# MathematicS Mathsin $A$. In Action 

François G. Ged

Moran model with simultaneous strong and weak selections: convergence towards a $\Lambda$-Wright-Fisher SDE
Volume 12 (2023), p. 87-116.
https://doi.org/10.5802/msia. 33
© Les auteurs, 2023.
(c) BY $\quad$ Cet article est mis à disposition selon les termes de la licence Creative Commons attribution 4.0.
http://creativecommons.org/licenses/by/4.0/


MERSENNE
MathematicS In Action est membre $d u$

# Moran model with simultaneous strong and weak selections: convergence towards a $\Lambda$-Wright-Fisher SDE 

François G. Ged *<br>* Chair of Statistical Field Theory, École Polytechnique Fédérale de Lausanne, Switzerland<br>E-mail address: francois.ged@epfl.ch.


#### Abstract

We establish a connection between two population models by showing that one is the scaling limit of the other, as the population grows large. In the infinite population model, individuals are split into two subpopulations, carrying either a selective advantageous allele, or a disadvantageous one. The proportion of disadvantaged individuals in the population evolves according to the $\Lambda$-Wright-Fisher stochastic differential equation (SDE) with selection, and the genealogy is described by the so-called Bolthausen-Sznitman coalescent. This equation has appeared in the $\Lambda$-lookdown model with selection studied by Bah and Pardoux [1]. Schweinsberg in [16] showed that in a specific setting, due to the strong selection, the genealogy of the so-called Moran model with selection converges to the Bolthausen-Sznitman coalescent. By splitting the population into two adversarial subgroups and adding a weak selection mechanism, we show that the proportion of disadvantaged individuals in the Moran model with strong and weak selections converges to the solution of the $\Lambda$-Wright-Fisher SDE of [1].


## 1. Introduction

The Moran model is a classical model in population genetics. It describes the evolution in continuous time of a haploid population with constant size, where generations are overlapping. Every individual dies at rate 1 and is instantaneously replaced by a copy of an individual chosen uniformly at random in the remaining population, including the individual who just died. It is well-known that the genealogy of the Moran model is described by the so-called Kingman's coalescent, which is the only exchangeable ${ }^{1}$ coalescent process where merging events are only binary and non-simultaneous. Some kind of universality of Kingman's coalescent for the genealogy of discrete time population models with fixed size was established in [12]; this result is known as Möhle's Lemma. In [13], Möhle also obtained convergence results towards different coalescents, and even allowed the size of the population to vary.
In the Moran model of size $N$, when the population is split into two subgroups, say the individuals carrying allele $X$ and the ones carrying allele $Y$, the proportion $\left(X_{t}\right)_{t \geq 0}$ of allele $X$ in the whole population converges as $N \rightarrow \infty$, when speeding up the time by a factor $N$, towards the Wright-Fisher diffusion, that is the solution to the SDE

$$
\mathrm{d} X_{t}=\sqrt{X_{t}\left(1-X_{t}\right)} \mathrm{d} W_{t}
$$

where $W$ is a standard Brownian motion. Note the symmetry between the two alleles reflecting the fact that none of them has a selective advantage over the other. One of the implications is the well-known duality relation between the number of blocks in Kingman's coalescent and the Wright-Fisher diffusion, as stated in [2, Theorem 2.7]. Namely, denoting $K_{t}$ the number of

[^0]blocks in Kingman's coalescent at time $t$, it holds for all $x \in(0,1)$ and $k \in \mathbb{N}$ that
\[

$$
\begin{equation*}
\mathbb{E}\left(X_{t}^{k} \mid X_{0}=x\right)=\mathbb{E}\left(x^{K_{t}} \mid K_{0}=k\right) . \tag{1.1}
\end{equation*}
$$

\]

The duality actually holds for more general coalescents, namely $\Lambda$-coalescents, and some Fleming-Viot processes, as shown in [3, Equation (18)], that is, for $K_{t}$ the number of blocks at time $t$ in a $\Lambda$-coalescent and $\left(X_{t}\right)_{t \geq 0}$ the solution of some SDE.

The coalescence rates in a $\Lambda$-coalescent are characterized by a finite measure $\Lambda$ on $[0,1]$, such that if the coalescent contains $k$ blocks at some given time, any sub-family of size $\ell$ among the $k$ blocks merge at rate given by

$$
\lambda_{k, \ell}:=\int_{[0,1]} p^{\ell-2}(1-p)^{k-\ell} \Lambda(\mathrm{d} p) .
$$

In particular, the blocks are exchangeable, in the sense that all the possible combinations of $\ell$ blocks have the same rate of merging. Kingman's coalescent corresponds to the case $\Lambda=\delta_{0}$, the Dirac mass at 0 . Another instance of $\Lambda$-coalescent is the Bolthausen-Sznitman coalescent introduced in [6], corresponding to $\Lambda(\mathrm{d} p)=\mathrm{d} p$. Its importance is due to its connections to models such as spin glasses, continuous branching processes, travelling waves, some population models, see e.g. [2] and references therein. The populations where we expect to observe the Bolthausen-Sznitman coalescent are for instance populations undergoing strong selection [16], exploring uninhabited territories [7], or quickly adapting to the environment [9, 14]. In those cases, an individual sometimes reproduces more (or faster) and generates a family of size of the same order as the population size.

Moran model with selection and $\Lambda$-lookdown model. When a death occurs in the Moran model, instead of choosing an individual uniformly at random to reproduce, we can include a selection component in the dynamics and choose the parent proportionally to its fitness. An instance of a Moran model with selection has been studied by Schweinsberg in [15] and [16], where the individuals accumulate beneficial mutations increasing their reproduction rates; this is the model we are interested in this work and we will describe it in more details later on. The main result of [16] establishes that the genealogy of the Moran model with selection of Schweinsberg converges towards the Bolthausen-Sznitman coalescent, as the size of the population goes to infinity. Let us connect it with another population model.

In [1], the authors study an infinite size population model called the $\Lambda$-lookdown model with selection, whose genealogy is that of the corresponding $\Lambda$-coalescent. We are of course interested in the Bolthausen-Sznitman case $\Lambda(\mathrm{d} p)=\mathrm{d} p$ on $[0,1]$. Each individual carries either allele $X$ or allele $Y$, the selection advantaging the individuals of type ${ }^{2} X$. Theorem 3.5 in [1] shows that the proportion of carriers of $Y$ is the solution of the following SDE:

$$
\begin{equation*}
\mathcal{Y}_{t}=\mathcal{Y}_{0}-\alpha \int_{0}^{t} \mathcal{Y}_{s}\left(1-\mathcal{Y}_{s}\right) \mathrm{d} s+\int_{[0, t] \times[0,1]^{2}} p\left(\mathbb{1}_{\left\{u \leq \mathcal{Y}_{s-}\right\}}-\mathcal{Y}_{s-}\right) M(\mathrm{~d} s, \mathrm{~d} u, \mathrm{~d} p), \tag{1.2}
\end{equation*}
$$

where $M$ is a Poisson point process with intensity $\mathrm{d} s \otimes \mathrm{~d} u \otimes \frac{\mathrm{~d} p}{p^{2}}$, and $\alpha \geq 0$ represents the selective advantage of $X$ over $Y$. In [1], Equation (1.2) is called the $\Lambda$-Wright-Fisher SDE with selection. The previously mentioned duality (1.1) in this case is between the solution of (1.2) with $\alpha=0$ and the associated $\Lambda$-coalescent. One may wonder whether it is possible to split the individuals in the Moran model with selection into two adversarial subgroups ( $X$ versus $Y$ ), in order to observe the convergence of the proportion of the disadvantaged group $Y$ towards the solution of (1.2). The goal of this work is to answer this question ${ }^{3}$.

[^1]
## 2. Model and main result

### 2.1. Previous results

We describe more formally the Moran model with selection and the results of Schweinsberg in [15] and [16].

We consider a population of fixed size $N \in \mathbb{N}$. Each individual dies at rate 1, meaning that its lifetime is an exponential random variable with parameter 1 . At time 0 , the $N$ individuals carry no mutation. Each of them acquires a mutation that adds up to its current number of mutations at rate $\mu=\mu_{N}$ that can depend on $N$. We call the number of mutations carried by an individual its type. When a death occurs, say at time $t$, the individual is instantaneously replaced by a copy of an individual chosen in the population at time $t$, including the one who just died, independently from the past. The parent is chosen at random proportionally to its fitness at time $t$, as explained below, and the newborn individual then inherits the type of its parent.
For all $j \geq 0$ and $t \in \mathbb{R}_{+}$, we denote by $W_{j}(t)$ the number of individuals of type $j$ at time $t$ in the population. The average number of mutations at time $t$ is thus given by

$$
M(t):=\frac{1}{N} \sum_{j \geq 0} j W_{j}(t)
$$

Let $s=s_{N}>0$ be the coefficient of selection and let the fitness of the type $j$ at time $t$ be $\max (1+s(j-M(t)), 0)$. If a death occurs at time $t$, the probability that a particular individual of type $j$ reproduces is

$$
F_{j}(t):=\frac{\max (1+s(j-M(t)), 0)}{\sum_{i \geq 0} W_{i}(t) \max (1+s(i-M(t)), 0)},
$$

which becomes

$$
\begin{equation*}
F_{j}(t)=\frac{1+s(j-M(t))}{N} \tag{2.1}
\end{equation*}
$$

when all the fitnesses are positive. The neutral Moran model corresponds to the case where $s=0$, that is all the individuals have the same probability to reproduce. We stress that thus defined, in our model, every mutation is beneficial.

Define

$$
\begin{equation*}
k_{N}:=\frac{\log N}{\log (s / \mu)} \quad \text { and } \quad a_{N}:=\frac{\log (s / \mu)}{s} \tag{2.2}
\end{equation*}
$$

which are proven in [15] to be the scaling constants such that in $a_{N}$ units of time, the difference between the largest type at time $t$ and the largest type at time $t+a_{N}$ is of order $k_{N}$. The assumptions on the parameters of the model are the following:

$$
\begin{gather*}
\lim _{N \rightarrow \infty} \frac{k_{N}}{\log (1 / s)}=\infty .  \tag{1}\\
\lim _{N \rightarrow \infty} \frac{k_{N} \log k_{N}}{\log (s / \mu)}=0 .  \tag{2}\\
\lim _{N \rightarrow \infty} s k_{N}=0 . \tag{3}
\end{gather*}
$$

In particular, it implies that $s \rightarrow 0, k_{N}, a_{N} \rightarrow \infty$ as $N \rightarrow \infty$, and for any $a, b>0$,

$$
\begin{equation*}
\frac{1}{N^{a}} \ll \mu \ll s^{b} . \tag{2.3}
\end{equation*}
$$

We refer to [15] and [16] for more detailed discussions on these assumptions.
We will recall later on the fact from [15] that the number of types that have appeared before time $a_{N} T$ is of order at most $O\left(k_{N}\right)$ with high probability. Since $s k_{N} \rightarrow 0$ as $N \rightarrow \infty$ by

Assumption $\left(A_{3}\right)$, the fitnesses for these types are always positive with high probability, that is, the fitnesses are given by (2.1).

The main results of Schweinsberg in [15] concern the dynamics of the type distribution in the population as $N \rightarrow \infty$. Theorem 1.4 in [15] shows that after $a_{N}$ units of time, the distribution of the types starts looking like that of a Gaussian variable with vanishing variance. Theorem 1.2 in [15] states that $M\left(a_{N} t\right) / k_{N}$ converges in probability and uniformly on compact sets of $(0,1) \cup$ $(1, \infty)$ towards a function $m$. A similar convergence holds for the difference between the fittest individuals (the highest type alive) and the mean type, as shown in [15, Theorem 1.1], towards the map $q: \mathbb{R}_{+} \rightarrow \mathbb{R}_{+}$that satisfies

$$
q(t)= \begin{cases}e^{t} & \text { if } t \in(0,1)  \tag{2.4}\\ \int_{t-1}^{t} q(u) \mathrm{d} u & \text { if } t>1\end{cases}
$$

This fully describes the dynamics of the type distribution as $N \rightarrow \infty$, forward in time. It enabled Schweinsberg in [16] to show, when looking backward in time, that the genealogy of the process converges in finite-dimensional distributions towards the Bolthausen-Sznitman coalescent.

Following the fittest type. An important result in [16] is that, when sampling $n$ individuals in the population at time $a_{N} T$ and looking backwards in time, after $a_{N}$ units of time, all the sampled individuals essentially share the same type with high probability and it goes on for their ancestors, the common type being the fittest (i.e. largest) type in the population. This can be rephrased as follows: after $a_{N}$ units of time forward, only the individuals that were among the fittests have begotten a non-negligeable offspring. In order to track the fittest type, Schweinsberg discretises the time at stopping times defined as follows: for all $j \geq 1$, let

$$
\begin{equation*}
\tau_{j}:=\inf \left\{t \geq 0: W_{j-1}(t)>s / \mu\right\} \tag{2.5}
\end{equation*}
$$

In words, $\tau_{j}$ is approximately the time when type $j$ mutations start occurring, making $j-1$ the fittest type in the population at time $\tau_{j}$; see [15, Equation (3.16)] and the associated discussion. Note that $s / \mu \rightarrow \infty$ as $N \rightarrow \infty$ by (2.3), but $\frac{s / \mu}{N} \rightarrow 0$. Roughly speaking, it means that although the largest type represents a fraction of the whole population close to 0 , once this type reaches a size $\lceil s / \mu\rceil:=\inf \{n \geq 1: n \geq s / \mu\}$, it starts evolving in a very predictable way, which is the reason why this discretisation is powerful. In particular, it is when a type $j$ mutation occurs relatively shortly after $\tau_{j}$ that a large family is likely to descend from it, due to the fact that the fitness is relative to the mean (meaning that individuals mutating faster than usual are getting strongly advantaged for reproducing). To follow the largest type, we introduce the random index

$$
\begin{equation*}
j(t):=\sup \left\{j \geq 1: \tau_{j} \leq a_{N} t\right\} \tag{2.6}
\end{equation*}
$$

which is adapted to the natural filtration of the process $\left(W_{j}(t) ; j \geq 0\right)_{t \geq 0}$. We stress that the notation $j(t)$ refers to another quantity in $[15,16]$.

### 2.2. Adding the weak selection dynamics

Recall that we want to divide our population into two adversarial subgroups, say $X$ and $Y$, giving a selective advantage to $X$ such that the proportion of $Y$-individuals converges towards the solution of (1.2) as $N \rightarrow \infty$. It is important to note that this new selection between groups $X$ and $Y$ should leave unchanged the selection between the different types. Henceforth, we will use the name type without further precisions to refer to the number of mutations carried by an individual, never for his group $X$ or $Y$ alone. Nonetheless we will sometimes use the condensed type $(Y, j)$ to refer to both the group and type of an individual.

For technical reasons due to the fact that the population takes about $a_{N}$ units of time to reach the Bolthausen-Sznitman dynamics, we study the proportion of $Y$-individuals starting only from time $\tau_{j(2)}$, when the types' distribution already looks like a Gaussian distribution. Let $\left(y_{N}\right)_{N \in \mathbb{N}}$
be a sequence in $(0,1)$ such that $y_{N} \rightarrow y \in(0,1)$ as $N \rightarrow \infty$ and $y_{N}\lceil s / \mu\rceil \in\{1, \ldots,\lceil s / \mu\rceil\}$. At time $\tau_{j(2)}$, we mark uniformly at random exactly $y_{N}\lceil s / \mu\rceil$ type $j(2)-1$ individuals to be in the group $Y$. Each individual of type $j<j(2)-1$ is marked with probability $y_{N}$. All the individuals in the population that are not in the $Y$ group form the $X$ group. The usual reproduction mechanism is left unchanged by the membership of $X$ or $Y$. During a reproduction event, the child inherits the group of its parent.

Recall the definition of the map $q$ in (2.4). Starting from some deterministic integer $k_{N}^{*}$ defined above the forthcoming equation (3.7), Schweinsberg introduced deterministic times approximating the random times $\tau_{j}, j \geq k_{N} *+1$, recursively defined by $\tau_{k_{N}^{*}+1}^{*}=0$,

$$
\begin{equation*}
\tau_{j+1}^{*}=\tau_{j}^{*}+\frac{a_{N}}{k_{N} q\left(\tau_{j}^{*} / a_{N}\right)}, \quad j \geq k_{N}^{*}+1 \tag{2.7}
\end{equation*}
$$

We then define

$$
\begin{equation*}
r_{j}:=\frac{\tau_{j+1}^{*}-\tau_{j}^{*}}{a_{N}} \tag{2.8}
\end{equation*}
$$

We add a new selection mechanism, that we call weak selection, operating between groups $X$ and $Y$ as follows. Set the weak selection coefficient $\alpha \geq 0$, that does not depend on $N$. Let $Y_{j}(t)$ be the number of $(Y, j)$-individuals at time $t$ for $j \geq j(2)$, and define $X_{j}(t)$ similarly for the $(X, j)$-individuals. Every time a $(Y, j)$-individual acquires a $j+1$-th mutation, say at a time $t$, it is instead killed with probability

$$
\begin{equation*}
\frac{\alpha}{r_{j+1}} \cdot \frac{X_{j}(t)}{X_{j}(t)+Y_{j}(t)} \tag{2.9}
\end{equation*}
$$

Each killing is immediately compensated by choosing an individual uniformly at random among the $X_{j}(t)(X, j)$-individuals to give birth to a $(X, j+1)$-individual. Note that the dynamics of the types (i.e. number of mutations) thus remain unchanged.

Let $\left(\mathcal{Y}_{t}^{N}\right)_{t \geq 2}$ be the càdlàg version of the process which, informally, follows the proportion of $Y$-individuals among the fittest ones, that is

$$
\mathcal{Y}_{t}^{N}:=\frac{Y_{j(t)-1}\left(\tau_{j(t)}\right)}{\lceil s / \mu\rceil}
$$

Note that the process $\left(\mathcal{Y}_{t}^{N}\right)_{t \geq 2}$ is not Markovian, due to the interactions between the types. Moreover, the individuals counted in $Y_{j}\left(\tau_{j+1}\right)$ need not be children of the individuals in $Y_{j-1}\left(\tau_{j}\right)$

We now state our main result.
Theorem 2.1. Given that $\mathcal{Y}_{2}^{N}=y_{N} \in(0,1)$, for all $T>2$, the process $\left(\mathcal{Y}_{t}^{N}\right)_{t \in[2, T]}$ converges weakly for the Skorokhod topology towards the unique solution of (1.2).

The strong uniqueness of the solution of (1.2) is proven in [8, Theorem 4.1].

Random scaling of the weak selection. The killing probability (2.9) involves deterministic scaling factors $r_{j}, j \geq k_{N}^{*}+1$. Throughout the paper, instead of deriving approximations of $r_{j}$ in every intermediate result, we shall rather use specific random factors to simplify the proofs; when establishing the convergence of Theorem 2.1, we will only need to make the approximation of $r_{j}$ once. We redefine the weak selection mechanism by replacing (2.9) with

$$
\begin{equation*}
\frac{\alpha}{q_{j+1}} \cdot \frac{X_{j}(t)}{X_{j}(t)+Y_{j}(t)} \tag{2.10}
\end{equation*}
$$

where $q_{j+1}$ is a random variable, defined later on in (3.8), that can be understood as the difference between the largest type and the average type in the population at the time when the type $j+1$ start appearing. It is measurable with respect to the natural filtration of the population at this time and we suppose independence of these killings with all randomness after this time. We
will see in the forthcoming Lemma 3.2 that $1 / q_{j+1}$ appears naturally as a scaling of the weak selection in order to observe a non-trivial limit as $N \rightarrow \infty$, and that $r_{j} \approx q_{j}$ in Equation (4.5), such that the two probabilities (2.9) and (2.10) are, in some sense, asymptotically equivalent.

## Organisation of the paper.

- Section 3 contains technical tools for the proof of Theorem 2.1, divided into three subsections:
(1) The first one recalls notation and results from [15] and [16]. Proposition 3.1 describes the evolution of the types and related quantitites, Lemma 3.2 controls the time lapse between the random discretization step $\tau_{j+1}-\tau_{j}$ defined in (2.5). Sometimes, an individual will have a much larger than usual number of descendents born between these times. Lemma 3.3 approximates the law of the size of such a large family.
(2) The second subsection introduces some notation and two new populations whose dynamics need to be described. Informally, the first population is identical to our model except that the most recent killings are cancelled, whereas the second population consists of ( $X, j$ )-individuals descending from a recent killing. They are of interest because one can retrieve the proportion of $Y$-individuals in the original population from these two. Lemma 3.4 shows that the weak selection does not interfere much with the strong selection, in the sense that with high probability, the same individual is not affected by the two mechanisms at the same time.
(3) The third subsection adapts the techniques of Schweinsberg based on martingales to investigate the fluctuations of different subpopulations. It is divided into 5 subsubsections, whose organisation is made precise at the beginning of the subsection. Lemmas 3.5 and 3.6 are technical results that serve to obtain the approximation of the $(Y, j-1)$-individuals when the $(Y, j)$-individuals start appearing in the population; this approximation is made in Lemma 3.7. Lemma 3.8 contains tools to control the numbers of $Y$-individuals as well as the number of individuals descending from killings. They will be used to derive Lemma 3.9 that describes the evolution of the proportion of $Y$-individuals when one individual in the population reproduces much more that the others, due to the strong selection. Lemma 3.10 shows that, in expectation, in the absence of weak selection and as long as no type $j$ individual appears too close to time $\tau_{j}$, the proportion of $Y$-individuals remains constant. The expectation of the effect of the weak selection on the proportions is obtained in Lemma 3.11.
- Section 4 is devoted to the proof of Theorem 2.1. To show the convergence of the process when the killing probability is given by (2.10), the strategy is the following. We first establish the tightness of $\mathcal{Y}^{N}$ in Lemma 4.1. Next, we show in Lemma 4.2 that the expectation of the increment of the proportion of $Y$-individuals from $j$ to $j+1$ is very close to the generator of the solution of (1.2). We then introduce a martingale problem in Lemma 4.3 which states that any weak limit of $\mathcal{Y}^{N}$ solves it. We continue the argument with Lemma 4.4, who states that this weak limit is therefore a solution of (1.2). The section ends with the extension of the convergence to the process when the killing probability is given by (2.9), through a coupling.


## 3. Toolbox

### 3.1. Schweinsberg's setting and notation

In this subsection, we introduce the notation used in $[15,16]$ and we recall some of the results we will need. Thus, what follows does not directly concern the dynamics of the two groups $X$
and $Y$, but rather that of the type distribution. Set $T>2$ a positive real number, arbitrarily large. Fix $\epsilon, \delta \in(0,1)$ such that

$$
\begin{equation*}
\delta<\min \left\{\frac{1}{100}, \frac{1}{19 T}, \epsilon^{3}\right\} \tag{3.1}
\end{equation*}
$$

as required in $\left[16\right.$, Equation (5.1)]. We will study the process up to time $a_{N} T$, and control its behaviour with a probability greater than $1-\epsilon$, with accuracy $\delta$. We shall denote by $C_{i}, i \in \mathbb{N}$, constants that can depend on $\delta, \epsilon, T$ whereas $C$ will always refer to a constant independent of those parameters, that may vary from line to line. We will also keep $C$ independent from the weak selection coefficient $\alpha$ of (2.10).

We introduce some tools to study the evolution of a type. Denote $B_{j}(t), D_{j}(t)$ respectively the birth-rate and death-rate per individual of type $j$ at time $t \leq a_{N} T$, that is:

$$
\begin{align*}
B_{j}(t) & :=\left(N-W_{j}(t)\right) F_{j}(t)  \tag{3.2}\\
D_{j}(t) & :=\mu+1-W_{j}(t) F_{j}(t) \tag{3.3}
\end{align*}
$$

and define

$$
\begin{equation*}
G_{j}(t):=B_{j}(t)-D_{j}(t) \tag{3.4}
\end{equation*}
$$

The value of $G_{j}(t)$ is the growth-rate per individual of type $j$ at time $t$. Thus, as in [15] and [16], we can define for all $j \geq 0$

$$
\begin{equation*}
\left(Z_{j}(t)\right)_{t \in\left[0, a_{N} T\right]}:=\left(e^{-\int_{0}^{t} G_{j}(v) \mathrm{d} v} W_{j}(t)-\int_{0}^{t} \mu W_{j-1}(u) e^{-\int_{0}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u-W_{j}(0)\right)_{t \in\left[0, a_{N} T\right]} \tag{3.5}
\end{equation*}
$$

Let $\left(\mathcal{F}_{t}^{N}\right)_{t \geq 0}$ denote the natural filtration of $\left(W_{j}(t), j \geq 0\right)_{t \geq 0}$. For all $j \geq 0$, the process $\left(Z_{j}(t)\right)_{t \in\left[0, a_{N} T\right]}$ is a square integrable martingale with respect to $\mathcal{F}^{N}$, the variance of which is given for $t \in\left[0, a_{N} T\right]$ by

$$
\begin{equation*}
\operatorname{Var}\left(Z_{j}(t)\right)=\mathbb{E}\left(\int_{0}^{t} e^{-2 \int_{0}^{u} G_{j}(v) \mathrm{d} v}\left(\mu W_{j-1}(u)+B_{j}(u) W_{j}(u)+D_{j}(u) W_{j}(u)\right) \mathrm{d} u\right) \tag{3.6}
\end{equation*}
$$

see [15, Proposition 5.1]. The role of $Z_{j}(t)$ is to control the fluctuations of $W_{j}(t)$ as follows: we rewrite (3.5) as

$$
\left(W_{j}(t)\right)_{t \in\left[0, a_{N} T\right]}:=\left(e^{\int_{0}^{t} G_{j}(v) \mathrm{d} v}\left(W_{j}(0)+Z_{j}(t)\right)+\int_{0}^{t} \mu W_{j-1}(u) e^{\int_{u}^{t} G_{j}(v) \mathrm{d} v} \mathrm{~d} u\right)_{t \in\left[0, a_{N} T\right]}
$$

Then one sees that if $Z_{j}(t)$ is much smaller that $e^{-\int_{0}^{t} G_{j}(v) \mathrm{d} v}$, then describing $W_{j}(t)$ reduces to describing $e^{\int_{0}^{t} G_{j}(v) \mathrm{d} v}$ and $W_{j-1}(u)$ up to time $t$. To show that $Z_{j}$ is small with high probability, the general strategy is to bound its variance given by (3.6). Roughly speaking, $Z_{j}$ is a martingale because $e^{\int_{0}^{t} G_{j}(v) \mathrm{d} v}$ is the expected number of individuals alive at time $t$ in a pure birth process starting from a single individual. Hence, one sees that $e^{-\int_{0}^{t} G_{j}(v) \mathrm{d} v} W_{j}(t)$ would be constant in expectation, in the absence of immigration by mutations. The integral in (3.5) is exactly the term needed to compensate these mutations and their offspring. The variance in (3.6) follows from stochastic calculus, as shown in [15, Section 5].

We will often work with variants of the martingales $Z_{j}$. We will always admit the fact that they are martingales, the reasons being the same as the one sketched above, as well as the formulas for their variances.

In $[15,16]$, Schweinsberg often distinguishes whether $j$ is greater or smaller than $k_{N}^{*}:=$ $\left\lceil k_{N}^{+}-1\right\rceil$, where

$$
\begin{equation*}
k_{N}^{+}:=k_{N}+\frac{2 k_{N} \log k_{N}}{\log (s / \mu)} \tag{3.7}
\end{equation*}
$$

This constant is, roughly speaking, the first type after which the type distribution looks like a Gaussian distribution. For $j \geq k_{N}^{*}+1$, let

$$
\begin{align*}
& q_{j}^{*}:= \begin{cases}j-k_{N} & \text { if } a_{N}-2 a_{N} / k_{N} \leq \tau_{j} \leq a_{N}+2 a_{N} / k_{N}, \\
j-M\left(\tau_{j}\right) & \text { otherwise. }\end{cases} \\
& q_{j}:=\max \left\{1, q_{j}^{*}\right\} . \tag{3.8}
\end{align*}
$$

As we mentioned before, all the individuals have type 0 at time 0 and the wave dynamics starts approximately around time $a_{N}$, which is why the condition in the definition of $q_{j}^{*}$ is needed in $[15,16]$. However, we will be mainly interested in types $j \geq j(2)$ for which $q_{j}^{*}=j-M\left(\tau_{j}\right)$. Recall that $\tau_{j}$ defined in (2.5) is approximately the time where one expects to see the first type $j$ mutations, hence $q_{j}$ is an approximation of the difference between $j$ and the average number of mutations when individuals of type $j$ start appearing. Set

$$
\begin{equation*}
b:=\log \frac{24000 T}{\delta^{2} \epsilon} . \tag{3.9}
\end{equation*}
$$

For $j \geq k_{N}^{*}+1$, define

$$
\begin{equation*}
\xi_{j}:=\max \left\{\tau_{j}, \tau_{j}+\frac{1}{s q_{j}} \log \left(\frac{1}{s q_{j}}\right)+\frac{b}{s q_{j}}\right\} . \tag{3.10}
\end{equation*}
$$

We shall work on a specific event, realized with high probability, such that it holds that $\tau_{j}<$ $\xi_{j}<\tau_{j+1}$ for all $j \geq k_{N}^{*}+1$ such that $\tau_{j+1}<a_{N} T$. The goal of $\xi_{j}$ is to distinguish whether a mutation is faster than usual: we call a type $j$ mutation an early type $j$ mutation if it occurs in the time interval $\left[\tau_{j}, \xi_{j}\right]$. The fitness being relative to the mean, the earlier a mutant is, the stronger is its advantage to reproduce immediately after the mutation. The individual acquiring an early type $j$ mutation, as well as its offspring, are called early type $j$ individuals. In general, we will speak of early type $j$ individuals during the time interval $\left[\tau_{j}, \tau_{j+1}\right]$ such that (on the high probability event we consider) they still have type $j$ when thus called. Schweinsberg showed that large families appear with the Bolthausen-Sznitman rates as a result of early mutations.

In $[15,16], \zeta=\zeta_{N}$ denotes a stopping time that is essentially the first time when the type distribution behaves atypically, i.e. differently from its large population limiting behavior. For example, it is defined such that before $\zeta$ and for some time after $\tau_{j}$, the dynamics of the number of type $j$ individuals has a specific exponential growth, see Proposition 3.1 (3) below). Its definition requires a lot of technical considerations that are not relevant for our purposes and we will merely recall the properties that we will need and that hold before time $\zeta$. The precise definition of $\zeta$ is given in [15, Section 3.3]. In particular, for $N$ large enough, it holds that $\mathbb{P}\left(\zeta>a_{N} T\right)>1-\epsilon$. Throughout the paper we will say that a property holds on some event $E$ if it is true for $\mathbb{P}$-almost every $\omega \in E$. Similarly, if we say that on the event $E, \mathbb{E}(\beta) \leq c$ for some random variable $\beta$ and some constant $c$, we mean that $\mathbb{E}(\beta \mid E) \leq c$.

We shall work on $\left\{\zeta>a_{N} T\right\}$ so that the properties of the next proposition hold. Besides, the listed properties in Proposition 3.1 below always hold under the conditions given in their respective statements, e.g. for all $j \geq k_{N}^{*}+1$ such that $\tau_{j+1} \leq \zeta \wedge a_{N} T$ means for all $\omega \in\left\{\tau_{j+1} \leq\right.$ $\left.\zeta \wedge a_{N} T\right\}$. The results it gathers are from [15] and [16] as follows:

- (1) is taken from both Proposition 3.3 (1) and Proposition 3.6 (3) in [15].
- (2) is taken from Proposition 3.3(2) in [15].
- (3) is taken from Proposition 3.3 (3) [15].
- (4) is taken from Proposition 4.4(1)-(3) in [16].
- (5) is taken from Lemma 4.5 in [15].


## Moran model with simultaneous strong and weak selections

Proposition 3.1. Recall that $k_{N}^{*}+1$ defined above Equation (3.7) is the first type we are interested in, and for all $j \geq 1, \tau_{j}$ defined in (2.5) is roughly the time when type $j$ start appearing in the population. The real numbers $\epsilon, \delta \in(0,1)$ do not depend on $N$ and were fixed such that (3.1) is satisfied. For $N$ large enough, the following hold:
(1) For all $j \geq k_{N}^{*}+1$ such that $\tau_{j+1} \leq \zeta \wedge a_{N} T$, no early type $j$ individual acquires a $j+1$-th mutation before time $\tau_{j+1}$. Furthermore, it holds that

$$
\frac{a_{N}}{3 k_{N}} \leq \tau_{j+1}-\tau_{j} \leq \frac{2 a_{N}}{k_{N}}
$$

and on $\left\{\zeta>a_{N} T\right\}$, we have $\tau_{J+1}>a_{N} T$ for $J:=3 T k_{N}+k_{N}^{*}+1$, so the types greater or equal to $J+1$ have not appeared at time $a_{N} T$ yet.
(2) For all $j \geq k_{N}^{*}+1$ and $t \in\left[\tau_{j}+\frac{a_{N}}{4 T k_{N}}, \tau_{j+1}\right] \cap\left[0, \zeta \wedge a_{N} T\right]$ :

$$
(1-4 \delta) e^{\int_{\tau_{j}}^{t} G_{j}(v) \mathrm{d} v} \leq \widetilde{W}_{j}(t) \leq(1+4 \delta) e^{\int_{\tau_{j}}^{t} G_{j}(v) \mathrm{d} v}
$$

where $\widetilde{W}_{j}$ denote the number of non-early type $j$. Moreover, the upper bound holds for all $t \in\left[\xi_{j}, \tau_{j+1}\right] \cap\left[0, \zeta \wedge a_{N} T\right]$, where $\xi_{j}$ was defined in (3.10).
(3) For all $j \geq k_{N}^{*}+1$ and $t \in\left[\tau_{j+1}, \tau_{\left\lceil j+\frac{k_{N}}{4}\right\rceil}+a_{N}\right] \cap\left[0, \zeta \wedge a_{N} T\right]$ :

$$
(1-\delta) \frac{s}{\mu} e^{\int_{\tau_{j+1}}^{t} G_{j}(v) \mathrm{d} v} \leq W_{j}(t) \leq(1+\delta) \frac{s}{\mu} e^{\int_{\tau_{j+1}}^{t} G_{j}(v) \mathrm{d} v}
$$

(4) For all $j \geq k_{N}^{*}+1$ and $t \in\left[\tau_{j}, \tau_{j+1}\right] \cap\left[0, \zeta \wedge a_{N} T\right]$ :

$$
\begin{aligned}
& s\left(q_{j}-C_{1}\right) \leq G_{j}(t) \leq s\left(q_{j}+C_{1}\right), \\
& s k_{N}(1-2 \delta) \leq G_{j}(t) \leq s k_{N}(e+2 \delta), \\
& k_{N}(1-2 \delta) \leq q_{j} \leq k_{N}(e+2 \delta),
\end{aligned}
$$

for some constant $C_{1}>0$.
(5) For all $j \geq k_{N}^{*}+1$ such that $\tau_{j+1} \leq \zeta \wedge a_{N} T$, we have

$$
\frac{s}{C_{2} \mu} \leq e^{\int_{\tau_{j}}^{\tau_{j}+1} G_{j}(v) \mathrm{d} v} \leq \frac{2 s}{\mu},
$$

for some constant $C_{2}>0$.
In [15, Proposition 3.6(1)] shows that on $\left\{\zeta>a_{N} T\right\}, \tau_{k_{N}^{*}+1} \leq 2 a_{N} / k_{N}$ so that $\tau_{j(2)}>\tau_{k_{N}^{*}+1}$. As explained in the introduction, it will be more convenient for us to study the process starting at time $\tau_{j(2)}$. We also note that by Proposition 3.1(1) above and Assumption $\left(A_{3}\right)$, on the event $\left\{\zeta>a_{N} T\right\}$, all the fitnesses of the individuals until time $a_{N} T$ are positive, i.e. (2.1) holds, and therefore

$$
\begin{equation*}
G_{j}(t)=s(j-M(t))-\mu, \quad \forall t \leq a_{N} T \text { and } j \leq J, \tag{3.11}
\end{equation*}
$$

where $J:=3 T k_{N}+k_{N}^{*}+1$ is from (1) of the above proposition. Moreover, for $N$ large enough, for all $t \leq a_{N} T$ and every $j \leq J$, on the event $\{\zeta>t\}$, one has

$$
\begin{align*}
B_{j}(t)+D_{j}(t) & =\frac{\left(N-2 W_{j}(t)\right)(1+s(j-M(t)))}{N}+1+\mu \leq 2+s J+\mu \\
& \leq 3 \tag{3.12}
\end{align*}
$$

by assumption $\left(A_{3}\right)$.

In [16], the study of the process backwards in time requires to consider only types $j$ 's that belong to some set $I \subset \mathbb{N}$, defined just before Lemma 6.2 in [16]. Its definition involves a fixed parameter $t_{0} \in(T-37, T-2)$. Choosing $t_{0}=T-3$, one gets

$$
I=\left\{j_{1}, \ldots, j_{2}\right\}
$$

with

$$
\begin{align*}
j_{1} & :=\max \left\{j: \tau_{j}^{*} \leq 2 a_{N}\right\}-\left\lfloor 9 \delta T k_{N}\right\rfloor  \tag{3.13}\\
j_{2} & :=\max \left\{j: \tau_{j}^{*} \leq a_{N}\left(T-1+19 / k_{N}\right)\right\}+\left\lfloor 9 \delta T k_{N}\right\rfloor
\end{align*}
$$

where the $\tau_{j}^{*}$ 's are some deterministic times, approximating the random $\tau_{j}$ 's (see [16, Equation (6.1)]). The relevant informations for our purposes are given by Lemma 6.2 in [16], which shows that on the event $\left\{\zeta>a_{N} T\right\}$, it holds that $\tau_{j_{1}}<2 a_{N}$, and $j_{2} \geq L+9$, where $L$ is defined in Lemma 5.1 of [16] as

$$
L:=\inf \left\{j: \tau_{j} \geq a_{N}(T-1)-3 a_{N} / k_{N}\right\}
$$

It entails that $\tau_{j_{2}} \geq \tau_{L}+9 a_{N} / 3 k_{N}$ by Proposition $3.1(1)$. Hence, $\tau_{j_{2}}>a_{N}(T-1)$. We thus have that on the event $\left\{\zeta>a_{N} T\right\}, j(2) \in I$ and $j(T-1) \in I$, so that for $j(2) \leq j \leq j(T-1)$, we can use the results of Schweinsberg proven for $j \in I$, since then $2 a_{N} \leq \tau_{j} \leq a_{N}(T-1)$. In particular, on the event $\left\{\zeta>a_{N} T\right\}$, the estimates in Proposition 3.1 hold for $j \in I$ and we will thus apply the proposition for $j \in I$ without recalling that this ensures $\tau_{j} \leq a_{N} T$.

We deduce a result on the time length between $\tau_{j}$ and $\tau_{j+1}$ that will be useful later on.
Lemma 3.2. For all $j \in I$, conditionally given $\mathcal{F}_{\tau_{j}}^{N}$ and on the event $\left\{\zeta>\tau_{j+1}\right\}$, it holds that

$$
\frac{1-2 \delta}{q_{j}} \leq \frac{\tau_{j+1}-\tau_{j}}{a_{N}} \leq \frac{1+2 \delta}{q_{j}}
$$

Proof. By Proposition 3.1(4), we know that $\sup _{t \in\left[\tau_{j}, \tau_{j+1}\right]}\left|G_{j}(t)-s q_{j}\right| \leq s C_{1}$. Equation (8.32) in [16] states that

$$
(1-\delta) a_{N} \leq \int_{\tau_{j}}^{\tau_{j+1}} \frac{G_{j}(v)}{s} \mathrm{~d} v \leq(1+\delta) a_{N}
$$

Therefore we have that

$$
\frac{1-\delta}{q_{j}+C_{1}} \leq \frac{\tau_{j+1}-\tau_{j}}{a_{N}} \leq \frac{1+\delta}{q_{j}-C_{1}}
$$

which implies for $N$ large enough that

$$
\frac{1-2 \delta}{q_{j}} \leq \frac{\tau_{j+1}-\tau_{j}}{a_{N}} \leq \frac{1+2 \delta}{q_{j}}
$$

since by Proposition $3.1(4), q_{j} \geq(1-2 \delta) k_{N} \rightarrow \infty$ as $N \rightarrow \infty$.
We conclude the first subsection of the toolbox with a reformulation of the result of Schweinsberg in [16] showing that the law of the number of early type $j$ individuals at time $\tau_{j+1}$ can be well approximated by the rates corresponding to Bolthausen-Sznitman's coalescent.

Lemma 3.3. For $N$ large enough, for all $j \in I, j \geq j(2)$, conditionally given $\mathcal{F}_{\tau_{j}}^{N}$ and on the event $\left\{\zeta>\tau_{j+1}\right\}$, for any $g \in \mathcal{C}^{\infty}([0,1])$, it holds that

$$
\left|q_{j} \int_{(\epsilon, 1)} g(x) p_{S_{j}}(\mathrm{~d} x)-\int_{\epsilon}^{1} g(x) \frac{\mathrm{d} x}{x^{2}}\right| \leq C \epsilon\left(\|g\|_{\infty}+\left\|g^{\prime}\right\|_{\infty}\right)
$$

where $S_{j}$ is the proportion of early type $j$ individuals at time $\tau_{j+1}$ among the type $j$ individuals and $p_{S_{j}}$ its probability distribution supported on $\{0,1 /\lceil s / \mu\rceil, \ldots, 1\}$.

Proof. Let $\nu(\mathrm{d} x)=\mathrm{d} x / x^{2}, x \in(0,1]$. Lemma 7.8 in [16] shows that for all $y \in(\epsilon, 1-\delta]$, it holds that

$$
\begin{equation*}
\left|q_{j} p_{S_{j}}((y, 1])-\nu((y, 1))\right| \leq 14 \delta \nu((y, 1]) \leq 14 \frac{\delta}{\epsilon} . \tag{3.14}
\end{equation*}
$$

We implicitely used that the event in [16, Equation (7.48)] has probability going to 1 as $N \rightarrow \infty$, see Lemmas 7.4 and 7.7 of the same paper. Roughly speaking, on this event the early mutants are coupled with a branching process introduced in Section 7.2 of the same paper, allowing to approximate the law of $S_{j}$. We write

$$
\begin{aligned}
& \left|q_{j} \int_{(\epsilon, 1)} g(x) p_{S_{j}}(\mathrm{~d} x)-\int_{\epsilon}^{1} g(x) \nu(\mathrm{d} x)\right|=\left|\int_{\epsilon}^{1} g^{\prime}(y)\left(q_{j} p_{S_{j}}-\nu\right)((y, 1])-g(\epsilon)\left(q_{j} p_{S_{j}}-\nu\right)((\epsilon, 1]) \mathrm{d} y\right| \\
& \left.\quad \leq 14 \frac{\delta}{\epsilon}\left\|g^{\prime}\right\|_{\infty}+\left\|g^{\prime}\right\|_{\infty} \int_{1-\delta}^{1} \mathrm{~d} y\left(q_{j} p_{S_{j}}+\nu\right)((1-\delta, 1])\right)+14 \frac{\delta}{\epsilon}\|g\|_{\infty} \leq C \frac{\delta}{\epsilon}\left(\|g\|_{\infty}+\left\|g^{\prime}\right\|_{\infty}\right) .
\end{aligned}
$$

We conclude using that $\delta / \epsilon<\epsilon$ by (3.1).

### 3.2. Splitting strategy to study the weak selection

Recall that, among the type $j(2)-1$ individuals at time $\tau_{j(2)}$, we assigned $y_{N}\lceil s / \mu\rceil$ of them to group $Y$, and $\left(1-y_{N}\right)\lceil s / \mu\rceil$ to group $X$, with the weak selection mechanism explained above Theorem 2.1. One sees when $\alpha=0$ that

$$
\left\{W_{j}(t): t \leq a_{N} T, j \geq 0\right\}=\left\{X_{j}(t)+Y_{j}(t): t \leq a_{N} T, j \geq 0\right\}
$$

is exactly the model of Schweinsberg. Moreover, when $\alpha \neq 0$, the type distribution remains unchanged (only the genealogy is altered). To make the proofs, we will study the fluctuations of each group as if there was no weak selection and then combine it with estimates on the number of killings. We thus introduce the notation $\check{Y}_{j}(t)$ for $j \in I, j \geq j(2)$ and $t \in\left[\tau_{j}, \tau_{j+1}\right]$ for the number of $(Y, j)$-individuals if we had cancelled the killings of the weak selection previously described between $\left[\tau_{j}, \tau_{j+1}\right]$, and only those ones ${ }^{4}$. In particular, for $t \in\left[\tau_{j}, \tau_{j+1}\right]$, denoting $\check{X}_{j}(t)$ the total number of individuals at time $t$ descending from killings between $\left[\tau_{j}, t\right]$, one can write $Y_{j}(t)=\check{Y}_{j}(t)-\check{X}_{j}(t)$. Hence, our strategy is to control $\check{X}_{j}(t)$ and $\check{Y}_{j}(t)$ separately before combining them to obtain control on $Y_{j}(t)$.

To obtain (1.2), the weak selection should have asymptotically no effect on the strong selection.
Schweinsberg in [15] is able to couple the early type $j$ individuals and their progeny with a branching process, for a certain amount of time. This allows him to bound the probability that an early mutation survives. Let $E_{j}$ be the event that a $Y$-individual is killed by the weak selection during an early type $j$ mutation in $\left[\tau_{j}, \xi_{j}\right]$, and that the resulting $(X, j)$-individual has descendents that are alive at time $\tau_{j+1}$. We complete the filtration $\mathcal{F}^{N}$ to take into account the groups $X$ and $Y$ of the individuals. The following lemma shows that these problematic events occur with negligible probabilities.

Lemma 3.4. For all $j \in I, j \geq j(2)$, on the event $\left\{\zeta>\tau_{j}\right\}$, it holds that

$$
\mathbb{P}\left(E_{j} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq \frac{C}{\epsilon q_{j}^{2}} .
$$

Proof. By independence, the probability of $E_{j}$ is the product of the probabilities of a surviving early mutation and a killing, the former being upper bounded by Lemma 7.8 in [16] (this Lemma

[^2]actually bounds the probability of survival up to a time $\tau_{j}^{\prime}$, that is anyway smaller than $\tau_{j+1}$ on the event $\left\{\zeta>\tau_{j+1}\right\}$ ). By combining this bound and (2.10), we obtain
$$
\mathbb{P}\left(E_{j} \mid \mathcal{F}_{\tau_{j}}^{N},\left\{\zeta>\tau_{j}\right\}\right) \leq C \frac{1}{\epsilon q_{j}} \frac{\alpha}{q_{j}} \leq \frac{C \alpha}{\epsilon q_{j}^{2}},
$$
as claimed.
From Lemma 3.4, we see that
$$
\mathbb{P}\left(\bigcup_{\substack{j \in I \\ j \geq j(2)}} E_{j} \mid\left\{\zeta>a_{N} T\right\}\right) \leq \sum_{\substack{j \in I \\ j \geq j(2)}} \frac{C \alpha}{\epsilon q_{j}^{2}} \leq \frac{C \alpha T}{\epsilon k_{N}} \rightarrow 0,
$$
as $N \rightarrow \infty$, where we used that the number of elements in $I$ is smaller than $J \leq 4 k_{N} T$, see the discussion after (3.13). Hence, we redefine $\zeta$ to include the first time at which an event $E_{j}$ occurs and with this new definition, one can still choose $N$ large enough such that $\mathbb{P}\left(\zeta>a_{N}(T-1)\right)>$ $1-\epsilon$, in particular no $E_{j}$ occurs for any $j \in I$ with high probability.

### 3.3. Expected fluctuations of the proportions

We divide this section into 5 parts: in the first subsection, we will look at the effect of the weak selection on type $(j-1)$, that is the second fittest type during $\left[\tau_{j}, \tau_{j+1}\right]$ when the fittest type $j$ starts building up. Then, in the second subsection, we will study the non-early type $(\check{Y}, j)$ individuals during $\left[\xi_{j}, \tau_{j+1}\right]$ In the third subsection we will describe the impact of an early mutation on the proportion of $\check{Y}$ individuals. In the fourth subsection, we will introduce the discrete process indexed by $j$ following the proportion of $(Y, j)$-individuals at time $\tau_{j}$. The importance of weak selection, that is the expected number of killings of $(Y, j)$-individuals occurring in $\left[\tau_{j}, \tau_{j+1}\right]$, will be discussed in the fifth subsection.

### 3.3.1. The type $j-1$ during $\left[\tau_{j}, \tau_{j+1}\right]$

Let $\check{X}_{j-1}^{\tau_{j}}(t)$ be the number of $(X, j-1)$-individuals at time $t \geq \tau_{j}$ descending from a killing that occurred after time $\tau_{j}$. In order to properly estimate $\check{X}_{j}\left(\tau_{j+1}\right)$, one needs to control $\breve{X}_{j-1}^{\tau_{j}}(t)$ for all $t \in\left[\tau_{j}, \tau_{j+1}\right]$, since type $j$ individuals can come from mutants of these type $j-1$ individuals. The next lemma enables us to do so.

Lemma 3.5. For $N$ large enough, on the event $\left\{\zeta>\tau_{j+1}\right\}$, for all $j \in I, j \geq j(2)$, with probability $1-o(\sqrt{\mu})$, it holds that

$$
\sup _{t \in\left[\tau_{j}, \tau_{j+1}\right]} \frac{\check{X}_{j-1}^{\tau_{j}}(t)}{W_{j-1}(t)} \leq \frac{3 \alpha}{q_{j-1}} .
$$

Proof. We admit the two following statements without proof, referring to [15, Section 5] for details on how to prove them: For all $j \in I, j \geq j(2)$, we have that

- the process defined for $t<\tau_{j}$ by $\check{Z}_{j-1}^{\tau_{j}}(t)=0$ and for $t \geq \tau_{j}$ by

$$
\begin{equation*}
\check{Z}_{j-1}^{\tau_{j}}(t):=e^{-\int_{\tau_{j}}^{t} G_{j-1}(v) \mathrm{d} v} \check{X}_{j-1}^{\tau_{j}}(t)-\int_{\tau_{j}}^{t} e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u) Y_{j-2}(u)}{W_{j-2}(u)} \mathrm{d} u \tag{3.15}
\end{equation*}
$$

[^3]
## Moran model with simultaneous strong and weak selections

- its conditional variance is given by

$$
\begin{align*}
\operatorname{Var}\left(\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta) \mid \mathcal{F}_{\tau_{j}}^{N}\right) & =\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} \mathrm{~d} u e^{-2 \int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v}\right. \\
& \left.\left.\times\left(\mu \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u) Y_{j-2}(u)}{W_{j-2}(u)}+\left(B_{j-1}(u)+D_{j-1}(u)\right) \check{X}_{j-1}^{\tau_{j}}(u)\right) \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \tag{3.16}
\end{align*}
$$

The first step of the proof is to bound (3.16). Since $X_{j-2}(t)$ and $Y_{j-2}(t)$ are smaller than $W_{j-2}(t)$ by definition, Proposition 3.1 (3) then (5) entail that

$$
\begin{aligned}
& \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} \mathrm{~d} u e^{-2 \int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u) Y_{j-2}(u)}{W_{j-2}(u)} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} \mathrm{~d} u e^{-2 \int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v}(1+\delta) s \frac{\alpha}{q_{j-1}} e^{\int_{\tau_{j-1}}^{u} G_{j-2}(v) \mathrm{d} v} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} \mathrm{~d} u(1+\delta) \frac{2 s^{2}}{\mu} \frac{\alpha}{q_{j-1}} e^{-\int_{\tau_{j}}^{u} G_{j}(v) \mathrm{d} v} e^{-s\left(\tau_{j}-\tau_{j-1}\right)} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right),
\end{aligned}
$$

where we used that $G_{j-1}(v)=G_{j-2}(v)+s$ for all $v \in\left[\tau_{j}, \tau_{j+1} \wedge \zeta\right]$ (3.11). Proposition 3.1(4) allows us to write $\int_{\tau_{j} \wedge \zeta}^{t \wedge \zeta} e^{-\int_{\tau_{j}}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u \leq \frac{1}{s\left(q_{j}-C_{1}\right)}\left(1-e^{-s\left(q_{j}-C_{1}\right)\left(t \wedge \zeta-\tau_{j} \wedge \zeta\right)}\right) \leq \frac{2}{s q_{j}}$ since $s \rightarrow 0$ as $N \rightarrow \infty$. Hence, we have that

$$
\begin{equation*}
\mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} \mathrm{~d} u e^{-2 \int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v}{ }_{\mu} \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u) Y_{j-2}(u)}{W_{j-2}(u)} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \leq C \alpha \frac{s}{\mu} \frac{e^{-s\left(\tau_{j}-\tau_{j-1}\right)}}{q_{j} q_{j-1}} . \tag{3.17}
\end{equation*}
$$

On the other hand, $B_{j-1}(u)+D_{j-1}(u) \leq 3$ for all $u \geq \tau_{j-1}$ by (3.12). We apply Proposition $3.1(4)$, then use (3.15) and the martingale property of $\check{Z}_{j-1}^{\tau_{j}}$ to write

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} e^{-2 \int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v}\left(B_{j-1}(u)+D_{j-1}(u)\right) \check{X}_{j-1}^{\tau_{j}}(u) \mathrm{d} u \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \quad \leq 3 \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j}\right\}} \int_{\tau_{j}}^{t \wedge \zeta} e^{-s\left(q_{j}-C_{1}-1\right)\left(u-\tau_{j}\right)} e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \check{X}_{j-1}^{\tau_{j}}(u) \mathrm{d} u \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq 3 \int_{\tau_{j}}^{t} \mathrm{~d} u e^{-s\left(q_{j}-C_{1}-1\right)\left(u-\tau_{j}\right)} \\
& \quad \times \mathbb{E}\left(\left.\check{Z}_{j-1}^{\tau_{j}}(u \wedge \zeta)+\mathbb{1}_{\{\zeta>t\}} \int_{\tau_{j}}^{u} \mathrm{~d} r e^{-\int_{\tau_{j}}^{r} G_{j-1}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(r) Y_{j-2}(r)}{W_{j-2}(r)} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \quad \leq 3 \int_{\tau_{j}}^{t} \mathrm{~d} u e^{-s\left(q_{j}-C_{1}-1\right)\left(u-\tau_{j}\right)} \mathbb{E}\left(\left.\mathbb{1}_{\{\zeta>t\}} \int_{\tau_{j}}^{u} \mathrm{~d} r e^{-\int_{\tau_{j}}^{r} G_{j-1}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(r) Y_{j-2}(r)}{W_{j-2}(r)} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \quad \leq \frac{C \alpha}{s q_{j}} \frac{s}{\mu} \frac{e^{-s\left(\tau_{j}-\tau_{j-1}\right)}}{q_{j} q_{j-1}},
\end{aligned}
$$

where the last inequality follows from (3.17). Thus, applying Proposition 3.1 (5) and coming back to (3.16), we have shown thanks to (3.17) and $\left(A_{3}\right)$ that

$$
\operatorname{Var}\left(\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq C \alpha \frac{s}{\mu} \frac{e^{-s\left(\tau_{j}-\tau_{j-1}\right)}}{s k_{N}^{3}} .
$$

Proposition $3.1(1)$ gives $s\left(\tau_{j}-\tau_{j-1}\right) \geq s \frac{a_{N}}{3 k_{N}}=\log (s / \mu) / 3 k_{N}$, which tends to $\infty$ as $N \rightarrow \infty$ as a consequence of assumption $\left(A_{2}\right)$ and $k_{N} \rightarrow \infty$. In particular, on $\left\{\zeta>\tau_{j}\right\}$, it holds that

$$
\begin{equation*}
e^{-s\left(\tau_{j}-\tau_{j-1}\right)}=o\left(\frac{1}{k_{N}}\right) \tag{3.18}
\end{equation*}
$$

Therefore, Doob's maximal inequality for square integrable martingales yields that

$$
\begin{equation*}
\mathbb{P}\left(\left.\sup _{t \in\left[\tau_{j}, \tau_{j+1}\right]} \frac{\left|\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta)\right|}{\lceil s / \mu\rceil}>\frac{1}{k_{N}^{2}} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \leq C \alpha \frac{\mu}{s^{2}} k_{N} e^{-s\left(\tau_{j}-\tau_{j-1}\right)}=o\left(\mu^{1 / 2}\right), \tag{3.19}
\end{equation*}
$$

since $\mu \ll s^{a}$ for any $a>0$ by (2.3). Then, we use Proposition 3.1(3) to write

$$
\begin{aligned}
\frac{\check{X}_{j-1}^{\tau_{j}}(t \wedge \zeta)}{W_{j-1}(t \wedge \zeta)} & \leq \frac{1}{1-\delta} e^{-\int_{\tau_{j}}^{t \wedge \zeta} G_{j-1}(v) \mathrm{d} v} \frac{\check{X}_{j-1}^{\tau_{j}}(t \wedge \zeta)}{\lceil s / \mu\rceil} \\
& =\frac{1}{1-\delta}\left(\frac{\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta)}{\lceil s / \mu\rceil}+\int_{\tau_{j}}^{t \wedge \zeta} e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u) Y_{j-2}(u)}{W_{j-2}(u)\lceil s / \mu\rceil} \mathrm{d} u\right) \\
& \leq \frac{1}{1-\delta}\left(\frac{\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta)}{\lceil s / \mu\rceil}+\mu e^{\int_{\tau_{j-1}}^{\tau_{j}} G_{j-1}(v) \mathrm{d} v} \frac{\alpha}{q_{j-1}} \int_{\tau_{j}}^{t \wedge \zeta} e^{-s\left(u-\tau_{j}\right)} \frac{X_{j-2}(u) Y_{j-2}(u)}{W_{j-2}(u)^{2}(1-\delta)} \mathrm{d} u\right) \\
& \leq \frac{1}{1-\delta}\left(\frac{\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta)}{\lceil s / \mu\rceil}+2 s \frac{\alpha}{q_{j-1}} \int_{\tau_{j}}^{t \wedge \zeta} \frac{e^{-s\left(u-\tau_{j}\right)}}{1-\delta} \mathrm{d} u\right)
\end{aligned}
$$

where we used Proposition 3.1(5), we then obtain

$$
\leq \frac{1}{1-\delta}\left(\frac{\check{Z}_{j-1}^{\tau_{j}}(t \wedge \zeta)}{\lceil s / \mu\rceil}+\frac{2}{1-\delta} \frac{\alpha}{q_{j-1}}\right)
$$

This and (3.19) together conclude the proof.
The following lemma will allow us to control the fluctuations of the $(Y, j+1)$-individuals by controlling that of the $(Y, j)$-individuals between $\left[\tau_{j+1}, \tau_{j+2}\right]$.
Lemma 3.6. For all $j \in I, j \geq j(2)$, the following process is a square integrable martingale:

$$
\begin{aligned}
& \left(Z_{j}^{Y}(t)\right)_{t \in\left[\tau_{j+1}, \tau_{j+2}\right)} \\
& \quad:=Y_{j}(t) e^{-\int_{\tau_{j+1}}^{t} G_{j}(v) \mathrm{d} v}-Y_{j}\left(\tau_{j+1}\right)-\int_{\tau_{j+1}}^{t}\left(1-\frac{\alpha}{q_{j-1}} \frac{X_{j-1}(u)}{W_{j-1}(u)}\right) \mu Y_{j-1}(u) e^{-\int_{\tau_{j+1}}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u .
\end{aligned}
$$

Moreover, for $N$ large enough, for all $t \in\left[\tau_{j+1}, \tau_{j+2} \wedge \zeta\right)$, one has the following upper bound for its conditional variance:

$$
\operatorname{Var}\left(Z_{j}^{Y}(t) \mid \mathcal{F}_{\tau_{j+1}}^{N}\right) \leq \frac{21}{\mu k_{N}}
$$

Proof. Again, we admit that $Z_{j}^{Y}$ is a square integrable martingale. We have $\frac{\alpha}{q_{j-1}} \frac{X_{j-1}(u)}{W_{j-1}(u)} \leq \frac{\alpha}{q_{j-1}}$, which tends to 0 as $N \rightarrow \infty$ by Proposition 3.1(4). One then gets the upper bound for the variance as a direct consequence of Lemma 9.27 in [15] (the process $Z^{\prime \prime}$ used in that Lemma is introduced in [15, p. 85, below Equation (9.82)]).

The next lemma shows that the evolution of $Y_{j-1}\left(\tau_{j}+t\right)$ until $\tau_{j+1} \wedge \zeta$ remains predictable.
Lemma 3.7. For all $j \in I, j \geq j(2)+1$, on the event $\left\{\zeta>\tau_{j}\right\}$, it holds that

$$
\left.\begin{array}{rl}
\mathbb{P}\left(\left|Y_{j-1}(t)-Y_{j-1}\left(\tau_{j}\right) e^{\int_{\tau_{j}}^{t} G_{j-1}(v) \mathrm{d} v}\right| \leq \delta\lceil s / \mu] e^{\int_{\tau_{j}}^{t} G_{j-1}(v) \mathrm{d} v},\right. & \forall t \in\left[\tau_{j}, \tau_{j+1} \wedge \zeta\right] \mid \mathcal{F}_{\tau_{j}}^{N}
\end{array}\right) .
$$

## Moran model with simultaneous strong and weak selections

Furthermore, the same statement holds with $X_{j-1}$ instead of $Y_{j-1}$.
Proof. The statements (2), (3), (4) of Proposition 3.1 hold up to time $t$ on the event $\{\zeta>t\}$. Write

$$
\begin{equation*}
Y_{j-1}(t)=e^{\int_{\tau_{j}}^{t} G_{j-1}(v) \mathrm{d} v}\left(Y_{j-1}\left(\tau_{j}\right)+\int_{\tau_{j}}^{t}\left(1-\frac{\alpha}{q_{j}} \frac{X_{j-2}(u)}{W_{j-2}(u)}\right) \mu Y_{j-2}(u) e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \mathrm{~d} u+Z_{j-1}^{Y}(t)\right) . \tag{3.20}
\end{equation*}
$$

We shall bound the two last terms in the above parentheses. By Proposition 3.1 (3), one has

$$
\left(1-\frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u)}{W_{j-2}(u)}\right) \mu Y_{j-2}(u) e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \leq(1+\delta) s e^{\int_{\tau_{j-1}}^{u} G_{j-2}(v) \mathrm{d} v} e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v}
$$

now remark that $G_{j-1}(v)=G_{j-2}(v)+s$ for all $v \in\left[\tau_{j}, t\right]$ (3.11) and obtain

$$
=(1+\delta) s e^{\int_{\tau_{j-1}}^{\tau_{j}} G_{j-1}(v) \mathrm{d} v} e^{-s\left(u-\tau_{j-1}\right)},
$$

then using Proposition 3.1(5),

$$
\leq(1+\delta) \frac{2 s^{2}}{\mu} e^{-s\left(u-\tau_{j-1}\right)}
$$

Hence,

$$
\int_{\tau_{j}}^{t}\left(1-\frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u)}{W_{j-2}(u)}\right) \mu Y_{j-2}(u) e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \mathrm{~d} u \leq(1+\delta) \frac{2 s}{\mu}\left(e^{-s\left(\tau_{j}-\tau_{j-1}\right)}-e^{-s\left(t-\tau_{j-1}\right)}\right) .
$$

By (3.18), we thus have shown that for $N$ large enough, on the event $\{\zeta>t\}$,

$$
\begin{equation*}
\int_{\tau_{j}}^{t}\left(1-\frac{\alpha}{q_{j-1}} \frac{X_{j-2}(u)}{W_{j-2}(u)}\right) \mu Y_{j-2}(u) e^{-\int_{\tau_{j}}^{u} G_{j-1}(v) \mathrm{d} v} \mathrm{~d} u \leq 3 \frac{s}{\mu} \times o\left(1 / k_{N}\right) \tag{3.21}
\end{equation*}
$$

Furthermore, using Lemma 3.6 and Doob's maximal inequality for squared integrable martingales, one has

$$
\begin{aligned}
\mathbb{P}\left(\left.\sup _{t \in\left[\tau_{j}, \tau_{j+1}\right)}\left|Z_{j-1}^{Y}(t \wedge \zeta)\right|>\frac{\delta}{2}\lceil s / \mu\rceil \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) & \leq \frac{4}{\delta^{2}} \operatorname{Var}\left(Z_{j-1}^{Y}\left(\tau_{j+1} \wedge \zeta\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq \frac{4 \mu^{2}}{\delta^{2} s^{2}} \frac{21}{\mu k_{N}}=o\left(\frac{1}{k_{N}}\right)
\end{aligned}
$$

by (2.3). Combined with (3.20) and (3.21), this shows that conditionally given $\zeta>\tau_{j}$, with probability $1-o\left(1 / k_{N}\right)$, it holds that

$$
\left|Y_{j-1}(t)-Y_{j-1}\left(\tau_{j}\right) e^{\int_{\tau_{j}}^{t} G_{j-1}(v) \mathrm{d} v}\right| \leq \delta\lceil s / \mu] e^{\int_{\tau_{j}}^{t \wedge \zeta} G_{j-1}(v) \mathrm{d} v}, \quad \forall t \in\left[\tau_{j}, \tau_{j+1} \wedge \zeta\right]
$$

which completes the proof of the statement for $Y_{j-1}$.
The proof of the statement for $X_{j-1}$ is identical.

### 3.3.2. The non-early individuals

We will need the following lemma to control the non-early individuals.
Lemma 3.8. For $j \in I, j \geq j(2)$, let $\widetilde{W}_{j}$ be the process which counts the number of non-early individuals, i.e. that obtain a jth mutation during $\left[\xi_{j}, \tau_{j+1}\right)$ and their descendants of type $j$, and $\widetilde{Z}_{j}$ the associated martingale, that is for all $t \in\left[\xi_{j}, \tau_{j+1}\right]$ :

$$
\widetilde{Z}_{j}(t):=e^{-\int_{\xi_{j}}^{t} G_{j}(v) \mathrm{d} v} \widetilde{W}_{j}(t)-\int_{\xi_{j}}^{t} \mu W_{j-1}(u) e^{-\int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u
$$

Then, its conditional variance at time $\tau_{j+1}$ satisfies

$$
\operatorname{Var}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \widetilde{Z}_{j}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\xi_{j}}^{N}\right) \leq \frac{C e^{\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v}}{s k_{N}^{2}} .
$$

Moreover, denoting $\left(Y_{j}^{\prime}(t)\right)_{t \in\left[\xi_{j}, \tau_{j+1}\right]}$ the process following the number of non-early $(\check{Y}, j)$-individuals, the same upper bound holds for the martingale defined by

$$
Z_{j}^{\prime}(t):=e^{-\int_{\xi_{j}}^{t} G_{j}(v) \mathrm{d} v} Y_{j}^{\prime}(t)-\int_{\xi_{j}}^{t} \mu Y_{j-1}(u) e^{-\int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u
$$

Finally, denoting $\left(\check{X}_{j}^{\prime}(t)\right)_{t \in\left[\xi_{j}, \tau_{j+1}\right]}$ the number of non-early $(\check{X}, j)$-individuals, the following process, defined for $t \in\left[\xi_{j}, \tau_{j+1}\right]$, is a mean zero square integrable martingale

$$
\check{Z}_{j}^{X}(t):=e^{-\int_{\xi_{j}}^{t} G_{j}(v) \mathrm{d} v} \check{X}_{j}^{\prime}(t)-\int_{\xi_{j}}^{t} \mu \frac{\alpha}{q_{j}} \frac{X_{j-1}(u) Y_{j-1}(u)}{W_{j-1}(u)} e^{-\int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u,
$$

whose conditional variance at time $\tau_{j+1}$ satisfies

$$
\operatorname{Var}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \check{Z}_{j}^{X}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\xi_{j}}^{N}\right) \leq \frac{C \alpha e^{\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v}}{s k_{N}^{3}}
$$

Proof. We only show the bound of the variance of $\check{Z}_{j}^{X}$ since the other statements follow from Lemma 9.12 in [15] (the ideas of the proof therein are similar to those presented here).

On the event $\left\{\zeta>\xi_{j}\right\}$, the formula for the variance is

$$
\begin{align*}
& \operatorname{Var}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \check{Z}_{j}^{X}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\xi_{j}}^{N}\right) \\
& =\mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-2 \int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v}\left(\mu \frac{\alpha}{q_{j}} \frac{X_{j-1}(u) Y_{j-1}(u)}{W_{j-1}(u)}+\left(B_{j}(u)+D_{j}(u)\right) \check{X}_{j}^{\prime}(u)\right) \right\rvert\, \mathcal{F}_{\xi_{j}}^{N}\right) \tag{3.22}
\end{align*}
$$

We focus on the first term in the parentheses. Since both $X_{j-1}(u)$ and $Y_{j-1}(u)$ are smaller than $W_{j-1}(u)$ by definition, Proposition 3.1 (3) shows that

$$
\begin{aligned}
& \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-2 \int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \mu \mu \frac{\alpha}{q_{j}} \frac{X_{j-1}(u) Y_{j-1}(u)}{W_{j-1}(u)} \\
& \leq(1+\delta) \mu \frac{\alpha}{q_{j}}\lceil s / \mu\rceil e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-\int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} e^{-s\left(u-\xi_{j}\right)} \\
& \leq C s \frac{\alpha}{q_{j}} e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-\int_{\xi_{j}}^{u}} G_{j}(v) \mathrm{d} v \\
& e \\
&-s\left(u-\xi_{j}\right)
\end{aligned},
$$

we then use Proposition 3.1 (4) and we get

$$
\begin{aligned}
& \leq C s \frac{\alpha}{q_{j}} e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-s\left(q_{j}-C_{1}+1\right)\left(u-\xi_{j}\right)} \\
& \leq C \frac{\alpha}{q_{j}} e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v} \frac{1}{q_{j}-C_{1}+1}
\end{aligned}
$$

By Proposition 3.1(4), we obtain that on the event $\zeta>\tau_{j+1}$,

$$
\begin{equation*}
\int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-2 \int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j}} \frac{X_{j-1}(u) Y_{j-1}(u)}{W_{j-1}(u)} \leq C \alpha \frac{e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v}}{k_{N}^{2}} \tag{3.23}
\end{equation*}
$$

For the other term of the variance, by (3.12), we have

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-2 \int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v}\left(B_{j}(u)+D_{j}(u)\right) \check{X}_{j}^{\prime}(u) \mid \mathcal{F}_{\xi_{j}}^{N}\right) \\
& \leq 3 \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-2 \int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \check{X}_{j}^{\prime}(u) \mid \mathcal{F}_{\xi_{j}}^{N}\right) \\
&= 3 \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \int_{\xi_{j}}^{\tau_{j+1}} \mathrm{~d} u e^{-\int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v}\right. \\
&\left.\left.\times\left(\check{Z}_{j}^{X}(u)+\mu \frac{\alpha}{q_{j}} \int_{\xi_{j}}^{u} e^{-\int_{\xi_{j}}^{r} G_{j}(v) \mathrm{d} v} \frac{X_{j-1}(r) Y_{j-1}(r)}{W_{j-1}(r)} \mathrm{d} r\right) \right\rvert\, \mathcal{F}_{\xi_{j}}^{N}\right)
\end{aligned}
$$

We bound the term with $X_{j-1} Y_{j-1} / W_{j-1} \leq W_{j-1}$ using Proposition 3.1 (3) and we upper bound $-G_{j}(v) \leq-s\left(q_{j}-C_{1}\right)$ with the Proposition 3.1(4). Integrating $u$ from $\xi_{j}$ to $\infty$, we can use the martingale property of $\check{Z}_{j}^{X}$ to bound from above the right-hand side by

$$
C s \frac{\alpha}{q_{j}} e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v} \int_{\xi_{j}}^{\infty} \mathrm{d} u e^{-s\left(q_{j}-C_{1}\right)\left(u-\xi_{j}\right)} \int_{\xi_{j}}^{u} \mathrm{~d} r e^{-s\left(r-\xi_{j}\right)}
$$

We use that $\int_{\xi_{j}}^{u} \mathrm{~d} r e^{-s\left(r-\xi_{j}\right)}=\frac{1}{s}\left(1-e^{-s\left(u-\xi_{j}\right)}\right)$ to write

$$
\begin{aligned}
& \mathbb{E}\left(\int_{\xi_{j}}^{\tau_{j+1} \wedge \zeta} \mathrm{~d} u e^{-2 \int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v}\left(B_{j}(u)+D_{j}(u)\right) \check{X}_{j}^{\prime}(u) \mid \mathcal{F}_{\xi_{j}}^{N}\right) \\
& \leq \frac{C \alpha}{q_{j}} e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v} \int_{\xi_{j}}^{\infty} \mathrm{d} u\left(e^{-s\left(q_{j}-C_{1}\right)\left(u-\xi_{j}\right)}-e^{-s\left(q_{j}-C_{1}+1\right)\left(u-\xi_{j}\right)}\right) \\
& \leq \frac{C \alpha e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v}}{q_{j}}\left(\frac{1}{s\left(q_{j}-C_{1}\right)}-\frac{1}{s\left(q_{j}-C_{1}+1\right)}\right) \\
& \leq \frac{C \alpha e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v}}{s q_{j}^{3}} \leq \frac{C \alpha e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v}}{s k_{N}^{3}}
\end{aligned}
$$

where we used Proposition 3.1 (4) for the last inequality. The claim follows since we now bounded the two terms of the variance $(3.22)$, the first one in $(3.23)$, thus concluding the proof.

### 3.3.3. After an early mutation

We now investigate the impact of an early mutation on the proportions. The next lemma essentially states that the difference between the proportion of $(\mathscr{Y}, j-1)$-individuals among the type $j-1$ individuals at time $\tau_{j}$ to the proportion of $(Y, j)$ among the type $j$ individuals at time $\tau_{j+1}$ is mostly determined by the number of early type $j$ individuals.

Lemma 3.9. Let $j \in I, j \geq j(2)$ and let $S$ be the proportion of early type $j$ individuals at time $\tau_{j+1}$. Conditionally given $\mathcal{F}_{\tau_{j}}^{N}$, on the event $\left\{\zeta>\tau_{j+1}\right\}$, the following holds with probability at least $1-C \epsilon / k_{N}$ : If $S>0$, then the early individuals have the same ancestor at time $\tau_{j}$, and if it belongs to the group $X$, then

$$
Y_{j-1}\left(\tau_{j}\right)(1-S)-8 \delta\lceil s / \mu\rceil \leq \check{Y}_{j}\left(\tau_{j+1}\right) \leq Y_{j-1}\left(\tau_{j}\right)(1-S)+8 \delta\lceil s / \mu\rceil
$$

Similarly, if the ancestor belongs to the group $Y$, then

$$
S\lceil s / \mu\rceil+Y_{j-1}\left(\tau_{j}\right)(1-S)-8 \delta\lceil s / \mu\rceil \leq \check{Y}_{j}\left(\tau_{j+1}\right) \leq S\lceil s / \mu\rceil+Y_{j-1}\left(\tau_{j}\right)(1-S)+8 \delta\lceil s / \mu\rceil
$$

## François G. Ged

Proof. Recall the notation $\widetilde{W}_{j}$ and $Y_{j}^{\prime}$ for the processes following the non-early type $j$ individuals, respectively the non-early type ( $\check{Y}, j$ )-individuals. We write

$$
\frac{Y_{j}^{\prime}\left(\tau_{j+1}\right)}{\widetilde{W}_{j}\left(\tau_{j+1}\right)}=\frac{e^{\int_{\xi_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v}\left(\int_{\xi_{j}}^{\tau_{j+1}} \mu Y_{j-1}(u) e^{-\int_{\xi_{j}}^{u} G_{j}(v) \mathrm{d} v} \mathrm{~d} u+Z_{j}^{\prime}\left(\tau_{j+1}\right)\right)}{\widetilde{W}_{j}\left(\tau_{j+1}\right)}
$$

where $Z_{j}^{\prime}$ is a martingale defined in Lemma 3.8. Using Proposition 3.1(2) and Lemma 3.7, one has with probability at least $1-o\left(1 / k_{N}\right)$ that

$$
\begin{equation*}
\frac{Y_{j}^{\prime}\left(\tau_{j+1}\right)}{\widehat{W}_{j}\left(\tau_{j+1}\right)} \leq \frac{e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v}}{1-4 \delta}\left(\left(\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}+\delta\right) \int_{\xi_{j}}^{\tau_{j+1}} s e^{-s\left(u-\xi_{j}\right)} \mathrm{d} u e^{\int_{\tau_{j}}^{\xi_{j}} G_{j-1}(v) \mathrm{d} v}+Z_{j}^{\prime}\left(\tau_{j+1}\right)\right) . \tag{3.24}
\end{equation*}
$$

We compute $\int_{\xi_{j}}^{\tau_{j+1}} s e^{-s\left(u-\xi_{j}\right)} \mathrm{d} u=1-e^{-s\left(\tau_{j+1}-\xi_{j}\right)}$, and claim that this converges to 1 as $N \rightarrow$ $\infty$. Indeed, we first write $e^{-s\left(\xi_{j}-\tau_{j}\right)}=\left(s q_{j}\right)^{1 / q_{j}} e^{-b / q_{j}}$, and use that on the event $\left\{\zeta>\tau_{j+1}\right\}$, Proposition 3.1(4) shows that

$$
\begin{equation*}
\left.\left(s k_{N}(1-2 \delta)\right)^{1 / k_{N}(e+2 \delta)} e^{-b / k_{N}(1-2 \delta)} \leq\left(s q_{j}\right)^{1 / q_{j}} e^{-b / q_{j}} \leq\left(s k_{N}(e+2 \delta)\right)\right)^{1 / k_{N}(1-2 \delta)} . \tag{3.25}
\end{equation*}
$$

Taking the logarithm of the left-hand side, we get

$$
\frac{1}{(1-2 \delta) k_{N}}\left(\log s+\log k_{N}+\log (1-2 \delta)\right) .
$$

The second and third terms vanish as $N \rightarrow \infty$ since $k_{N} \rightarrow \infty$, and this is also the case of the first term thanks to Assumption $\left(A_{1}\right)$, hence the lower bound of (3.25) tends to 1 as $N \rightarrow \infty$. The same reasoning shows that the upper bound of (3.25) converges to 1 as well. Thus, on the event $\left\{\zeta>\tau_{j+1}\right\}$, it holds that

$$
\begin{equation*}
e^{-s\left(\xi_{j}-\tau_{j}\right)}=\left(s q_{j}\right)^{1 / q_{j}} e^{-b / q_{j}} \underset{N \rightarrow \infty}{\longrightarrow} 1 . \tag{3.26}
\end{equation*}
$$

We then write

$$
e^{-s\left(\tau_{j+1}-\xi_{j}\right)}=e^{-s\left(\tau_{j+1}-\tau_{j}\right)} e^{s\left(\xi_{j}-\tau_{j}\right)}=o\left(1 / k_{N}\right),
$$

where we have used (3.18). Coming back to (3.24), this gives

$$
\begin{equation*}
\frac{Y_{j}^{\prime}\left(\tau_{j+1}\right)}{\widetilde{W}_{j}\left(\tau_{j+1}\right)} \leq \frac{1}{1-4 \delta}\left(\left(\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}+\delta\right)+e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} Z_{j}^{\prime}\left(\tau_{j+1}\right)\right) . \tag{3.27}
\end{equation*}
$$

Applying Lemma 3.8 and Doob's maximal inequality for squared integrable martingales, one has

$$
\mathbb{P}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left|Z_{j}^{\prime}\left(\tau_{j+1}\right)\right|>\delta e^{\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \mid \mathcal{F}_{\xi_{j}}^{N}\right) \leq C \frac{e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v}}{\delta^{2} s k_{N}^{2}} .
$$

On the event $\left\{\zeta>\tau_{j+1}\right\}$ a double application of Proposition 3.1(4) gives

$$
e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \leq e^{-s\left(q_{j}-C_{1}\right)\left(\xi_{j}-\tau_{j}\right)} \leq s q_{j} e^{-b}(1+\delta) \leq C s k_{N} e^{-b} .
$$

(Since $s k_{N} \rightarrow 0$ as $N \rightarrow \infty$, the constant $C_{1}$ has been absorbed in $C$, which does not depend on the parameters $\epsilon, \delta, T$.) Hence,

$$
\mathbb{P}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left|Z_{j}^{\prime}\left(\tau_{j+1}\right)\right|>\delta e^{\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \mid \mathcal{F}_{\xi_{j}}^{N}\right) \leq \frac{C e^{-b}}{\delta^{2} k_{N}} \leq \frac{C \epsilon}{k_{N}},
$$

where we used (3.1) and (3.9) for the last inequality (since we chose the constant $T>2$, we replaced it by 2 in the denominator to absorb it in $C$ ). This result combined with (3.27) entails that, on the event $\left\{\zeta>\tau_{j+1}\right\}$, with a probability greater than $1-C \epsilon / k_{N}$ :

$$
\frac{Y_{j}^{\prime}\left(\tau_{j+1}\right)}{\widetilde{W}_{j}\left(\tau_{j+1}\right)} \leq \frac{1}{1-4 \delta}\left(\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}+2 \delta\right) \leq \frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}+7 \delta
$$

For the lower bound, the same reasoning as for the upper bound gives

$$
\frac{Y_{j}^{\prime}\left(\tau_{j+1}\right)}{\widetilde{W}_{j}\left(\tau_{j+1}\right)} \geq \frac{1}{1+4 \delta}\left(\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}-3 \delta\right) \geq \frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}-8 \delta .
$$

Since $\widetilde{W}\left(\tau_{j+1}\right)=(1-S)\lceil s / \mu\rceil$, homogenizing the bounds, we get

$$
\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}-8 \delta \leq \frac{Y_{j}^{\prime}\left(\tau_{j+1}\right)}{(1-S)\lceil s / \mu\rceil} \leq \frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}+8 \delta
$$

To conclude, conditionally given $\mathcal{F}_{\tau_{j}}^{N}$, Lemma 7.5 in [16] bounds from above the probability that two early mutations survive by $2 e^{2 b} / q_{j}^{2} \leq 3 e^{2 b} / k_{N}^{2}$. Then, excluding this event, if there is an early mutation in the group $X$, easy calculations lead to

$$
\begin{aligned}
& Y_{j-1}\left(\tau_{j}\right)(1-S)-8 \delta\lceil s / \mu\rceil \leq \check{Y}_{j}\left(\tau_{j+1}\right) \leq Y_{j-1}\left(\tau_{j}\right)(1-S)+8 \delta\lceil s / \mu\rceil \\
& S\lceil s / \mu\rceil+X_{j-1}\left(\tau_{j}\right)(1-S)-8 \delta\lceil s / \mu\rceil \leq X_{j}\left(\tau_{j+1}\right)-\check{X}_{j}\left(\tau_{j+1}\right) \\
& \leq S\lceil s / \mu\rceil+X_{j-1}\left(\tau_{j}\right)(1-S)+8 \delta\lceil s / \mu\rceil
\end{aligned}
$$

The cases where the early mutant is a $Y$-individual is identical, which concludes the proof.

### 3.3.4. The discrete proportions process

To ensure that our description of the evolution of the proportions is accurate enough, we introduce a stopped discrete time process as follows. For all $j \in I, j \geq j(2)-1$, we define

$$
\mathrm{Y}_{j}^{N}:=\frac{Y_{j \wedge\left(j\left(\zeta / a_{N}\right)-1\right)}\left(\tau_{(j+1) \wedge j\left(\zeta / a_{N}\right)}\right)}{\lceil s / \mu\rceil}
$$

which follows the proportion of $Y$-individuals among the fittest at each time $\tau_{j}$ stopped at the last type $j$ before $j\left(\zeta / A_{N}\right)$ (recall that $j\left(\zeta / a_{N}\right)$ defined in (2.6) is the largest $j$ such that $\left.\tau_{j} \leq a_{N} \zeta\right)$. The reason to stop the process at $j\left(\zeta / a_{N}\right)$ is to ensure that the results above and in particular Proposition 3.1 apply to $Y_{j}$.

Similarly, we denote by $\check{Y}_{j}^{N}$ the process defined with $\check{Y}_{j \wedge\left(j\left(\zeta / a_{N}\right)-1\right)}\left(\tau_{(j+1) \wedge j\left(\zeta / a_{N}\right)}\right)$ in place of $Y_{j \wedge\left(j\left(\zeta / a_{N}\right)-1\right)}\left(\tau_{(j+1) \wedge j\left(\zeta / a_{N}\right)}\right)$, i.e. the process following the proportion of $(Y, j)$-individuals where the killings of the type $j$ mutations have been cancelled (and only those ones). We stress that the event $\left\{\zeta>\tau_{j+1}\right\}$ is included in the above definition, in the sense that the process stops evolving at the last $\tau_{j} \leq \zeta$, and this fact will be kept implicit when working with $Y$ or $\check{Y}$.

We now give a lemma controlling the two first moments of the proportions' increments in the absence of weak selection, when there is no early mutation, or one that does not generate a too large family.
Lemma 3.10. Let $S_{j}$ be the proportion of early type $j$ individuals at time $\tau_{j+1}$ among the type $j$ individuals (potentially, $S_{j}$ can be 0 ). For all $j \in I, j \geq j(2)$, it holds that

$$
\begin{aligned}
\mathbb{E}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N} \mid \mathcal{F}_{\tau_{j}}^{N}\right) & =o\left(1 / k_{N}\right) \\
\mathbb{E}\left(\mathbb{1}_{\left\{S_{j} \leq \epsilon\right\}}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2} \mid \mathcal{F}_{\tau_{j}}^{N}\right) & \leq \frac{C \epsilon}{k_{N}}
\end{aligned}
$$

## François G. Ged

Proof. Throughout the proof, we say population for the population of individuals for which the weak selection between the last time interval $\left[\tau_{j}, \tau_{j+1}\right]$ has been cancelled (but for which the killings that occurred before $\tau_{j}$ are kept). We will thus speak of type $(\check{Y}, j)$-individuals.

Fix $j \in I, j \geq j(2)$. We first note that

$$
\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}=\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{\check{Y}_{j}\left(\tau_{j+1}\right)-Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}
$$

We call a $Y$-individual of type $j$ at time $\tau_{j+1}$ good if his ancestor at time $\tau_{j}$ is of type $j-1$. We denote by $\widehat{Y}_{j}\left(\tau_{j+1}\right)$ the number of good $\check{Y}$ individuals at time $\tau_{j+1}$, and $K_{j}$, respectively $K_{\check{Y}, j}$ the number of type $j$ individuals in the population, respectively of type $\left(\check{Y}, j\right.$ ), at time $\tau_{j+1}$ that are not good. We have

$$
\check{Y}_{j}\left(\tau_{j+1}\right)=K_{\check{Y}, j}+\widehat{Y}_{j}\left(\tau_{j+1}\right)
$$

We pick an individual uniformly at random among the $\lceil s / \mu\rceil$ individuals of type $j$ at time $\tau_{j+1}$. Note that he belongs to the group of good $\check{Y}$-individuals if and only if his ancestor at time $\tau_{j}$ is in the group $Y$ with type $j-1$, and we call this event $B$. Let $j_{\text {anc }} \in \mathbb{N}$ be the type of his ancestor at time $\tau_{j}$. We have in particular that $\mathbb{P}\left(B \mid \mathcal{F}_{\tau_{j}}^{N}, j_{\text {anc }}=j-1, \zeta>\tau_{j+1}\right)=Y_{j-1}\left(\tau_{j}\right) /\lceil s / \mu\rceil$, since his ancestor is chosen uniformly at random among the $\lceil s / \mu\rceil$ individuals of type $j-1$ at time $\tau_{j}$. Using that $\left\{\zeta>\tau_{j+1}\right\}$ is $\mathcal{F}_{\tau_{j+1}}^{N}$-measurable, we then write

$$
\begin{aligned}
\mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{\widehat{Y}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) & =\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \mathbb{P}\left(B \mid \mathcal{F}_{\tau_{j+1}}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)=\mathbb{P}\left(B \cap\left\{\zeta>\tau_{j+1}\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& =\mathbb{P}\left(B \mid \mathcal{F}_{\tau_{j}}^{N}, j_{\mathrm{anc}}=j-1, \zeta>\tau_{j+1}\right) \mathbb{P}\left(j_{\mathrm{anc}}=j-1, \zeta>\tau_{j+1} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& =\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil} \mathbb{P}\left(j_{\mathrm{anc}}=j-1, \zeta>\tau_{j+1} \mid \mathcal{F}_{\tau_{j}}^{N}\right)
\end{aligned}
$$

Basic properties of probability measures entail that

$$
\begin{aligned}
\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}\left(\mathbb{P}\left(\zeta>\tau_{j+1} \mid \mathcal{F}_{\tau_{j}}^{N}\right)-\mathbb{P}(\zeta\right. & \left.\left.>\tau_{j+1}, j_{\mathrm{anc}} \neq j-1 \mid \mathcal{F}_{\tau_{j}}^{N}\right)\right) \\
& \leq \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{\widehat{Y}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \leq \frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil} \mathbb{P}\left(\zeta>\tau_{j+1} \mid \mathcal{F}_{\tau_{j}}^{N}\right)
\end{aligned}
$$

Since $\mathbb{P}\left(\zeta>\tau_{j+1}, j_{\text {anc }} \neq j-1 \mid \mathcal{F}_{\tau_{j}}^{N}\right)=\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} K_{j} /\lceil s / \mu\rceil \mid \mathcal{F}_{\tau_{j}}^{N}\right)$, we have shown that

$$
\begin{align*}
\left|\mathbb{E}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N} \mid \mathcal{F}_{\tau_{j}}^{N}\right)\right| & \leq \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{K_{\check{Y}, j}}{\lceil s / \mu\rceil} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)+\mathbb{P}\left(\zeta>\tau_{j+1}, j_{\mathrm{anc}} \neq j-1 \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq 2 \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{K_{j}}{\lceil s / \mu\rceil} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \leq\left(\frac{\mu}{s}\right)^{1 / 3 k_{N}} \tag{3.28}
\end{align*}
$$

where the last inequality is from [16, Lemma 6.3] and, taking the logarithm and using Assumption $\left(A_{2}\right)$, one can show that the bound is $o\left(1 / k_{N}\right)$.

We turn our attention to the moment of order two. Suppose now that we independently sample two individuals, possibly the same, uniformly at random among the $\lceil s / \mu\rceil$ individuals of type $j$ at time $\tau_{j+1}$. Denote $j_{\text {anc }}$ and $j_{\text {anc }}^{\prime}$ the types of their respective ancestors at time $\tau_{j}$ and let $B^{\prime}$ be the event that they both belong to the good $\check{Y}$ group. Let $D$ be the event that the two ancestors are different with $j_{\mathrm{anc}}=j_{\mathrm{anc}}^{\prime}=j-1$. Recall that given $D$, the ancestors are exchangeable. In particular, given $D$, the ancestor of the first individual is chosen uniformly at random among the $\lceil s / \mu\rceil$ individuals of type $j-1$ at time $\tau_{j}$, and then the ancestor of the

## Moran model with simultaneous strong and weak selections

second one is chosen uniformly at random among the $\lceil s / \mu\rceil-1$ that remain, the two ancestors being independent from $S_{j}$. We get that

$$
\begin{align*}
& \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\}}\left(\frac{\widehat{Y}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil}\right)^{2} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)
\end{aligned} \begin{aligned}
&=\mathbb{P}\left(B^{\prime} \cap\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
&=\mathbb{P}\left(B^{\prime} \mid \mathcal{F}_{\tau_{j}}^{N}, D \cap\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\}\right) \mathbb{P}\left(D \cap\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
&+\mathbb{P}\left(B^{\prime} \cap D^{c} \cap\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq\left(\frac{Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}\right)^{2} \mathbb{P}\left(\zeta>\tau_{j+1}, S_{j} \leq \epsilon \mid \mathcal{F}_{\tau_{j}}^{N}\right)+\mathbb{P}\left(B^{\prime} \cap D^{c} \cap\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \tag{3.29}
\end{align*}
$$

Note that $B^{\prime} \cap D^{c}$ is included in the event that the two sampled individuals have the same ancestor of type $j-1$ at time $\tau_{j}$. The probability to pick twice the same individual is $1 /\lceil s / \mu\rceil$. In particular, Equation (8.16) in [16] implies that the probability that two type $j$ individuals at time $\tau_{j+1}$ have the same ancestor at time $\tau_{j}$ is bounded by $C \epsilon / k_{N}$, so that $\mathbb{P}\left(B^{\prime} \cap D^{c} \cap\{\zeta>\right.$ $\left.\left.\tau_{j+1}, S_{j} \leq \epsilon\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq C \epsilon / k_{N}$. However, this probability of coalescence is not computed explicitly; we give the notation and arguments in order to read Equation (8.16) in [16] and deduce from it the bound for the probability of coalescence (we do not reprove the claim):

- $\Pi_{N}^{*}$ is a coalescent process that coincides at all times $\tau_{\ell}, \ell \in I$ with high probability (see [16, Lemma 8.2] and use the bounds derived in its proof before summing over $j \in I$ ) with the coalescent $\Pi_{N}$ that describes the genealogy of the population,
- $Y_{j}$ corresponds to our $S_{j}$,
- the event $\Psi_{j}$ is defined in the beginning of the proof of Lemma 8.8, with probability converging to 1 as $N \rightarrow \infty$ thanks to Lemma 7.4 and 7.7 in [16].

Therefore, the bound in [16] Equation (8.16) holds true with $\Pi_{N}$ in place of $\Pi_{N}^{*}$ and it follows that $\mathbb{P}\left(B^{\prime} \cap D^{c} \cap\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq C \epsilon / k_{N}$.

Hence, by (3.29), we obtain that

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{S_{j} \leq \epsilon\right\}}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq 2 \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(\frac{K_{j}}{\lceil s / \mu\rceil}\right)^{2} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
&+2 \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}, S_{j} \leq \epsilon\right\}} \frac{1}{\lceil s / \mu\rceil^{2}}\left(\widehat{Y}_{j}\left(\tau_{j+1}\right)^{2}-2 \widehat{Y}_{j}\left(\tau_{j+1}\right) Y_{j-1}\left(\tau_{j}\right)+Y_{j-1}\left(\tau_{j}\right)^{2}\right) \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq 2 \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(\frac{K_{j}}{\lceil s / \mu\rceil}\right) \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)^{2}+\frac{C \epsilon}{k_{N}} \\
& \leq \frac{C \epsilon}{k_{N}}
\end{aligned}
$$

as claimed, where we used the last inequality in (3.28). This concludes the proof.

### 3.3.5. Weak selection

For Lemma 3.10 to be useful, it needs to be combined with a description of the number of type $j$ individuals at time $\tau_{j+1}$ descending from killings between $\left[\tau_{j}, \tau_{j+1}\right]$, that is $\check{X}_{j}\left(\tau_{j+1}\right)$. Note that in Lemma 3.8, the martingale $\check{Z}_{j}^{X}$ is constructed from $\check{X}_{j}^{\prime}$, namely the ( $X, j$ )-individuals coming from non-early killings, that is to say, killings occurring after time $\xi_{j}$. Nonetheless, on the event
$\left\{\zeta>\tau_{j+1}\right\}$, it holds that $\check{X}_{j}^{\prime}\left(\tau_{j+1}\right)=\check{X}_{j}\left(\tau_{j+1}\right)$, by definition of $\zeta$, see the discussion following Lemma 3.4. Hence, the expected effect of the weak selection on the proportions from $\tau_{j}$ to $\tau_{j+1}$ can be estimated:

Lemma 3.11. For all $j \in I, j \geq j(2)$, on the event $\left\{\zeta>\tau_{j}\right\}$, it holds that

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v} \check{X}_{j}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)=\frac{\alpha}{q_{j}} \frac{X_{j-1}\left(\tau_{j}\right) Y_{j-1}\left(\tau_{j}\right)}{\left\lceil s / \mu 7^{2}\right.}+O\left(\frac{\epsilon(\alpha+\sqrt{\alpha})}{k_{N}}\right), \\
& \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{\check{X}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)=\frac{\alpha}{q_{j}} \frac{X_{j-1}\left(\tau_{j}\right) Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil^{2}}+O\left(\frac{\epsilon(1+\alpha+\sqrt{\alpha})}{k_{N}}\right), \\
& \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{\check{X}_{j}\left(\tau_{j+1}\right)^{2}}{\lceil s / \mu\rceil^{2}} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)=o\left(\frac{1}{k_{N}}\right) .
\end{aligned}
$$

Proof. We address the first claim. Using the martingale $\check{Z}_{j}^{X}$ from Lemma 3.8, we can then combine Proposition 3.1(3) and Lemma 3.7 and see that

$$
\begin{aligned}
& \mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v} \check{X}_{j}\left(\tau_{j+1}\right) \\
& =\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \check{Z}_{j}^{X}\left(\tau_{j+1}\right)+\int_{\xi_{j}}^{\tau_{j+1}} e^{-\int_{\tau_{j}}^{u} G_{j}(v) \mathrm{d} v} \mu \frac{\alpha}{q_{j}} \frac{X_{j-1}(u) Y_{j-1}(u)}{W_{j-1}(u)} \mathrm{d} u\right) \\
& =\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \check{Z}_{j}^{X}\left(\tau_{j+1}\right)+\mu \frac{\alpha}{q_{j}}\left(\frac{X_{j-1}\left(\tau_{j}\right) Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil}+O\left(\delta \frac{s}{\mu}\right)\right) \int_{\xi_{j}}^{\tau_{j+1}} e^{-s\left(u-\tau_{j}\right)} \mathrm{d} u\right) \\
& =\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \check{Z}_{j}^{X}\left(\tau_{j+1}\right)+\frac{\alpha}{q_{j}} \frac{X_{j-1}\left(\tau_{j}\right) Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil^{2}}\left(e^{-s\left(\xi_{j}-\tau_{j}\right)}-e^{-s\left(\tau_{j+1}-\tau_{j}\right)}\right)+\frac{O(\delta \alpha)}{q_{j}}\right) .
\end{aligned}
$$

Recall (3.26) and (3.18), so that $e^{-s\left(\xi_{j}-\tau_{j}\right)}-e^{-s\left(\tau_{j+1}-\tau_{j}\right)} \rightarrow 1$ as $N \rightarrow \infty$. Using Proposition 3.1 (4), the above hence becomes

$$
\begin{align*}
& \mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v} \check{X}_{j}\left(\tau_{j+1}\right) \\
& \quad=\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \check{Z}_{j}^{X}\left(\tau_{j+1}\right)+\frac{\alpha}{q_{j}} \frac{X_{j-1}\left(\tau_{j}\right) Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil^{2}}+O\left(\frac{\delta \alpha}{k_{N}}\right)\right) . \tag{3.30}
\end{align*}
$$

We use Cauchy-Schwarz inequality and Lemma 3.8 to bound the expectation of the first term:

$$
\begin{aligned}
& \left|\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \check{Z}_{j}^{X}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)\right| \\
& \quad \leq C \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-2 \int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \mid \mathcal{F}_{\tau_{j}}^{N}\right)^{1 / 2} \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{\alpha e^{\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v}}{s q_{j}^{3}} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)^{1 / 2} .
\end{aligned}
$$

By Proposition 3.1 (4), we have

$$
\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-2 \int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq e^{-2 s\left(q_{j}-C_{1}\right)\left(\xi_{j}-\tau_{j}\right)}
$$

Recalling the definition of $\xi_{j}$ in (3.10) and using Proposition 3.1(4), the right-hand side above reads as

$$
\left(s q_{j}\right)^{2 s\left(q_{j}-C_{1}\right) / s q_{j}} e^{-2 b} \leq\left(s k_{N}(e+2 \delta)\right)^{2\left(1-C_{1} / q_{j}\right)} e^{-2 b} \leq C\left(s k_{N}\right)^{2}\left(s k_{N}\right)^{-2 C_{1} /(1-2 \delta) k_{N}} e^{-2 b} .
$$

Moreover, since $s k_{N} \rightarrow 0$ by Assumption $\left(A_{3}\right)$, we have $\left(s k_{N}\right)^{-2 C_{1} /(1-2 \delta) k_{N}} \leq s^{-2 C_{1} /(1-2 \delta) k_{N}}$, which tends to 1 since by Assumption $\left(A_{1}\right), k_{N} / \log (1 / s) \rightarrow \infty$ as $N \rightarrow \infty$ so that $s^{1 / k_{N}} \rightarrow 1$. Similarly with Proposition 3.1 (4), one can show that $e^{\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \leq C\left(s k_{N}\right)^{-1} e^{b}$ and we see that

$$
\begin{aligned}
\left|\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-\int_{\tau_{j}}^{\xi_{j}} G_{j}(v) \mathrm{d} v} \check{Z}_{j}^{X}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)\right| & \leq C \sqrt{\alpha}\left(\left(s k_{N}\right)^{2} e^{-2 b}\right)^{1 / 2}\left(s^{-2} k_{N}^{-4} e^{b}\right)^{1 / 2} \\
& \leq \frac{C \sqrt{\alpha}}{k_{N}} e^{-b / 2} \leq \frac{C \sqrt{\alpha} \delta}{k_{N}}
\end{aligned}
$$

by Assumption $\left(A_{3}\right)$ and the definition of $b$ in (3.9). Therefore taking the expectation in (3.30) yields

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v} \check{X}_{j}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\tau_{j}}^{N},\left\{\zeta>\tau_{j}\right\}\right) \\
&=\frac{\alpha}{q_{j}} \frac{X_{j-1}\left(\tau_{j}\right) Y_{j-1}\left(\tau_{j}\right)}{\lceil s / \mu\rceil^{2}} \mathbb{P}\left(\zeta>\tau_{j+1} \mid \mathcal{F}_{\tau_{j}}^{N},\left\{\zeta>\tau_{j}\right\}\right)+O\left(\frac{\delta(\alpha+\sqrt{\alpha})}{k_{N}}\right)
\end{aligned}
$$

Since $\mathbb{P}\left(\zeta>\tau_{j+1}\right) \geq \mathbb{P}\left(\zeta>a_{N} T\right) \geq 1-\epsilon$ and $\delta<\epsilon$ by (3.1), the proof of the first statement of the lemma is complete.

To prove the second claim, let $S_{j}$ be the number of early type $j$ individuals alive at time $\tau_{j+1}$. It suffices to show that the following is of order $\epsilon / k_{N}$ :

$$
\begin{align*}
& \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left|e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v}-\frac{1}{\lceil s / \mu\rceil}\right| \check{X}_{j}\left(\tau_{j+1}\right) \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \quad=\mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}}\left(\mathbb{1}_{\left\{S_{j}=0\right\}}+\mathbb{1}_{\left\{S_{j}>0\right\}}\right)\left|e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v}-\frac{1}{\lceil s / \mu\rceil}\right| \check{X}_{j}\left(\tau_{j+1}\right) \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \tag{3.31}
\end{align*}
$$

Then, by Proposition $3.1(2)$, the term with $\mathbb{1}_{\{S=0\}}$ is bounded from above by

$$
\frac{4 \delta}{1-4 \delta} e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v}-\frac{1}{\lceil s / \mu\rceil} \check{X}_{j}\left(\tau_{j+1}\right)
$$

where we also got rid of $\mathbb{1}_{\left\{S_{j}=0\right\}}$. Taking the expectation, we get

$$
\mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}, S_{j}=0\right\}}\left|e^{-\int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v}-\frac{1}{\lceil s / \mu\rceil}\right| \check{X}_{j}\left(\tau_{j+1}\right) \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \leq C \delta\left(\frac{1}{q_{j}}+\frac{\epsilon}{k_{N}}\right)=O\left(\frac{\epsilon}{k_{N}}\right)
$$

where we used Proposition 3.1 (4) and (3.1) for the last inequality.
It remains to bound the term with $\mathbb{1}_{\left\{S_{j}>0\right\}}$ in (3.31). By Cauchy-Schwarz's inequality, we can bound its expectation by

$$
\mathbb{P}\left(\zeta>\tau_{j+1}, S_{j}>0 \mid \mathcal{F}_{\tau_{j}}^{N}\right)^{1 / 2} \mathbb{E}\left(\left.\left(\mathbb{1}_{\left\{S_{j}>0\right\}} \frac{\check{X}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil}\right)^{2} \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right)^{1 / 2}
$$

By Lemma 7.5 in [16], we know that $\mathbb{P}\left(S_{j}>0 \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq C e^{b} / q_{j}$, which can further be bounded by $C e^{b} / k_{N}$ using Proposition 3.1 (4). Therefore, using the third claim of the lemma (proved below), we see that the above is bounded by $o\left(1 / k_{N}\right)$. Using (3.31) with the first claim proves the second claim of the lemma.

We now show the last identity of the lemma. We simply note that elevating to the square then taking the expectation in (3.30) and using the same approximations as throughout the proof
implies that

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} e^{-2 \int_{\tau_{j}}^{\tau_{j+1}} G_{j}(v) \mathrm{d} v} \check{X}_{j}\left(\tau_{j+1}\right)^{2} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& \leq\left(C\left(s q_{j}\right)^{2} e^{-2 b} \operatorname{Var}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \check{Z}_{j}^{X}\left(\tau_{j+1}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)+\frac{C \alpha^{2}}{k_{N}^{2}}\right)=O\left(\frac{1}{k_{N}^{2}}\right)
\end{aligned}
$$

The use of Proposition 3.1 (5) concludes the proof.

## 4. Convergence towards the SDE.

In this section, our strategy to show the convergence is very common when showing convergence of a Markov process to the solution of a SDE: we first establish tightness, then look at the infinitesimal generator, and show that any weak limit solves a martingale problem associated to the SDE. The next lemma addresses the tightness of $\left(\mathcal{Y}_{t}^{N}\right)_{t \in[2, T-1]}$.

Lemma 4.1. Recall that $\mathcal{Y}_{2}^{N}=y_{N}$ with $y_{N} \rightarrow y \in(0,1)$. The sequence $\left\{\left(\mathcal{Y}^{N}(t)\right)_{t \in[2, T-1]} ; N \in\right.$ $\mathbb{N}\}$ is tight.

Proof. The proof uses Aldous' criterion for tightness, stated e.g. in [10, Chapter VI Theorem 4.5]. Let $\lambda, \theta>0$ and let $\sigma, \sigma^{\prime}$ denote any two stopping times with respect to the filtration $\mathcal{F}^{N}$, that are bounded by $T-1$, and such that $\sigma \leq \sigma^{\prime} \leq \sigma+\theta$. Splitting the following probability on the events $\left\{\zeta>a_{N} T\right\}$ and its complement entails that

$$
\begin{align*}
\mathbb{P}\left(\left|\mathcal{Y}_{\sigma^{\prime}}^{N}-\mathcal{Y}_{\sigma}^{N}\right|>\lambda\right) & \leq \mathbb{P}\left(\zeta>a_{N} T,\left|\mathcal{Y}_{\sigma^{\prime}}^{N}-\mathcal{Y}_{\sigma}^{N}\right|>\lambda\right)+\epsilon \\
& \leq \lambda^{-2} \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>a_{N} T\right\}}\left(\mathcal{Y}_{\sigma^{\prime}}^{N}-\mathcal{Y}_{\sigma}^{N}\right)^{2} \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right)\right)+\epsilon \tag{4.1}
\end{align*}
$$

We rewrite the conditional expectation as

$$
\begin{align*}
\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>a_{N} T\right\}}\right. & \left.\left(\sum_{j=j(\sigma)}^{j\left(\sigma^{\prime}\right)-1} \mathrm{Y}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2} \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \\
& \leq 2 \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>a_{N} T\right\}}\left(\left(\sum_{j=j(\sigma)}^{j\left(\sigma^{\prime}\right)-1} \check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2}+\left(\sum_{j=j(\sigma)}^{j\left(\sigma^{\prime}\right)-1} \frac{\check{X}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil}\right)^{2}\right) \right\rvert\, \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right), \tag{4.2}
\end{align*}
$$

where we used that $(x+y)^{2} \leq 2 x^{2}+2 y^{2}$. Recall that on the event $\left\{\zeta>a_{N} T\right\}$, the number of $j \geq 1$ such that $a_{N} \sigma \leq \tau_{j} \leq a_{N} \sigma^{\prime}$ is at most $\left\lceil 3 k_{N} \theta\right\rceil$ by Proposition 3.1 (1). We bound the first sum above by

$$
\begin{aligned}
& \mathbb{E}\left(\left(\sum_{j=j(\sigma)}^{j\left(\sigma^{\prime}\right)-1} \check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2} \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \\
& \quad \leq 2 \mathbb{E}\left(\sum_{j=j(\sigma)}^{j(\sigma)+\left\lceil 3 k_{N} \theta\right\rceil}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2}+2 \sum_{\substack{j, \ell=j(\sigma) \\
j<\ell}}^{j\left(\sigma^{\prime}\right)-1}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)\left(\check{\mathrm{Y}}_{\ell}^{N}-\mathrm{Y}_{\ell-1}^{N}\right) \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right)
\end{aligned}
$$

We now bound the first sum of the right-hand side. Let $S_{j}$ be the number of early type $j$ individuals at time $\tau_{j+1}$. Using our Lemma 3.10 and Lemma 7.8 in [16], we have the following
bound

$$
\begin{aligned}
& \sum_{j=j(\sigma)}^{j(\sigma)+\left\lceil 3 k_{N} \theta\right\rceil} \mathbb{E}\left(\mathbb{E}\left(\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \\
& \leq \sum_{j=j(\sigma)}^{j(\sigma)+\left\lceil 3 k_{N} \theta\right\rceil}\left(\frac{C \epsilon}{k_{N}}+\mathbb{E}\left(\mathbb{P}\left(S_{j}>\epsilon, \zeta>\tau_{j+1} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right)\right) \leq C \epsilon \theta+\frac{\theta}{\epsilon} \leq C \frac{\theta}{\epsilon} .
\end{aligned}
$$

We next turn our attention to the double sum, we write

$$
\sum_{\substack{j, \ell=j(\sigma) \\ j<\ell}}^{j\left(\sigma^{\prime}\right)-1}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)\left(\check{\mathrm{Y}}_{\ell}^{N}-\mathrm{Y}_{\ell-1}^{N}\right) \leq \sup _{n \leq\left\lceil 3 k_{N} \theta\right\rceil} \sum_{\substack{j, \ell=j(\sigma) \\ j<\ell}}^{j(\sigma)+n}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)\left(\check{\mathrm{Y}}_{\ell}^{N}-\mathrm{Y}_{\ell-1}^{N}\right),
$$

therefore we have that

$$
\begin{aligned}
\mid \mathbb{E}\left(\sum_{\substack{, \ell=j(\sigma) \\
j<\ell}}^{j\left(\sigma^{\prime}\right)-1}\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)\right. & \left.\left(\check{\mathrm{Y}}_{\ell}^{N}-\mathrm{Y}_{\ell-1}^{N}\right) \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \mid \\
& \leq \sup _{n \leq\left\lceil 3 k_{N} \theta\right\rceil}\left|\sum_{\substack{, \ell=j(\sigma) \\
j<\ell}}^{j(\sigma)+n} \mathbb{E}\left(\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right) \mathbb{E}\left(\check{\mathrm{Y}}_{\ell}^{N}-\mathrm{Y}_{\ell-1}^{N} \mid \mathcal{F}_{\tau_{\ell}}^{N}\right) \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right)\right| \\
& =\sum_{\substack{j, \ell=j(\sigma) \\
j<\ell}}^{j(\sigma)+\left\lceil 3 k_{N} \theta\right\rceil} \mathbb{E}\left(\left|\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right| \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \times o\left(\frac{1}{k_{N}}\right)=o(1),
\end{aligned}
$$

by Lemma 3.10.
We now bound the second sum in (4.2). Proceeding similarly as before, we obtain

$$
\begin{aligned}
\mathbb{E}\left(\mathbb{1}_{\left\{\zeta>a_{N} T\right\}}( \right. & \left.\left.\sum_{j=j(\sigma)}^{j\left(\sigma^{\prime}\right)-1} \frac{\check{X}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil}\right)^{2} \mid \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \\
\leq & \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>a_{N} T\right\}} \sum_{j=j(\sigma)}^{j(\sigma)+\left\lceil 3 k_{N} \theta\right\rceil}\left(\frac{\check{X}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil}\right)^{2}\right. \\
& \left.\left.+\sum_{j, \ell=j(\sigma)}^{j(\sigma)+\left\lceil 3 k_{N} \theta\right\rceil} \frac{\check{X}_{j}\left(\tau_{j+1}\right)}{\lceil s / \mu\rceil} \mathbb{E}\left(\left.\mathbb{1}_{\left\{\zeta>a_{N} T\right\}} \frac{\check{X}_{\ell}\left(\tau_{\ell+1}\right)}{\lceil s / \mu\rceil} \right\rvert\, \mathcal{F}_{\tau_{\ell}}^{N}\right) \right\rvert\, \mathcal{F}_{\tau_{j(\sigma)}}^{N}\right) \\
\leq & \left\lceil 3 k_{N} \theta\right\rceil o\left(\frac{1}{k_{N}}\right)+\left\lceil 3 \theta k_{N}\right\rceil^{2} \times \frac{C(1+\alpha+\sqrt{\alpha})}{k_{N}^{2}} \leq \theta C(1+\alpha+\sqrt{\alpha}),
\end{aligned}
$$

by Lemma 3.11. Hence, coming back to (4.1), we have shown that for all $\lambda>0$, it holds that

$$
\lim _{\theta \rightarrow 0} \limsup _{N \rightarrow \infty} \sup _{\sigma, \sigma^{\prime}} \mathbb{P}\left(\left|\mathcal{Y}_{\sigma^{\prime}}^{N}-\mathcal{Y}_{\sigma}^{N}\right|>\lambda\right) \leq \lambda^{-2} \lim _{\theta \rightarrow 0}\left(C(1+\alpha+\sqrt{\alpha}) \theta+C \frac{\theta}{\epsilon}+\epsilon\right)=\lambda^{-2} \epsilon .
$$

Since the left-hand side does not depend on $\epsilon$, its value is simply 0 , which shows that the sequence is tight by Aldous' criterion for tightness.

Define

$$
\begin{equation*}
\Delta_{j}:=\frac{\tau_{j+1}-\tau_{j}}{a_{N}} \tag{4.3}
\end{equation*}
$$

Even though $\left\{\mathrm{Y}_{j}^{N} ; j \in I, j \geq j(2)-1\right\}$ is not Markovian, we shall mimic a classical method for showing convergence of a Markov process through its infinitesimal generator.

Lemma 4.2. On the event $\left\{\zeta>\tau_{j+1}\right\}$, for all $f \in \mathcal{C}^{\infty}([0,1])$, all $j \in I, j \geq j(2)$ and $N$ large enough, it holds that

$$
\begin{aligned}
& \left\lvert\, \frac{1}{\Delta_{j}} \mathbb{E}\left(f\left(\mathrm{Y}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)+\alpha \mathrm{Y}_{j-1}^{N}\left(1-\mathrm{Y}_{j-1}^{N}\right) f^{\prime}\left(\mathrm{Y}_{j-1}^{N}\right)\right. \\
&-\int_{0}^{1} \frac{\mathrm{~d} x}{x^{2}} \int_{0}^{1} \mathrm{~d} u\left(f \left(\mathrm{Y}_{j-1}^{N}(1-x)\right.\right.\left.\left.+x \mathbb{1}_{\left\{u \leq \mathrm{Y}_{j-1}^{N}\right\}}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \\
& \leq C \epsilon(1+\alpha+\sqrt{\alpha})\left(\|f\|_{\infty}+\left\|f^{\prime}\right\|_{\infty}+\left\|f^{\prime \prime}\right\|_{\infty}\right) .
\end{aligned}
$$

Proof. Let $\check{X}_{j}^{N}:=\check{X}_{j}\left(\tau_{j+1}\right) /\lceil s / \mu\rceil$. We write

$$
\begin{aligned}
& \mathbb{E}\left(f\left(\mathrm{Y}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)=\mathbb{E}\left(\left(f\left(\check{\mathrm{Y}}_{j}^{N}-\check{X}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
&=\mathbb{E}\left(\left(f\left(\check{\mathrm{Y}}_{j}^{N}-\check{\mathrm{X}}_{j}^{N}\right)-f\left(\check{\mathrm{Y}}_{j}^{N}\right)+f\left(\check{\mathrm{Y}}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
&=\mathbb{E}\left(\left(-\check{\mathrm{X}}_{j}^{N} f^{\prime}\left(\check{\mathrm{Y}}_{j}^{N}\right)+O\left(\left(\check{\mathrm{X}}_{j}^{N}\right)^{2}\right)\left\|f^{\prime \prime}\right\|_{\infty}+f\left(\check{\mathrm{Y}}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right),
\end{aligned}
$$

where we used the Taylor-Lagrange formula. The second term is

$$
\mathbb{E}\left(O\left(\left(\check{X}_{j}\right)^{2}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)=o\left(\frac{1}{k_{N}}\right),
$$

by using Lemma 3.11. We then focus on the first term. We have

$$
\begin{aligned}
\mathbb{E}\left(\check{\mathrm{X}}_{j}^{N} f^{\prime}\left(\check{\mathrm{Y}}_{j}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) & =\mathbb{E}\left(\check{\mathrm{X}}_{j}^{N} f^{\prime}\left(\mathrm{Y}_{j-1}^{N}\right)+O\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& =\frac{\alpha}{q_{j}} \mathrm{Y}_{j-1}^{N}\left(1-\mathrm{Y}_{j-1}^{N}\right) f^{\prime}\left(\mathrm{Y}_{j-1}^{N}\right)+O\left(\frac{\epsilon(1+\alpha+\sqrt{\alpha})}{k_{N}}\right)+o\left(\frac{1}{k_{N}}\right),
\end{aligned}
$$

by Lemmas 3.11 and 3.10. We thus have shown that

$$
\begin{align*}
\mathbb{E}\left(f\left(\mathrm{Y}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)=-\frac{\alpha}{q_{j}} \mathrm{Y}_{j-1}\left(1-\mathrm{Y}_{j-1}\right) f^{\prime}\left(\mathrm{Y}_{j-1}^{N}\right) & +O\left(\frac{\epsilon(1+\alpha+\sqrt{\alpha})}{k_{N}}\right) \\
& +\mathbb{E}\left(f\left(\check{\mathrm{Y}}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) . \tag{4.4}
\end{align*}
$$

We address the last term as follows: let $S_{j}$ be the number of early type $j$ individuals at time $\tau_{j+1}$, we use Taylor-Lagrange Formula to get the existence of $\xi$ strictly between $\check{Y}_{j}^{N}$ and $\mathrm{Y}_{j-1}^{N}$ such that

$$
\mathbb{1}_{\left\{S_{j}<\epsilon\right\}}\left(f\left(\check{\mathrm{Y}}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right)=\mathbb{1}_{\left\{S_{j}<\epsilon\right\}}\left(\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right) f^{\prime}\left(\mathrm{Y}_{j-1}^{N}\right)+\left(\check{\mathrm{Y}}_{j}^{N}-\mathrm{Y}_{j-1}^{N}\right)^{2} f^{\prime \prime}(\xi)\right)
$$

Therefore, applying Lemma 3.10, we see that

$$
\left|\mathbb{E}\left(\mathbb{1}_{\left\{S_{j}<\epsilon\right\}}\left(f\left(\check{\mathrm{Y}}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)\right| \leq o\left(\frac{1}{k_{N}}\right)\left\|f^{\prime}\right\|_{\infty}+O\left(\frac{\epsilon}{k_{N}}\right)\left\|f^{\prime \prime}\right\|_{\infty}
$$

We now turn our attention to the difference when an early mutation generates a large family. In the proof of Lemma 3.10, we introduced the notion of "good" type $j$ individuals at time $\tau_{j+1}$, that are those whose ancestor at time $\tau_{j}$ is of type $j-1$. We denoted the number of individuals of type $j$ at time $\tau_{j+1}$ that are not good by $K_{j}$. Recall that by Proposition 3.1(1), we know that the ancestors at time $\tau_{j}$ of these $K_{j}$ individuals are not of type greater or equal to $j$. On
the other hand, Lemma 6.3 in $[16]$ shows that $\mathbb{E}\left(K_{j} \mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \mid \mathcal{F}_{\tau_{j}}^{N}\right) \leq 5(s / \mu)^{1-1 / 3 k_{N}}$. Markov's Inequality thus yields that

$$
\mathbb{P}\left(\left.\mathbb{1}_{\left\{\zeta>\tau_{j+1}\right\}} \frac{K_{j}}{s / \mu}>\epsilon \right\rvert\, \mathcal{F}_{\tau_{j}}^{N}\right) \leq \frac{5}{\epsilon}\left(\frac{\mu}{s}\right)^{1 / 3 k_{N}}=o\left(\frac{1}{\epsilon k_{N}}\right),
$$

where the last estimate can be derived by taking the logarithm and using Assumption $\left(A_{2}\right)$.
Let $p_{S_{j}}$ denote the conditional distribution of $S_{j}$ given $\mathcal{F}_{\tau_{j}}^{N}$, supported on $\{0,1 /\lceil s / \mu\rceil, \ldots, 1\}$. Note that if an early mutation occurs as described in Lemma 3.9, on the event $\left\{\tau_{j+1}<\zeta\right\}$ and given $\mathcal{F}_{\tau_{j}}^{N}$, the individual who generates the large family, conditionally given that it is a good type $j$ individual, is chosen uniformly at random among the $\lceil s / \mu\rceil$ individuals of type $j-1$ at time $\tau_{j}$. Hence, the probability that the early individual is in group $Y$, respectively $X$, (given that there was an early type $j$ mutation) is $\mathrm{Y}_{j-1}^{N}$, respectively $1-\mathrm{Y}_{j-1}^{N}$, up to a term of order $o\left(1 / \epsilon k_{N}\right)$, as discussed above. Thanks to Lemma 3.9, we can write

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}, S_{j}>\epsilon\right\}}\left(f\left(\check{\mathrm{Y}}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
& =\int_{(\epsilon, 1]} p_{S_{j}}(\mathrm{~d} x)\left(\mathrm{Y}_{j-1}^{N} f\left(\mathrm{Y}_{j-1}^{N}(1-x)+x\right)+\left(1-\mathrm{Y}_{j-1}^{N}\right) f\left(\mathrm{Y}_{j-1}^{N}(1-x)\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \\
& \\
& +E+o\left(\frac{1}{\epsilon k_{N}}\right),
\end{aligned}
$$

where $E$ is the error coming from the approximation in Lemma 3.9, and the probability that this approximation does not hold. In particular, we have that

$$
\begin{aligned}
&|E| \leq p_{S_{j}}((\epsilon, 1])\left(\sup _{-8 \delta<z<8 \delta}\left|f\left(\mathrm{Y}_{j-1}^{N}(1-x)+x+z\right)-f\left(\mathrm{Y}_{j-1}^{N}(1-x)+x\right)\right|\right. \\
&\left.+\sup _{-8 \delta<z<8 \delta}\left|f\left(\mathrm{Y}_{j-1}^{N}(1-x)+z\right)-f\left(\mathrm{Y}_{j-1}^{N}(1-x)\right)\right|\right)+\frac{C \epsilon}{k_{N}}\|f\|_{\infty} \\
& \leq \frac{1+13 \delta}{\epsilon q_{j}} 16 \delta\left\|f^{\prime}\right\|_{\infty}+\frac{C \epsilon}{k_{N}}\|f\|_{\infty} \leq C \frac{\epsilon}{q_{j}}\left(\|f\|_{\infty}+\left\|f^{\prime}\right\|_{\infty}\right),
\end{aligned}
$$

where we used Lemma 7.8 in [16], Proposition 3.1(4), and that $\epsilon^{3}>\delta$ by (3.1). Lemma 3.3 allows us to write

$$
\begin{aligned}
& \mid \mathbb{E}\left(\mathbb{1}_{\left\{\zeta>\tau_{j+1}, S_{j}>\epsilon\right\}}\left(f\left(\check{\mathrm{Y}}_{k}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right) \\
&-\frac{1}{q_{j}} \int_{\epsilon}^{1} \frac{\mathrm{~d} x}{x^{2}} \int_{0}^{1} \mathrm{~d} u\left(f\left(\mathrm{Y}_{j-1}^{N}(1-x)+x \mathbb{1}_{\left\{u \leq \mathrm{Y}_{j-1}^{N}\right\}}\right)\right.\left.-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \mid \\
& \leq C \frac{\epsilon}{q_{j}}\left(\left\|f^{\prime}\right\|_{\infty}+\|f\|_{\infty}\right) .
\end{aligned}
$$

We leave to the reader the proof of the following bound:

$$
\left|\int_{0}^{\epsilon} \frac{\mathrm{d} x}{x^{2}} \int_{0}^{1} \mathrm{~d} u\left(f\left(\mathrm{Y}_{j-1}^{N}(1-x)+x \mathbb{1}_{\left\{u \leq Y_{j-1}^{N}\right\}}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right)\right| \leq \epsilon\left\|f^{\prime \prime}\right\|_{\infty}
$$

Recall (4.4), we thus have shown that

$$
\begin{aligned}
& \mid q_{j} \mathbb{E}\left(f\left(\mathrm{Y}_{j}^{N}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right) \mid \mathcal{F}_{\tau_{j}}^{N}\right)+\alpha \mathrm{Y}_{j-1}^{N}\left(1-\mathrm{Y}_{j-1}^{N}\right) f^{\prime}\left(\mathrm{Y}_{j-1}^{N}\right) \\
& \left.-\int_{0}^{1} \frac{\mathrm{~d} x}{x^{2}} \int_{0}^{1} \mathrm{~d} u\left(f\left(\mathrm{Y}_{j-1}^{N}(1-x)+x \mathbb{1}_{\left\{u \leq Y_{j-1}^{N}\right\}}\right)-f\left(\mathrm{Y}_{j-1}^{N}\right)\right) \right\rvert\, \quad \leq C \epsilon(1+\alpha+\sqrt{\alpha})\left(\|f\|_{\infty}+\left\|f^{\prime}\right\|_{\infty}+\left\|f^{\prime \prime}\right\|_{\infty}\right) .
\end{aligned}
$$

On the event $\left\{\zeta>\tau_{j+1}\right\}$, thanks to Lemma 3.2 , we can replace $q_{j}$ by $1 / \Delta_{j}$, modifying slightly the constant $C$, which nonetheless would not depend on $\epsilon, \delta, T$ and $\alpha$. This concludes the proof.

Lemma 4.3. Any weak limit $\left(\mathcal{Y}_{t}\right)_{t \in[2, T-1]}$ of $\left(\mathcal{Y}_{t}^{N}\right)_{t \in[2, T-1]}$ solves the following martingale problem

$$
\begin{aligned}
& M_{t}=f\left(\mathcal{Y}_{t}\right)-f(y)-\int_{2}^{t}-\alpha \mathcal{Y}_{v}\left(1-\mathcal{Y}_{v}\right) f^{\prime}\left(\mathcal{Y}_{v}\right) \mathrm{d} v \\
&-\int_{2}^{t} \mathrm{~d} v \int_{[0,1]^{2}} \mathrm{~d} u \frac{\mathrm{~d} p}{p^{2}}\left(f\left(\mathcal{Y}_{v}+p\left(\mathbb{1}_{\left\{u \leq \mathcal{Y}_{v}\right\}}-\mathcal{Y}_{v}\right)\right)-f\left(\mathcal{Y}_{v}\right)-p\left(\mathbb{1}_{\left\{u \leq \mathcal{Y}_{v}\right\}}-\mathcal{Y}_{v}\right) f^{\prime}\left(\mathcal{Y}_{v}\right)\right)
\end{aligned}
$$

Proof. Suppose that $\phi: \mathbb{N} \rightarrow \mathbb{N}$ defines a subsequence such that $\mathcal{Y}^{\phi(N)} \rightarrow \mathcal{Y}$ in distribution, as $N \rightarrow \infty$. We define a sequence of random processes derived from the usual model with varying $\epsilon$ and $\delta$. More specifically, for all $\ell \geq 1$ let $\epsilon_{\ell}:=1 / \ell^{2}$ and $\delta_{\ell}=O\left(\epsilon_{\ell}^{3}\right)$ such that (3.1) is satisfied (e.g. $1 / 2 \ell^{6}$ ). Denote $\zeta_{\ell}$ the stopping time associated to the model with parameters $T, \epsilon_{\ell}, \delta_{\ell}$ and define $N_{\ell}$ large enough such that $\mathbb{P}\left(\zeta_{\ell}>a_{N_{\ell}} T\right)<\epsilon_{\ell}$ and Lemma 4.2 holds, and such that $\left(N_{\ell}\right)_{\ell \geq 1}$ is a subsequence of $\phi$. Let $\left(\mathrm{Y}_{j}^{(\ell)}\right)_{j(2) \leq j \leq j(T-1)}$ be the process stopped at time $\zeta_{\ell}$, defined as previously, but with the varying parameters $\epsilon_{\ell}, \delta_{\ell}, N_{\ell}$. Since it does not depend on $\epsilon_{\ell}$ and $\delta_{\ell}$, the corresponding non-stopped continuous-time process is simply $\mathcal{Y}^{N_{\ell}}$. Note that

$$
\left(\mathbb{1}_{\left\{\zeta_{\ell}>a_{N_{\ell}} T\right\}} Y_{j(t)-1}^{(\ell)}\right)_{t \in[2, T-1]}=\left(\mathbb{1}_{\left\{\zeta_{\ell}>a_{N_{\ell}} T\right\}} \mathcal{Y}_{t}^{N_{\ell}}\right)_{t \in[2, T-1]}
$$

Hence, using that $\epsilon_{\ell} \rightarrow 0$ as $\ell \rightarrow \infty$, it is straightforward that $\left(\mathrm{Y}_{j(t)-1}^{(\ell)}\right)_{t \in[2, T-1]} \rightarrow \mathcal{Y}$ in distribution as $\ell \rightarrow \infty$. Thanks to Skorokhod's Representation Theorem (see e.g. [5, Theorem 6.7 p. 70]), we can assume without loss of generality that the convergence holds almost surely as $\ell \rightarrow \infty$.

We define

$$
M_{t}^{(\ell)}:=f\left(\mathrm{Y}_{j(t)-1}^{(\ell)}\right)-f\left(\mathrm{Y}_{j(2)-1}^{(\ell)}\right)-\sum_{j=j(2)}^{j(t) \wedge j\left(\zeta_{\ell}\right)-1} \mathbb{E}\left(f\left(\mathrm{Y}_{j}^{(\ell)}\right)-f\left(\mathrm{Y}_{j-1}^{(\ell)}\right) \mid \mathcal{F}_{\tau_{j}}^{N_{\ell}}\right)
$$

Note that $\zeta_{\ell} \leq a_{N_{\ell}} T$ for finitely many $\ell$ almost surely, as a consequence of the choice of $\epsilon_{\ell}$. Therefore, almost surely, for all $t \in[2, T-1]$, it holds that

$$
\begin{aligned}
\lim _{\ell \rightarrow \infty} M_{t}^{(\ell)} & =\lim _{\ell \rightarrow \infty}\left(f\left(\mathrm{Y}_{j(t)-1}^{(\ell)}\right)-f\left(\mathrm{Y}_{j(2)-1}^{(\ell)}\right)-\sum_{j=j(2)}^{j(t)-1} \mathbb{E}\left(f\left(\mathrm{Y}_{j}^{(\ell)}\right)-f\left(\mathrm{Y}_{j-1}^{(\ell)}\right) \mid \mathcal{F}_{\tau_{j}}^{N_{\ell}}\right)\right) \\
& =M_{t}
\end{aligned}
$$

thanks to Lemma 4.2.
We now show that $\left(M_{t}^{(\ell)}\right)_{t \in[2, T-1]}$ is a martingale with respect to its natural filtration, which will readily extend to its almost sure limit $\left(M_{t}\right)_{t \in[2, T-1]}$. Let $2 \leq t<t+r \leq T-1$ and write

$$
\begin{aligned}
& \mathbb{E}\left(M_{t+r}^{(\ell)}-M_{t}^{(\ell)} \mid\left(M_{u}^{(\ell)}\right)_{u \leq t}\right) \\
& =\sum_{n=j(t)}^{\infty} \mathbb{P}\left(j(t+r) \wedge j(\zeta)=n \mid\left(M_{u}^{(\ell)}\right)_{u \leq t}\right) \\
& \quad \times \sum_{j=j(t)}^{n-1} \mathbb{E}\left(f\left(\mathrm{Y}_{j}^{(\ell)}\right)-f\left(\mathrm{Y}_{j-1}^{(\ell)}\right)-\mathbb{E}\left(f\left(\mathrm{Y}_{j}^{(\ell)}\right)-f\left(\mathrm{Y}_{j-1}^{(\ell)}\right) \mid \mathcal{F}_{\tau_{j}}^{N_{\ell}}\right) \mid\left(M_{u}^{(\ell)}\right)_{u \leq t}\right)
\end{aligned}
$$

All the terms in the last sum are null, since the information given by $\left(M_{u}^{(\ell)}\right)_{u \leq t}$ is contained in $\mathcal{F}_{\tau_{j}}^{N_{\ell}}$ for all $j \geq j(t)$. We thus deduce that $M$ is a martingale, which entails the claim.

Lemma 4.4. Any solution of the martingale problem of Lemma 4.3 is a solution of the $S D E(1.2)$.

Lemma 4.4 follows from Theorem 2.3 in [11], which addresses the question of when a solution of a martingale problem is also a solution of the associated SDE, for general Markov processes. For a more specific treatment of this question in our setting, the reader may read Section 3.3 of [1], that sketches an adaptation of an elegant duality argument from [4] (see proof of Lemma 1 therein).

Concluding the proof of Theorem 2.1. We have established the convergence of $\left(\mathcal{Y}_{t}^{N}\right)_{[2, T-1]}$ towards the solution of (1.2), when the killing probability is set to (2.10), the scaling factor $q_{j+1}$ being a $\mathcal{F}_{\tau_{j+1}}^{N}$-measurable random variable. In order to prove Theorem 2.1, we only need to extend the result to the process with the probability of killing (2.9), where the factor $r_{j+1}$ is deterministic.

By Lemma 6.1 in [16], for all $j \in I, j \geq j(2)$, on the event $\left\{\zeta>\tau_{j+1}\right\}$, it holds that

$$
\left|\Delta_{j}-r_{j}\right| \leq 10 \delta T,
$$

where we recall the definition of $\Delta_{j}$ in (4.3). Thanks to Lemma 3.2, we can write

$$
\begin{equation*}
\left|q_{j}-r_{j}\right| \leq C \delta T \tag{4.5}
\end{equation*}
$$

Let $\left(\mathrm{Y}_{j}^{N, *}\right)_{j(2) \leq j \leq j(T-1)}$ be constructed from the population with probability of killing given by (2.9), analogously to $\left(\mathrm{Y}_{j}^{N}\right)_{j(2) \leq j \leq j(T-1)}$ from the population with probability of killing (2.10). We now define two processes that bound from above and below the process $\left(\mathrm{Y}_{j}^{N, *}\right)_{j(2) \leq j \leq j(T-1)}$.

Let $\eta \in(0, \alpha)$ be arbitrary small. We construct $\left(\mathrm{Y}_{j}^{N, 1}\right)_{j(2) \leq j \leq j(T-1)}$ from the same population as $\left(\mathrm{Y}_{j}^{N, *}\right)_{j(2) \leq j \leq j(T-1)}$ as follows: for every killing happening, say, at the instant $t$ of a type $j$-mutation, we cancel it with probability

$$
1-\frac{(\alpha-\eta) r_{j}}{\alpha q_{j}} .
$$

Note that for $N$ large enough, the above is indeed positive thanks to (4.5) and Proposition 3.1 (4), and that this process has the same law as that of $\left(\mathrm{Y}_{j}^{N}\right)_{j(2) \leq j \leq j(T-1)}$, with weak selection coefficient $\alpha-\eta$ instead of $\alpha$. Similarly, we construct $\left(\mathrm{Y}_{j}^{N, 2}\right)_{j(2) \leq j \leq j(T-1)}$ by adding a killing at every time $t$ of a type $j$-mutation such that no killing occurred at time $t$, with probability

$$
\left(\frac{\alpha+\eta}{q_{j}}-\frac{\alpha}{r_{j}}\right) \frac{X_{j}(t)}{X_{j}(t)+Y_{j}(t)},
$$

and this process has the same law as that of $\left(\mathrm{Y}_{j}^{N}\right)_{j(2) \leq j \leq j(T-1)}$ with weak selection coefficient $\alpha+\eta$ instead of $\alpha$ (again, the above is positive by (4.5)). Furthermore, by construction, for all $j \in I, j(2) \leq j \leq j(T-1)$, it holds that

$$
\mathrm{Y}_{j}^{N, 1} \leq \mathrm{Y}_{j}^{N} \leq \mathrm{Y}_{j}^{N, 2}
$$

We know that $\left(\mathrm{Y}_{j}^{N, 1}\right)_{j(2) \leq j \leq j(T-1)}$, respectively $\left(\mathrm{Y}_{j}^{N, 2}\right)_{j(2) \leq j \leq j(T-1)}$, converges to the solution of (1.2) with $\alpha-\eta$ instead of $\alpha$, respectively $\alpha+\eta$ instead of $\alpha$. Since this is true for all $\eta \in(0, \alpha)$ we deduce that the process $\left(Y_{j}^{N}\right)_{j(2) \leq j \leq j(T-1)}$ defined with the killing probability (2.9) converges to the solution of (1.2), thus proving Theorem 2.1.

## References

[1] Boubacar Bah and Etienne Pardoux. The $\lambda$-lookdown model with selection. Stochastic Processes Appl., 125(3):1089-1126, 2015.
[2] Nathanael Berestycki. Recent Progress in Coalescent Theory, volume 16 of Ensaios Matematicos. Sociedade Brasileira de Matemática, 2009.

## François G. Ged

[3] Jean Bertoin and Jean-François Le Gall. Stochastic flows associated to coalescent processes. Probab. Theory Relat. Fields, 126(2):261-288, 2003.
[4] Jean Bertoin and Jean-François Le Gall. Stochastic flows associated to coalescent processes II: Stochastic differential equations. Annales de l'Institut Henri Poincare (B) Probability and Statistics, 41(3):307-333, 2005.
[5] Patrick Billingsley. Convergence of probability measures. John Wiley \& Sons, 2013.
[6] Erwin Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. Commun. Math. Phys., 197(2):247-276, 1998.
[7] Éric Brunet, Bernard Derrida, Alfred H. Mueller, and Stéphane Munier. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. Phys. Rev. E, 76(4):041104, 2007.
[8] Donald A. Dawson and Zenghu Li. Stochastic equations, flows and measure-valued processes. Ann. Probab., 40(2):813-857, 2012.
[9] Michael M. Desai, Aleksandra M. Walczak, and Daniel S. Fisher. Genetic diversity and the structure of genealogies in rapidly adapting populations. Genetics, 193(2):565-585, 2013.
[10] Jean Jacod and Albert Shiryaev. Limit theorems for stochastic processes, volume 288. Springer, 2013.
[11] Thomas G. Kurtz. Equivalence of stochastic equations and martingale problems. In Stochastic analysis 2010, pages 113-130. Springer, 2011.
[12] Martin Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. Adv. Appl. Probab., 32(4):983-993, 2000.
[13] Martin Möhle. The coalescent in population models with time-inhomogeneous environment. Stochastic Processes Appl., 97(2):199-227, 2002.
[14] Richard A. Neher and Oskar Hallatschek. Genealogies of rapidly adapting populations. Proc. Natl. Acad. Sci. USA, 110(2):437-442, 2013.
[15] Jason Schweinsberg. Rigorous results for a population model with selection I: evolution of the fitness distribution. Electron. J. Probab., 22:1-94, 2017.
[16] Jason Schweinsberg. Rigorous results for a population model with selection II: genealogy of the population. Electron. J. Probab., 22:1-54, 2017.


[^0]:    Keywords: Moran model with selection, Bolthausen-Sznitman's coalescent, $\Lambda$-Wright-Fisher SDE.
    2020 Mathematics Subject Classification: 60J80, 92D15, 92D25, 60H10.
    ${ }^{1}$ Exchangeable refers to the property that permuting the labels of several individuals in the sample leaves the law of the process unchanged.

[^1]:    ${ }^{2}$ In our work, type will refer to another concept. We will get rid of the ambiguity when introducing our model.
    ${ }^{3}$ Although the duality nourished the intuition that there is a link between the Moran model with selection and the $\Lambda$-lookdown model, we do not make use of any duality argument throughout the work.

[^2]:    ${ }^{4}$ Note that $\check{Y}_{j}$ does not correspond to $Y_{j}$ in a population with $\alpha=0$, since the killings of type $j^{\prime} \leq j$ individuals that occurred before time $\tau_{j}$ are kept when counting $\check{Y}_{j}$.

[^3]:    is a mean zero, square integrable martingale;

