  $t_{5+} < T_7 \wedge T_8$ , we have

$$\begin{aligned} \lambda(t_{5+} - t_4) &= \int_0^{t_{5+} - t_4} \tilde{X}_3(t_4 + v) dv \\ &> \left( 1 - \frac{\lfloor 4\delta^2 N \rfloor}{2N} \right) (t_{5+} - t_4) \geq (1 - 2\delta^2)(t_{5+} - t_4) \\ &= \frac{1}{s} \ln(Ns). \end{aligned}$$

It follows that

$$\begin{aligned} P(t_{5+} < T_7 \wedge T_8 | \mathcal{F}_{t_4}) &= P(\{Y^*(\lambda(t_{5+} - t_4)) > 0\} \cap \{t_{5+} < T_7 \wedge T_8\} | \mathcal{F}_{t_4}) \\ &\leq P\left(Y^*\left(\frac{1}{s} \ln(Ns)\right) > 0 \middle| \mathcal{F}_{t_4}\right). \end{aligned}$$

By the same reason used to obtain (5.37), which gives the probability that the birth-death process survives until time  $t$  when the process starts with one individual, we can generalize the result to the process that begins with any finite number of individuals. If  $k \leq 5\delta^2 N/2$ , then

$$\begin{aligned} &P\left(Y^*\left(\frac{1}{s} \ln(Ns)\right) > 0 \middle| Y^*(0) = k\right) \\ &= 1 - \left(1 - \frac{(1 - 2s + r) - (1 - s - r)}{(1 - 2s + r) - (1 - s - r)e^{-((1-2s+r)-(1-s-r)) \cdot \frac{1}{s} \ln(Ns)}}\right)^k \\ &= 1 - \left(1 - \frac{s - 2r}{(1 - s - r)e^{-\frac{2r}{s} \ln(Ns)} Ns - (1 - 2s + r)}\right)^k \\ &\leq 1 - \left(1 - \frac{s}{(1 - s - r)e^{-\frac{2r}{s} \ln(Ns)} Ns - (1 - 2s + r)}\right)^{5\delta^2 N/4}, \end{aligned} \tag{9.4}$$

and note that this upper bound does not depend on  $k$ . Now, by using the facts that  $r \ll s \ll 1$  and  $1 \ll Ns$ , along with (2.5), when  $N$  is sufficiently large, on the event  $A_{(4)}$  on which we know that  $Y^*(0) = 2N - X_3(t_4) \leq 5\delta^2 N/2$  from Proposition 2.5, we have

$$P\left(Y^*\left(\frac{1}{s} \ln(Ns)\right) > 0 \middle| \mathcal{F}_{t_4}\right) \leq 1 - \left(1 - \frac{s}{0.5Ns}\right)^{5\delta^2 N/2} = 1 - \left(1 - \frac{2}{N}\right)^{5\delta^2 N/2}. \tag{9.5}$$

Note that when  $N \rightarrow \infty$ , by using that  $\delta \in (0, \frac{1}{4})$ , we have

$$1 - \left(1 - \frac{2}{N}\right)^{5\delta^2 N/2} \rightarrow 1 - e^{-5\delta^2/2} \leq \frac{5\delta^2}{2} < \delta.$$

This fact and (9.5) prove the inequality that we want to show. □

**Lemma 9.3.** *For sufficiently large  $N$ , we have  $P(t_{5-} < T_7 \wedge T_8 | \mathcal{F}_{t_4}) \geq 1 - 2\epsilon$  on the event  $A_{(4)}$ .*

*Proof:* The proof is similar to the proof of Lemma 9.2. In this proof, we are going to consider the process  $(X_1(t) + X_2(t), t \geq t_4)$ . For  $t \geq t_4$ , let  $B(t)$  and  $D(t)$  be the rates at which the process increases or decreases by 1. We will now give a lower bound for  $B(t)$  and an upper bound for  $D(t)$ . For the increasing rate, one way to increase  $X_1(t) + X_2(t)$  is by having a type 0 or type 3 individual die, which occurs at the total rate  $X_0(t) + (1 - 2s)X_3(t)$ , and the new individual is type 1 or 2 that is created without recombination, which occurs with probability  $(1 - r)(\tilde{X}_1(t) + \tilde{X}_2(t))$ . Then,

$$\begin{aligned} B(t) &\geq (X_0(t) + (1 - 2s)X_3(t)) \cdot (1 - r)(\tilde{X}_1(t) + \tilde{X}_2(t)) \\ &\geq (1 - 2s)(1 - r)(X_0(t) + X_3(t))(\tilde{X}_1(t) + \tilde{X}_2(t)) \\ &\geq (1 - 2s - r)(\tilde{X}_0(t) + \tilde{X}_3(t))(X_1(t) + X_2(t)). \end{aligned}$$

One way to decrease  $X_1(t) + X_2(t)$  is by having a type 1 or type 2 dies, which occurs at total rate  $(1-s)(X_1(t) + X_2(t))$ , and the new individual cannot be type 1 or 2, which occurs with probability bounded above by  $1 - (1-r)(\tilde{X}_1(t) + \tilde{X}_2(t))$ . Another way to decrease  $X_1(t) + X_2(t)$  by having a type 1 or 2 mutate to type 3, which occurs at rate  $\mu(X_1(t) + X_2(t))$ . So,

$$\begin{aligned} D(t) &\leq (1-s)(X_1(t) + X_2(t)) \cdot (1 - (1-r)(\tilde{X}_1(t) + \tilde{X}_2(t))) + \mu(X_1(t) + X_2(t)) \\ &= ((1-s)(\tilde{X}_0(t) + \tilde{X}_3(t) + r(\tilde{X}_1(t) + \tilde{X}_2(t))) + \mu) \cdot (X_1(t) + X_2(t)). \end{aligned}$$

When  $t \in [t_4, T_7 \wedge T_8]$ , we have  $X_3(t) > N$ . So,  $\tilde{X}_1(t) + \tilde{X}_2(t) \leq \tilde{X}_0(t) + \tilde{X}_3(t)$  and  $\mu \leq 2(\tilde{X}_0(t) + \tilde{X}_3(t))\mu$ . Hence, when  $t \in [t_4, T_7 \wedge T_8]$ ,

$$\begin{aligned} D(t) &\leq (1-s)(1+r+2\mu)(\tilde{X}_0(t) + \tilde{X}_3(t))(X_1(t) + X_2(t)) \\ &\leq (1-s+r+2\mu)(\tilde{X}_0(t) + \tilde{X}_3(t))(X_1(t) + X_2(t)). \end{aligned}$$

Let  $(Y(t), t \geq t_4)$  be a birth-death process in which each individual gives birth at rate

$$b(t) = (1-2s-r)(\tilde{X}_0(t) + \tilde{X}_3(t))$$

and dies at rate

$$d(t) = (1-s+r+2\mu)(\tilde{X}_0(t) + \tilde{X}_3(t)).$$

Also, we let  $Y(t_4) = X_1(t_4) + X_2(t_4)$ . We can couple this process with the process  $(X_1(t) + X_2(t), t \geq t_4)$  in a way that for  $t \in [t_4, T_7 \wedge T_8]$ ,  $Y(t) \leq X_1(t) + X_2(t)$ ; that is if  $Y(t) > 0$ , then  $X_1(t) + X_2(t) > 0$ .

Next, we construct a time change. We define  $\lambda(t) = \int_0^t (\tilde{X}_0(t_4+v) + \tilde{X}_3(t_4+v)) dv$  for  $t \in [0, (T_7 \wedge T_8) - t_4]$ . Note that  $\lambda(t) \leq t$  for all  $t \in [0, (T_7 \wedge T_8) - t_4]$ . Now, we define  $Y^*(t) = Y(\lambda^{-1}(t))$ . It follows that the process  $(Y^*(t), t \in [0, \lambda((T_7 \wedge T_8) - t_4)])$  is a birth-death process such that  $Y^*(0) = X_1(t_4) + X_2(t_4)$ , in which each individual gives birth at rate  $1-2s-r$  and dies at rate  $1-s+r+2\mu$ . We extend the process  $Y^*$  so that is defined for the entire time interval  $[0, \infty)$  with constant birth and death rates.

From the definition of  $T_7$  in (9.3), if  $T_7 \leq t_{5-} \wedge T_8$ , then  $Y^*(t_{5-} - t_4) = 0$ , because  $Y^*(\lambda(T_7 - t_4)) = Y(T_7 - t_4) = 0$  and  $\lambda(T_7 - t_4) \leq T_7 - t_4 \leq t_{5-} - t_4$ . Thus,

$$\begin{aligned} P(Y^*(t_{5-} - t_4) > 0 | \mathcal{F}_{t_4}) &\leq P(T_7 > t_{5-} \wedge T_8 | \mathcal{F}_{t_4}) \\ &\leq P(t_{5-} < T_7 \wedge T_8 | \mathcal{F}_{t_4}) + \leq P(T_7 \geq T_8 | \mathcal{F}_{t_4}). \end{aligned}$$

By Lemma 9.1, for sufficiently large  $N$ , on the event  $A_{(4)}$ ,

$$P(t_{5-} < T_7 \wedge T_8 | \mathcal{F}_{t_4}) \geq P(Y^*(t_{5-} - t_4) > 0 | \mathcal{F}_{t_4}) - \epsilon.$$

It is left show that for sufficiently large  $N$ , on the event  $A_{(4)}$ ,

$$P(Y^*(t_{5-} - t_4) > 0 | \mathcal{F}_{t_4}) = P\left(Y^*\left((1-\delta) \cdot \frac{1}{s} \ln(Ns)\right) > 0 \middle| \mathcal{F}_{t_4}\right) \geq 1 - \epsilon. \quad (9.6)$$

Similar to the way we get (9.4), if  $k \geq K_3 N$ , we get that

$$\begin{aligned}
& P\left(Y^*\left((1-\delta) \cdot \frac{1}{s} \ln(Ns)\right) > 0 \mid Y^*(0) = k\right) \\
&= 1 - \left(1 - \frac{(1-2s-r) - (1-s+r+2\mu)}{(1-2s-r) - (1-s+r+2\mu)e^{-((1-2s-r)-(1-s+r+2\mu)) \cdot \frac{1-\delta}{s} \ln(Ns)}}\right)^k \\
&= 1 - \left(1 - \frac{s+2r+2\mu}{(1-s+r+2\mu)e^{\frac{2(1-\delta)r}{s} \ln(Ns) + \frac{2(1-\delta)\mu}{s} \ln(Ns)} (Ns)^{1-\delta} - (1-2s-r)}\right)^k \\
&\geq 1 - \left(1 - \frac{s}{(1-s+r+2\mu)e^{\frac{2(1-\delta)r}{s} \ln(Ns) + \frac{2(1-\delta)\mu}{s} \ln(Ns)} (Ns)^{1-\delta} - (1-2s-r)}\right)^{K_3 N},
\end{aligned}$$

and note that this lower bound does not depend on  $k$ . We know from Lemma 2.1 that  $\frac{r}{s} \ln(Ns) \ll 1$ . Also, since  $N\mu^2 \leq s$ , we get  $\frac{\mu}{s} \ln(Ns) \ll \frac{1}{\sqrt{Ns}} \ln(Ns) \ll 1$ . Hence, for sufficiently large  $N$ , if  $k \geq K_3 N$ , then

$$\begin{aligned}
P\left(Y^*\left((1-\delta) \cdot \frac{1}{s} \ln(Ns)\right) > 0 \mid \mathcal{F}_{t_4}\right) &\geq 1 - \left(1 - \frac{s}{2(Ns)^{1-\delta}}\right)^{K_3 N} \\
&= 1 - \left(1 - \frac{0.5(Ns)^\delta}{N}\right)^{K_3 N}. \tag{9.7}
\end{aligned}$$

From Proposition 2.5, we know that  $Y^*(0) = Y(t_4) = X_1(t_4) + X_2(t_4) \geq K_3 N$  on the event  $A_{(4)}$ . So, (9.7) holds on the event  $A_{(4)}$ . Because  $1 \ll Ns$ , the lower bound converges to 1 as  $N \rightarrow \infty$ . This fact and (9.7) prove (9.6).  $\square$

**Lemma 9.4.** *For sufficiently large  $N$ , we have*

$$P(A_{(4)} \cap \{t_{5-} < T_7 < t_{5+}\}) \geq 1 - 29\epsilon - 8\delta - \delta^2.$$

*Proof:* From Lemmas 9.1, 9.2, and 9.3, for sufficiently large  $N$ , on the event  $A_{(4)}$ ,

$$P(t_{5-} < T_7 < t_{5+} \mid \mathcal{F}_{t_4}) \geq 1 - 3\epsilon - \delta.$$

Hence, by Proposition 2.5, for sufficiently large  $N$ , we have  $P(A_{(4)} \cap \{t_{5-} < T_7 < t_{5+}\}) \geq 1 - 29\epsilon - 8\delta - \delta^2$ .  $\square$

Finally, we are now giving the proof of Theorem 1.

*Proof of Theorem 1.1:* First, for every subsequence  $(N_k)_{k=1}^\infty$ , there is a further subsequence that satisfies (2.1), or there is a further subsequence that satisfies (2.2). By a subsequence argument, it is enough to prove Theorem 1.1 in the recombination dominating case and in the mutation dominating case.

Now, recall that the stopping time  $T$  defined in Theorem 1.1 is the first time that type 3 individuals have fixated in the population. We will show that if  $\theta \in (0, 1)$ , then for sufficiently large  $N$ , we have

$$P((1-\theta)t_N^*(r_N) \leq T \leq (1+\theta)t_N^*(r_N)) \geq 1 - 38\epsilon.$$

We choose  $\delta$  to be small enough so that (1)  $\delta < \epsilon$ , (2)  $(1-\delta^2)^{-1} < 1+\theta$  and (3)  $1-2\delta > 1-\theta$ . From Lemma 9.4,

$$P(A_{(4)} \cap \{t_{5-} < T_7 < t_{5+}\}) \geq 1 - 29\epsilon - 8\delta - \delta^2 \geq 1 - 38\epsilon$$

for sufficiently large  $N$ . Note that from the definition of  $T_7$  in (9.3), we have  $T_7 = T \vee t_4$ . Also, by the definition of  $t_{5-}$ , for sufficiently large  $N$ , we have

$t_{5-} > t_4$ . Thus,  $P(t_{5-} < T < t_{5+}) = P(t_{5-} < T_7 < t_{5+}) \geq 1 - 38\epsilon$  when  $N$  is sufficiently large. By simple calculation, one can show that for sufficiently large  $N$ ,  $(1 - \theta)t_N^*(r_N) \leq t_{5-}$  and  $t_{5+} < (1 + \theta)t_N^*(r_N)$ . This completes the proof.  $\square$

### Acknowledgements

The author is grateful that Professor Jason Schweinsberg suggested the author to read one his previous paper; several ideas from the paper are used intensively in this article. The author also would like to thank him for reading the manuscript and giving some helpful advices.

Lastly, the author would like to thank anonymous referees for pointing out several mistakes and suggesting several related works, which improves the introduction part of this article significantly.

### References

- Athreya, K. B. and Ney, P. E. *Branching processes*. Springer-Verlag, New York-Heidelberg (1972). [MR0373040](#).
- Barton, N. H. and Otto, S. P. Evolution of recombination due to random drift. *Genetics*, **169** (4), 2353–2370 (2005). [DOI: 10.1534/genetics.104.032821](#).
- Berestycki, N. and Zhao, L. Z. The shape of multidimensional Brunet-Derrida particle systems. *Ann. Appl. Probab.*, **28** (2), 651–687 (2018). [MR3784486](#).
- Bossert, S. and Pfaffelhuber, P. The fixation probability and time for a doubly beneficial mutant. *Stochastic Process. Appl.*, **128** (12), 4018–4050 (2018). [MR3906977](#).
- Bürger, R. Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics*, **153** (2), 1055–1069 (1999). [www.genetics.org/content/153/2/1055](#).
- Charlesworth, B. Directional selection and the evolution of sex and recombination. *Genet. Res.*, **61** (3), 205–224 (1993a). [DOI: 10.1017/S0016672300031372](#).
- Charlesworth, B. The evolution of sex and recombination in a varying environment. *J. Hered.*, **84** (5), 345–350 (1993b). [DOI:10.1093/oxfordjournals.jhered.a111355](#).
- Cohen, E., Kessler, D. A., and Levine, H. Analytic approach to the evolutionary effects of genetic exchange. *Phys. Rev. E*, **73** (1), 016113 (2006). [DOI: 10.1103/PhysRevE.73.016113](#).
- Crow, J. F. and Kimura, M. Evolution in sexual and asexual populations. *Am. Nat.*, **99** (909), 439–450 (1965). [DOI: 10.1086/282389](#).
- Cuthbertson, C., Etheridge, A., and Yu, F. Fixation probability for competing selective sweeps. *Electron. J. Probab.*, **17**, no. 31, 36 (2012). [MR2915667](#).
- Darling, R. W. R. and Norris, J. R. Differential equation approximations for Markov chains. *Probab. Surv.*, **5**, 37–79 (2008). [MR2395153](#).
- Desai, M. M. and Fisher, D. S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics*, **176** (3), 1759–1798 (2007). [DOI: 10.1534/genetics.106.067678](#).
- Durrett, R. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer, New York, second edition (2008). ISBN 978-0-387-78168-6. [MR2439767](#).
- Felsenstein, J. The evolutionary advantage of recombination. *Genetics*, **78** (2), 737–756 (1974). [www.genetics.org/content/78/2/737](#).

- Fisher, D. S. Asexual evolution waves: fluctuations and universality. *J. Stat. Mech. Theory Exp.*, (1), P01011, 65 (2013). [MR3036201](#).
- Fisher, R. A. *The genetical theory of natural selection*. Oxford University Press, Oxford, variorum edition (1999). ISBN 0-19-850440-3. [MR1785121](#).
- Fogle, C. A., Nagle, J. L., and Desai, M. M. Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics*, **180** (4), 2163–2173 (2008). DOI: [10.1534/genetics.108.090019](#).
- Gerrish, P. J. and Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica*, **102**, 127–144 (1998). DOI: [10.1023/A:1017067816551](#).
- Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O., and Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (13), 4950–4955 (2012). DOI: [10.1073/pnas.1119910109](#).
- Haigh, J. The accumulation of deleterious genes in a population—Muller’s ratchet. *Theoret. Population Biol.*, **14** (2), 251–267 (1978). [MR514423](#).
- Hallatschek, O. The noisy edge of traveling waves. *Proc. Natl. Acad. Sci. U.S.A.*, **108** (5), 1783–1787 (2011). DOI: [10.1073/pnas.1013529108](#).
- Hartfield, M., Otto, S. P., and Keightley, P. D. The role of advantageous mutations in enhancing the evolution of a recombination modifier. *Genetics*, **184** (4), 1153–1164 (2010). DOI: [10.1534/genetics.109.112920](#).
- Hill, W. G. and Robertson, A. The effect of linkage on limits to artificial selection. *Genetical Research*, **8** (3), 269–294 (1966). DOI: [10.1017/S0016672300010156](#).
- Keightley, P. and Otto, S. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature*, **443**, 89–92 (2006). DOI: [10.1038/nature05049](#).
- Maynard Smith, J. Evolution in sexual and asexual populations. *Am. Nat.*, **102** (927), 469–473 (1968). DOI: [10.1086/282559](#).
- Muller, H. J. Some genetic aspects of sex. *Am. Nat.*, **66** (703), 118–138 (1932). DOI: [10.1086/280418](#).
- Neher, R. A., Kessinger, T. A., and Shraiman, B. I. Coalescence and genetic diversity in sexual populations under selection. *Proc. Natl. Acad. Sci. U.S.A.*, **110** (39), 15836–15841 (2013). DOI: [10.1073/pnas.1309697110](#).
- Neher, R. A., Shraiman, B. I., and Fisher, D. S. Rate of adaptation in large sexual populations. *Genetics*, **184** (2), 467–481 (2010). DOI: [10.1534/genetics.109.109009](#).
- Otto, S. P. and Feldman, M. W. Deleterious Mutations, Variable Epistatic Interactions, and the Evolution of Recombination. *Theor. Popul. Bio.*, **51** (2), 134–147 (1997). DOI: [10.1006/tpbi.1997.1301](#).
- Park, S. C., Simon, D., and Krug, J. The speed of evolution in large asexual populations. *J. Stat. Phys.*, **138** (1-3), 381–410 (2010). [MR2594902](#).
- Peck, J. R. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, **137** (2), 597–606 (1994). [www.genetics.org/content/137/2/597](#).
- Rouzine, I. M., Brune, E., and Wilke, C. O. The traveling wave approach to asexual evolution: Muller’s ratchet and speed of adaptation. *Theor Popul Biol.*, **73** (1), 24–46 (2008). DOI: [10.1016/j.tpb.2007.10.004](#).

- Rouzine, I. M. and Coffin, J. M. Multi-site adaptation in the presence of infrequent recombination. *Theor Popul Biol.*, **77** (3), 189–204 (2010). DOI: [10.1016/j.tpb.2010.02.001](https://doi.org/10.1016/j.tpb.2010.02.001).
- Rouzine, I. M., Wakeley, J., and Coffin, J. M. The solitary wave of asexual evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **100** (2), 587–592 (2003). DOI: [10.1073/pnas.242719299](https://doi.org/10.1073/pnas.242719299).
- Roze, D. and Barton, N. H. The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Genetics*, **173** (3), 1793–1811 (2006). DOI: [10.1534/genetics.106.058586](https://doi.org/10.1534/genetics.106.058586).
- Schweinsberg, J. Rigorous results for a population model with selection I: evolution of the fitness distribution. *Electron. J. Probab.*, **22**, Paper No. 37, 94 (2017). MR3646063.
- Takahata, N. Sexual recombination under the joint effects of mutation, selection, and random sampling drift. *Theoret. Population Biol.*, **22** (2), 258–277 (1982). MR679306.
- Van Valen, L. A New Evolutionary Law. *Evol. Theory*, **1** (1), 1–30 (1973). PASCALGEODEBRGM7720012907.
- Waxman, D. and Peck, J. R. Sex and adaptation in a changing environment. *Genetics*, **153** (2), 1041–1053 (1999). [www.genetics.org/content/153/2/1041](http://www.genetics.org/content/153/2/1041).
- Weissman, D. B. and Barton, N. H. Limits to the rate of adaptive substitution in sexual populations. *PLoS Genet.*, **8** (6), e1002740 (2012). DOI: [10.1371/journal.pgen.1002740](https://doi.org/10.1371/journal.pgen.1002740).
- Weissman, D. B. and Hallatschek, O. The rate of adaptation in large sexual populations with linear chromosomes. *Genetics*, **196** (4), 1167–1183 (2014). DOI: [10.1534/genetics.113.160705](https://doi.org/10.1534/genetics.113.160705).
- Wilke, C. O. The speed of adaptation in large asexual populations. *Genetics*, **167** (4), 2045–2053 (2004). DOI: [10.1534/genetics.104.027136](https://doi.org/10.1534/genetics.104.027136).