

THE EFFECTIVE STRENGTH OF SELECTION IN RANDOM ENVIRONMENT

BY ADRIÁN GONZÁLEZ CASANOVA^{1,a} , DARIO SPANÒ^{2,b}  AND MAITE WILKE-BERENGUER^{3,c} 

¹*School of Mathematical and Statistical Sciences, Arizona State University* ^aagonz591@asu.edu

²*Department of Statistics, University of Warwick*, ^bd.spano@warwick.ac.uk

³*Institut für Mathematik, Humboldt-Universität zu Berlin*, ^cmaite.wilkeberenguer@hu-berlin.de

We analyse a family of two-types Wright-Fisher models with selection in a random environment and skewed offspring distribution. We provide a calculable criterion to quantify the impact of different shapes of selection on the fate of the weakest allele, and thus compare them. The main mathematical tool is duality, which we prove to hold, also in presence of random environment (quenched and in some cases annealed), between the population's allele frequencies and genealogy, both in the case of finite population size and in the scaling limit for large size. Duality also yields new insight on properties of branching-coalescing processes in random environment, such as their long term behaviour.

1. Introduction and overview of content. In population genetics the selective fitness of a species, or an allele, is widely thought to be permeable to the influence of environmental factors which may vary randomly in time: in certain generations, the population may be subject to particularly stressful external conditions (extreme temperatures, cataclysms, abrupt invasions of pathogens, hurricanes etc) making the selective advantage of some allelic types unusually more pronounced than in other generations. This could be due to a better ability to secure resources, or to a lower sensitivity to stress, or to various other reasons [11, 17, 26, 25]. Thus if we consider a population comprising only two allelic types 0 and 1, say, where type 1 is assumed to be always fitter than type 0, it is intuitively clear that the possible occurrence of such “cataclysmic” generations would enhance the probability of extinction of type 0. It is, however, less obvious if rare but strong selective events put type 0 more at risk of extinction than a small but constant-in-time selective pressure. The problem is reminiscent of similar questions arising in experimental biology, where, for example, some detrimental substance (antibiotic) is inoculated in a population of bacteria, and there is an interest in determining whether a constant administration of the substance in low concentration dosage is more effective at wiping out the population, than a more occasional inoculation with higher dosages of varying concentration [34]. This paper aims to study the long term effects of rare selective events and quantify how big and frequent cataclysms owed to random environment must be, in order to wipe out a family as effectively as constant weak selection in a steady environment.

To this purpose, we will construct a suitable family of jump-diffusion allele-frequency models, generalising the well-known family of one-dimensional Lambda-Fleming-Viot models to encompass frequency-dependent selection in a time-varying random environment. It will turn out that, even in the simple case of independent and identically distributed (*i.i.d.*) environment, it is not sufficient to compare the total rate α_s , say, of rare selective events against the rate w of classical constant weak selection: it is necessary to take into account

MSC 2010 subject classifications: Primary 60G99; 60K3; 60K37; secondary 92D10; 92D11; 92D25.

Keywords and phrases: Duality, Wright-Fisher, Selection, Random Environment.

also information about the actual “shape” of rare selection, reflecting the specific action of the random environment. This will lead us to a rigorous derivation of a shape parameter $\alpha^* \geq 1$, obtained in Section 4 (formula (26)), as a non-trivial explicit function of both the random environment and of the offspring distribution influenced by it.

The central result in this regard is Theorem 4.2. We will demonstrate that the correct way to measure the *effective strength of selection*, combining the effect of both rare and weak selection, is via the quantity

$$(1) \quad \alpha_{\text{Eff}} := \alpha_s \alpha^* + w.$$

Indeed, we will find (Theorem 4.2) a critical value $c\beta^*$ (see formula (25)) for the total selection α_{Eff} , separating regimes leading to almost sure ultimate extinction ($\alpha_{\text{Eff}} > c\beta^*$), from regimes where the weaker type 0 has a chance to survive and fixate ($\alpha_{\text{Eff}} < c\beta^*$). This extends results by Foucart [13] and Griffiths [20] who independently discovered this threshold for the case of classical weak selection. While Foucart works with the branching-coalescing process of the population’s genealogy, Griffiths introduces a helpful representation of the generator of the forward-in-time diffusion process of the allele frequencies and combines this with a Lyapunov-function approach. We propose (Section 5.3) an extension of Griffiths’ approach to find the critical value for Lambda Fleming-Viot models with frequency-dependent selection and random environment. A non-trivial step in this endeavour requires closing a gap in [20] (caused by an exchange of limit and integration not justified *a priori*), which we have achieved with our Lemma 4.5. This Lemma, which also covers the case analysed by Griffiths, can be extended to include the models in [19] and it may be of independent interest for uses even beyond population genetics.

It will turn out that the critical value $c\beta^*$ for α_{Eff} does not depend on the selection parameters, nor on the environment. This has interesting and, to some extent, counter-intuitive implications: In a model with purely rare selective events ($w = 0$), the effective critical threshold for α_s to guarantee extinction is given by the ratio $c\beta^*/\alpha^*$, which is never smaller than $c\beta^*$. This implies that, given the same genetic drift, a higher minimum level of rare selective intensity α_s is required to guarantee extinction, compared to a model with purely constant weak selection ($\alpha_s = 0$) for which the same $c\beta^*$ is still the actual threshold for w . Due to the form of the shape parameter α^* , it will be apparent that this is the case even for choices of environment typically favouring large “cataclysms”, from which we would expect a faster pressure towards extinction. For a comparison among specific examples of rare selection mechanisms see Remark 4.4.

The analysis of the longterm effects is embedded in an extended analysis of aspects of the modelling of such cataclysms with further results that may be of independent interest and which we briefly describe here.

1.1. *Finite population model.* To derive our generalised Lambda Fleming-Viot process, we will propose, in Section 2, an individual-based construction, based on a “random graph” approach, first proposed in [19], namely: a graph with deterministic vertices, representing individuals at each generation, and random edges being thrown between any two vertices for which a potential parental relation exists (cf. Definition 2.1 in Section 2.1). On this graph, a suitable notion of selection is defined, associated to the idea that, at each generation, every individual chooses a random number of potential parents from the previous generation and inherits the fittest type among those found in their sampled pool. Thus the larger the typical pool’s size, the less likely it is to inherit the least fit type. In a population of size N , at each generation $g \in \mathbb{Z}$, selection is thus parametrised by the probability distribution Q , say,

of the vector of pool sizes $K_g = \{K_{(g,i)} : i = 1, \dots, N\}$. The distribution of Q will depend randomly upon time, through a sequence of $[0, 1]$ -valued random variables, representing the time-varying random environment.

The selection mechanism is thus fully specified in terms of the time-varying distribution of pool sizes of potential parents, but the actual impact of selection depends also on how each individual chooses each such pool, and this choice is in turn affected by whether or not, at any given generation, the population undergoes an extreme reproductive event (or Lambda-event). In generations without extreme reproductive events, all individuals will sample their own pool of potential parents independently at random; in those with extreme reproduction, choices made by distinct individuals may be correlated. The occurrence and size of extreme reproductive events at each generation will be modelled by an additional, independent sequence of $[0, 1]$ -valued random variables.

1.2. Infinite population model. We will then focus on the dynamics under an *i.i.d.* environment and, taking the population size N to infinity and rescaling time suitably, we will prove weak convergence of both allele frequencies (Theorems 3.3 and 3.5) and ancestral processes (Theorem 3.10). A brief description is the following. The limit ancestral process Z , counting the number of lineages ancestral, at any time, to a (sample from a) given generation, will evolve as a continuous-time Markov chain with state space $\mathbb{N} \cup \{\infty\}$, with positive jumps from n to $n + k$ occurring at rate

$$(2) \quad \int_{[0,1]} \mathbb{P} \left(\sum_{j=1}^n K_{y,j} = n + k \right) \frac{\Lambda_s(dy)}{\mathbb{E}[K_{y,1} - 1]} + wn\delta_{1,k}$$

for $k \in \mathbb{N}$ and with negative jumps from n to $n - k$ occurring at rate

$$(3) \quad c \int_{[0,1]} \binom{n}{k+1} y^{k+1} (1-y)^{n-k-1} \frac{\Lambda_c(dy)}{y^2} + \sigma \binom{n}{2} \delta_{1,k}$$

for all $k \in \{1, \dots, n-1\}$, where $w \geq 0$ and $c \geq 0$ are constants, Λ_s and Λ_c are two finite measures with no atoms in 0 and, for each $y \in [0, 1]$, $\{K_{y,j}\}_{j \in \mathbb{N}}$ is an *iid* family of $\mathbb{N} \cup \{\infty\}$ -valued random variables with common distribution $Q(y)$ parametrised by $y \in [0, 1]$, where $Q(y) = \delta_1$ if and only if $y = 0$. We will refer to $Z = (Z(t) : t \geq 0)$ as the *branching-coalescing process in random environment (BCRE)* (cf. Definition 3.6). Positive jumps are driven by Λ_s and correspond to a lineage branching off into one *actual* and one or more *virtual* parental lineages, as an effect of selective pressure, whereas negative jumps, driven by Λ_c , correspond to two or more lineages coalescing into a common ancestor, as an effect of pure genetic drift. The measures Λ_s and Λ_c are the intensities of two Poisson processes governing the occurrence and size of rare selective events (cataclisms) and of the extreme reproduction events, respectively. Loosely speaking, for every $y \in [0, 1]$, $Q(y)$ represents the distribution of the size of the pool of potential parents which a typical individual in the limit population will choose when the environment has value y . The choice of Q is arbitrary as long as it satisfies a minimal key condition of integrability, as explained later on in Section 1.4. The constant w is responsible for the classical form of *weak* genic selection, acting constantly and deterministically in time. Correspondingly, σ is a weight associated to the classical Kingman dynamics (binary mergers, i.e. non-extreme reproduction).

Known standard genealogical processes are thus subsumed as particular cases of our model: in particular when Λ_s and Λ_c are null on $(0, 1]$, the rates (2) and (3) describe the dynamics of the so-called Ancestral Selection Graph of Krone and Neuhauser [29].

As for the forward-in-time allele frequency process, we will prove that a scaling limit approximation of the 0-allele frequency evolution is described by a jump-diffusion process $(X(t) : t \geq 0)$ arising as the unique strong solution to the SDE

$$\begin{aligned} dX(t) = & \int (\mathbb{E}[X(t^-)^{K_y} \mid X(t^-)] - X(t^-)) \mathcal{N}_s(dt, dy) - wX(t)(1 - X(t))dt \\ & + c \int \int z(\mathbb{1}_{\{u \leq X(t^-)\}} - X(t^-)) \mathcal{N}_c(dt, dz, du) \\ (4) \quad & + \sqrt{\sigma X(t)(1 - X(t))} dB(t) \end{aligned}$$

where: $(B(t) : t \geq 0)$ is a standard Brownian motion; \mathcal{N}_s is a Poisson point measure on $[0, \infty[\times [0, 1]$ with intensity $ds \otimes (\mathbb{E}[K_y - 1])^{-1} \mathbb{1}_{[0, 1]} \Lambda_s(dy)$, with $K_y \equiv K_{y,1}$, $\Lambda_s(dy)$ and w exactly as in (2); \mathcal{N}_c is a Poisson point measure on $[0, \infty[\times [0, 1] \times [0, 1]$ with intensity $ds \otimes z^{-2} \Lambda_c(dz) \otimes du$, with $\Lambda_c(dz)$ exactly as in (3). We will refer to the process X as the *two-type Fleming-Viot process with weak and rare selection (FVWRS)* (cf. Definition 3.2). For $c = 0$ this process reduces to a model introduced by Bansaye, Caballero and Méléard in [4]. The classical Wright-Fisher diffusion process with genic selection is obtained when $c = 0$ and $\mathcal{N}_s([0, \infty[\times [0, 1]) = 0$.

1.3. Duality. The construction summarised above (and described in full details in Section 2.1) shows clearly how the dynamics (and fate) of the weak allele frequency is intrinsically linked to the dynamics of the ancestry of any sample from the population, which can be read off directly from the population's graph, as the connected component subtending the sample.

In models with selection in deterministic environment, a well-established method to study such a link is based on the notion of *stochastic duality*, which has attracted the interest of an increasing number of researchers in a variety of areas of probability [16, 23, 33], including population genetics [9, 12, 19, 2, 31]: If $X = (X(t) : t \geq 0)$ and $Z = (Z(t) : t \geq 0)$ are two stochastic processes, respectively with state space E and F , then X and Z are said to be *dual* to each other *with respect to the duality function* $h : E \times F \rightarrow \mathbb{R}$ if, for every $(x, y) \in E \times F$,

$$(5) \quad \forall t \geq 0 : \quad \mathbb{E}[h(X(t), y) \mid X(0) = x] = \mathbb{E}[h(x, Z(t)) \mid Z(0) = y].$$

We will extend the use of duality to analyse population models with time-varying random environment. We will first show in Lemma 2.6 that, for finite N , a *quenched* version of the duality relation (5) holds conditionally on the environment's sample path, when X describes the forward-in-time process of 0-allele frequencies and Z the corresponding backward-in-time ancestral process, tracking the number of surviving potential ancestral lineages. In this case, the duality function h is an appropriate generalisation of the so-called *sampling duality* function originally proposed in [31] and later adapted by [19] to models with deterministic selection. For the case of *i.i.d.* environment, it turns out that sampling duality also holds in an *annealed* form (i.e. unconditionally on the environment), as will be shown in Proposition 2.8. In addition, as observed in [31] and given in Proposition 2.9, more than one duality relation holds between X and Z .

We will then prove in Lemma 3.8 that sampling duality will converge to an *annealed moment duality* relation between the two limit processes, i.e. one of the form (5) with the choice $h(x, n) = x^n$.

Although random environment was introduced quite a few decades back both in the literature of branching processes (e.g. [3, 37]) and in population genetics models ([26, 25]), it is currently attracting a renewed interest in both communities (see [1, 5, 4, 10, 15, 14, 21] for branching processes and [4, 8, 10, 22] in the context of population genetics). While the present paper draws its main motivation from open problems about selection in population genetics, it is our opinion that the above-mentioned duality property is also interesting from the point of view of the theory of branching (and coalescing) processes in a random environment. In the present work, Lemma 3.7 shows that not only does the key Condition 1 ensure strong existence of the solution of Equation (4), but it also implies conservativeness of the BCRE defined by (2)-(3), and this time in presence of both coalescence and random environment. As a beautiful consequence of moment duality the solution of Equation (4) is Feller, cf. Lemma 3.9. We believe that the results and methods in this paper could help to shed further light on the connections between these two families and it is plausible that several of the results presented here can be extended to other types of non-*iid* environmental processes.

1.4. A key condition on the parameters of rare selection. The action of rare selection in random environment is parametrised by two objects: the rare selection mechanism is given by a kernel $\mathbb{Q} = \{Q(y) : y \in [0, 1]\}$ where, for every y , $Q(y)$ is a distribution on $\mathbb{N} \cup \{\infty\}$; the random environment itself is determined by a measure μ on $[0, 1]$, the space of parameters of \mathbb{Q} , whose one-to-one relation with Λ_s we now explain.

As a central assumption for all our results, we will require that the kernel $\mathbb{Q} = \{Q(y) : y \in [0, 1]\}$ and the measure μ satisfy the following condition:

CONDITION 1. The kernel \mathbb{Q} and the measure μ satisfy

$$(6) \quad \int_{[0,1]} \mathbb{E}[K_y - 1] \mu(dy) < \infty.$$

where K_y is a $(\mathbb{N} \cup \{\infty\})$ -valued random variable with distribution $Q(y)$.

In addition, assume that $Q(y) = \delta_1$ if and only if $y = 0$

Note that the main part of Condition 1 is (6). This is essential for most results, while the second part is mostly for simplicity. The results would still have analogous formulations where one would have to take into account that, e.g. singularities might appear for other values of y that $y = 0$. In addition, for the sake of simplicity we will also always assume that $Q(y) = \delta_1$ if and only if $y = 0$. Note that Condition 1 implies the representation $\mu = \mu(\{0\})\delta_0 + \mu^+$ where μ^+ has density $y \mapsto (\mathbb{E}[K_y - 1])^{-1} \mathbb{1}_{[0,1]}(y)$ with respect to a finite measure which we denote by Λ_s . It will become apparent, however, that the atom in zero of μ has no effect on the selection mechanism, hence we will assume, without loss of generality, $\mu(\{0\}) = 0$. Since μ and Λ_s are suitably equivalent, we will make use of both representations.

The choice of $[0, 1]$ as parameter space for $Q(y)$ is entirely arbitrary and one can replace it with any general parameter space \mathcal{Y} . The use of $[0, 1]$ has the advantage of being rich enough and interpretable, encompassing most known models of non-balancing selection. For example, by setting $Q(y)$ to be the Geometric distribution on \mathbb{N} with parameter $1 - y$ for $y \in [0, 1[$ and $Q(1) \equiv \delta_{\{\infty\}}$, and choosing $\Lambda_s = a\delta_{y^*}$, ($y^* \in]0, 1[$) one recovers known population models with selection in constant environment: For $a = 1$ and $y^* = 0$, the model is neutral (no selection), in which case the ancestry is described by a coalescent with multiple collisions [32, 35] with merger sizes governed by the parameter measure Λ_c and the frequency process is the two-type Fleming-Viot process [7]. For $a = y^* \rightarrow 0$ the model specialises to the haploid

weak selection model, whose genealogy is given by the Ancestral Selection Graph of [29] and its dual is the Wright-Fisher diffusion with weak selection [27]. If, in addition, $c = 0$ the model reduces to Kingman's coalescent process [28] with the Wright-Fisher diffusion as its dual [27]. With random environment (non-degenerate Λ_s), the duality property (5) has not been established before.

The duality implies that the allele 0 will become extinct with probability one if and only if its ancestral process does not admit a stationary distribution, which happens when branching events occur at a sufficiently high rate to ultimately outperform the coalescence events.

1.5. Outline of the paper. The structure of this paper follows the approach of first summarising all important results while postponing the proofs to the last section. The results are presented in the order logical from a modeling point of view. The reader mainly interested in the result on the longterm behaviour (Theorem 4.2), may go directly to Section 4 (briefly visiting Section 3.1 for the existence of the object under study). The paper is structured as follows. We begin, in Section 2.1, with the construction of the *Wright-Fisher graph with selection in random environment* incorporating the possible occurrence of highly skewed offspring distributions (Lambda reproduction mechanism); in Sections 2.2 and 2.3 we will define the frequency process of the 0-allele and the so-called block-counting process of the ancestry of a sample. Subsequently, in Section 2.4, we prove the quenched and annealed sampling duality results. Section 3 is concerned with the scaling limits of the processes introduced before. 3.1 establishes the two-type Fleming-Viot process with selection X as the scaling limit of the frequency process of the 0-allele, if the environment is chosen to be *iid*. Section 3.2 begins with the duality of X and the branching coalescing process in random environment Z . We prove the Feller property and conservativeness of Z and deduce the Feller property of X from this. Duality is then also used to prove that Z arises as the scaling limit of the block-counting process of the ancestry of the finite Wright-Fisher graph. The long-term behaviour of the scaling limits and their dependence on the strength of selection is then analysed in Section 4 subdivided in the result for X in Section 4.1 and its translation through duality for Z in Section 4.2. All proofs are postponed to Section 5, where they appear in the same order as previously the results.

2. Modeling Selection in random environment. This section is devoted to the construction and analysis of an individual-based, finite-population model. It is based on a random graph describing the ancestral relations between the generations in the populations, extending [19] to encompass randomly varying selection.

We first describe how selection, the random environment governing it and extreme reproductive events are parametrised in the graph. The population's forward-in-time allele frequency process (Section 2.2) and its backward-in-time genealogical process (Section 2.3) then both arise as functions of the same graph. This construction in particular then yields the (sampling) duality between the two processes (Section 2.4).

2.1. The discrete ancestral selection graph with random environment and skewed offspring distribution. Consider a population of fixed, finite size $N \in \mathbb{N}$, with discrete, non-overlapping generations indexed by $g \in \mathbb{Z}$. Denote: $[N] := \{1, \dots, N\}$ and $V_N := \mathbb{Z} \times [N]$. Each individual in the history of the population is identified by a point $v = (g, i) \in V_N$, and we will write $g(v) = g$ and $i(v) = i$ to indicate, respectively, the *generation* of v and its *label* in $[N]$. Negative and positive values of g will then index past and future generations, respectively, with respect to an arbitrarily chosen “present generation” $g = 0$.

V_N is the (deterministic) set of vertices of what is to become our random graph. The randomness, of course, lies in the ancestral relations: an edge will be drawn between any two vertices

where one vertex is a potential parent of the other. The dynamics of the random graph will have to incorporate two features: selection in a random environment, and possibly skewed offspring distribution when extreme reproductive events occur. The former is determined by *how many* potential parents an individual chooses, the latter by how the individuals choose the *labels* of their parents. Let us now describe the two mechanisms in more detail.

Selection in random environment At each generation g , we assign to every individual $v = (g, i) \in V$ a random number $K_v \in \mathbb{N} \cup \{\infty\}$, representing the number of potential parents it will choose - with replacement - from among the N individuals in generation $g - 1$. The distribution of the collection $\{K_v : v \in V_N\}$ will be fully specified by

- (i) the law of an environmental process, modelled as a sequence of random variables $\bar{Y} = (Y_g : g \in \mathbb{Z})$ taking values in a measurable state space \mathcal{Y} ;
- (ii) a given probability kernel $\{Q(y, \cdot) : y \in \mathcal{Y}\}$ from \mathcal{Y} to $\mathbb{N} \cup \{\infty\}$;

and the following

ASSUMPTION 1. *For every g , given $Y_g = y \in \mathcal{Y}$, the random variables $K_{(g,i)}, i \in [N]$, are conditionally iid with common distribution $Q(y, \cdot)$.*

In addition, conditioned on a realisation of \bar{Y} , the random vectors $(K_{(g,i)} : i \in [N]), g \in \mathbb{Z}$, are independent.

As mentioned in the introduction, throughout this paper we will mostly take $\mathcal{Y} = [0, 1]$ for convenience. Note that we do not specify the law of the environmental process in this section. The scaling limit results in Section 3 will focus mainly on the assumption of the environment \bar{Y} being an iid sequence.

Highly-skewed offspring distribution The highly-skewed offspring distribution is modelled through correlation in how the individuals choose the labels of their potential parents from the previous generation. The strength of the correlation will be determined by Λ , a distribution on $[0, 1]$ with $\Lambda(\{0\}) = 0$, and $c \in [0, 1]$. Denote with $\mathcal{U}_{[N]}$ the discrete uniform distribution on $[N]$.

The choice of parental labels is then modelled as follows.

ASSUMPTION 2. *Let $(C_g : g \in \mathbb{Z})$ be iid $[0, 1]$ -valued random variables with distribution*

$$(1 - c)\delta_0 + c\Lambda$$

and $(I_g : g \in \mathbb{Z})$ iid random variables with distribution $\mathcal{U}_{[N]}$, such that $(C_g : g \in \mathbb{Z})$, $(I_g : g \in \mathbb{Z})$ and $((K_{(g,i)} : i \in [N]) : g \in \mathbb{Z})$ are independent.

At each generation $g \in \mathbb{Z}$, given realisations $C_g = a$, $I_g = j$ and $(K_{(1,g)}, \dots, (K_{(N,g)}) = (k_{(1,g)}, \dots, k_{(N,g)})$ the labels of the $\sum_{i=1}^N k_{(g,i)}$ potential parents of all individuals in that generation are chosen with replacement from $[N]$ as iid random variables with distribution

$$(7) \quad a\delta_j + (1 - a)\mathcal{U}_{[N]}.$$

In addition, conditioned on a realisation of $((K_{(g,i)} : i \in [N]) : g \in \mathbb{Z})$, the choices of labels across generations are also independent.

In other words, the whole population at generation g , collectively chooses a distribution $Q(Y_g)$, a strength of correlation C_g and marks a special label I_g picked uniformly from $\{1, \dots, N\}$. Then every individual $v = (i, g)$ will independently sample its number of choices for potential parents K_v according to the same $Q(Y_g)$ and subsequently flip K_v iid coins with a random bias C_g . Whenever any such coin returns a success, the I_g -th individual from generation $g - 1$ is chosen as a potential parent for v ; for all coins returning failure, the

potential parent is picked uniformly at random, independently of all other random variables. By construction, ultimately each individual potential parent is, marginally, uniformly chosen from $[N]$.

Notice that, since choices are made *with replacement*, the actual number of *distinct* potential parents, chosen by an individual $v \in V_N$, might be less than K_v . Repeated choices will, however, be invisible to the graph, i.e. no multiple edges will be drawn between any two distinct vertices. We assume that all the above random variables are defined on the same underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote by $[w, v]$ a directed edge from w to v , for $w, v \in V_N$ and by \mathfrak{N} the Borel sigma-algebra of $\mathbb{N} \cup \{\infty\}$ under the discrete topology.

With this, we obtain the definition of the *Wright-Fisher graph with selection in random environment and skewed offspring distribution*.

DEFINITION 2.1. For every $N \in \mathbb{N}$ let $\bar{Y} = (Y_g)_{g \in \mathbb{Z}}$ be a $[0, 1]$ -valued process, $\mathbb{Q} : [0, 1] \times \mathfrak{N} \rightarrow [0, 1]$ a probability kernel, $c \in [0, 1]$, and Λ a distribution on $[0, 1]$. The *Wright-Fisher graph with selection \mathbb{Q} in random environment \bar{Y} and skewed offspring distribution governed by c and Λ* – denoted $\text{WF}(N, \bar{Y}, \mathbb{Q}, c, \Lambda)$ – is given by the graph

$$G_N = G_N(\omega) = (V_N, E_N(\omega)), \quad \omega \in \Omega,$$

with deterministic set of vertices V_N , and random set of edges $E_N = E_N(\omega)$ formed by the rule: $[u, v]$ is an edge of E_N if and only if u is a potential parent of v , where potential parents are chosen according to Assumptions 1 and 2.

On a given Wright-Fisher graph $\text{WF}(N, \bar{Y}, \mathbb{Q}, c, \Lambda)$, we now introduce the two-point type space $\{0, 1\}$, where 0 will be the *weak* (less fit) type and 1 the *strong* (fitter) type. The population's allele frequencies dynamics and the corresponding backward-in-time genealogy can then both be obtained as functions of the same $\text{WF}(N, \bar{Y}, \mathbb{Q}, c, \Lambda)$ graph as we see in the following sections.

2.2. Inheritance and two-type allele frequencies dynamics. We will assign arbitrarily types 0 or 1 to all the individuals in a fixed generation $g_0 \in \mathbb{Z}$, chosen to be the starting generation of our process.

The mechanism of inheritance is modelled as follows: in each subsequent generation, every individual will take on the type of its *fittest* potential parent, that is: it will inherit type 0 if and only if *all* its potential parents are of type 0. Clearly, this mechanism describes the advantage of the fitter type 1 and selection is fully explained in terms of multiple parents choice.

This rule assigns types to all vertices in $\{v \in V_N \mid g(v) \geq g_0\}$. Let $\xi(v) \in \{0, 1\}$ denote the type of v and note that for every $g > g_0$ the vector of types $(\xi(g, i) : i \in [N])$ is exchangeable. Define $[N]_0 := \{0, \dots, N\}$, $[N]_0/N := \{0, 1/N, \dots, 1\}$.

DEFINITION 2.2 (Allele-frequency process). The 0-allele frequency process started in $g_0 \in \mathbb{Z}$ in a Wright-Fisher graph $\text{WF}(N, \bar{Y}, \mathbb{Q}, c, \Lambda)$ is the $[N]_0/N$ -valued process $X^{N, g_0} = (X^{N, g_0}(g) : g \geq g_0)$ describing the proportion of 0-alleles in each generation $g \geq g_0$:

$$X^{N, g_0}(g) := \frac{1}{N} \sum_{\{v: g(v)=g\}} (1 - \xi(v)), \quad g \geq g_0.$$

The forward-in-time dynamics of the allele frequency process can be read off quite immediately from the graphical construction. For every $y \in [0, 1]$, denote by φ_y the probability generating function of the distribution $Q(y, \cdot)$:

$$\varphi_y(x) := \sum_j Q(y, \{j\}) x^j.$$

PROPOSITION 2.3. *Conditionally on a realisation of the environment $\bar{Y} = \bar{y} \in [0, 1]^{\mathbb{Z}}$ the frequency process X^{N, g_0} is a $[N]_0/N$ -valued time-inhomogeneous Markov chain with one-step transition probabilities*

$$(8) \quad \mathbb{P}(X^{N, g_0}(g+1) = z \mid X^{N, g_0}(g) = x, \bar{Y} = \bar{y}) \\ = (1-c)\text{Bin}(Nz \mid N, \varphi_{y_g}(x)) + c \mathbb{E} \left[\text{Bin}(Nz \mid N, \varphi_{y_g}(VB_x + (1-V)x)) \right].$$

for any $z, x \in [N]/N$, where: $\text{Bin}(\cdot \mid m, x)$ is the binomial probability mass function with parameter m, x ; $V \sim \Lambda$; B_x is a Bernoulli random variable with success parameter x ; and V and B_x are mutually independent.

The detailed explanation can, of course, be found in the proof in Section 5.1, but let us briefly shed some light on the role of the probability generating function in this context: Assume a generation g without skewed offspring distribution. Given $Y_g = y_g \in [0, 1]$ as the realisation of the random environment in that generation the individual $(g, 1)$ samples the number $K_{(g,1)}$ of choices of potential parents according to $Q(y_g, \cdot)$ and then independently samples as many potential parents from the previous generation. Given that the frequency of weak 0-type is x in the previous generation, the probability for each choice to be of the weak 0-type is x . Since this individual will only be of the weak 0-type, if *all* its choices are of the weak 0-type, the probability of this event is precisely $\mathbb{E}[x^{K_{(g,1)}}] = \varphi_{y_g}(x)$.

REMARK 2.4. Our graph encompasses several models of allele frequency evolution already known in the population genetics literature. Consider $\mathcal{Y} = [0, 1]$. Models without skewed offspring distribution are those where $c = 0$. The classical neutral Wright-Fisher model, for example, is obtained by setting $c = 0$ and $Q(y, \{1\}) = 1$ (i.e. $Q(y, \cdot) = \delta_1$) for every $y \in [0, 1]$. Indeed, in this case the environment does not affect random mating and every individual choose its only one parent independently, uniformly at random, so that $\varphi_y(x) = x$ for every x, y and the probability (8) becomes Binomial with parameter x . Focusing still on the case $c = 0$, a key role is played by the choice of geometric kernels Q given, for every $y \in [0, 1[$, by

$$Q(y, \cdot) = \sum_{i=1}^{\infty} y^{i-1}(1-y)\delta_i =: \text{Geo}(1-y)$$

and $Q(1, \cdot) := \delta_{\{\infty\}}$. Indeed, with such a choice of Q , if \bar{Y} is taken to be a constant, deterministic process, that is, for some $y \in [0, 1[$, $Y_g = y$ for every $g \in \mathbb{Z}$, then X^{N, g_0} reduces to the ordinary Wright-Fisher model with selection parameter y (see [19], Example 2.3) and the neutral Wright-Fisher model corresponds to the case $y = 0$. If, with the same choice of Q , the environment $\bar{Y} = (Y_g : g \in \mathbb{Z})$ is the N -th instance of a sequence of processes in the domain of attraction a spectrally negative Lévy process \bar{Y} , then it can be shown that the frequency process X^{N, g_0} falls within a class of Wright-Fisher models with selection with random environment recently introduced in [4].

2.3. *Backward-in-time ancestral process.* We now turn to defining the ancestral process arising from a $\text{WF}(N, \bar{Y}, \mathbb{Q}, c, \Lambda)$ graph. An individual $(g(v) - r, i) \in V_N$ living $r \geq 1$ generations before v is called a *potential ancestor of v* , if there exists a path in the graph (V_N, E_N) connecting $v_0 := (g(v) - r, i)$ to $v_r := v$, that is: there exist $v_1, \dots, v_{r-1} \in V_N$ such that $[v_l, v_{l+1}] \in E_N$ for every $l = 0, \dots, r-1$. Denote by $A^N(v)$ the set of *all* such ancestors of v . For $n \in [N]$, the *potential ancestry of the sample v_1, \dots, v_n* is the set $A^N(v_1, \dots, v_n) := \bigcup_{j=1}^n A^N(v_j)$.

Now, given an arbitrary collection $\bar{v} = \{v_1, \dots, v_n\} \subseteq \{g_0\} \times [N]$ of distinct individuals chosen from generation $g_0 \in \mathbb{Z}$, the *potential ancestors of the sample* $\{v_1, \dots, v_n\}$ *alive l generations back in time* are the vertices in the set

$$A_l^N(\bar{v}) := \{w \in A^N(v_1, \dots, v_n) \mid g(w) = g_0 - l\}.$$

Thus $A^N(\bar{v}) = \bigcup_{l=1}^{\infty} A_l^N(\bar{v})$.

The following function of the ancestry will play a key role.

DEFINITION 2.5. Let $\bar{v} \subseteq \{g_0\} \times [N]$ be a sample of individuals from generation g_0 . The *block-counting process of the potential ancestry of \bar{v}* is the $[N]_0$ -valued process $Z_{\bar{v}}^{N, g_0} = (Z_{\bar{v}}^{N, g_0}(g) : g \leq g_0)$ defined by $Z_{\bar{v}}^{N, g_0}(g_0) := |\bar{v}|$ and

$$Z_{\bar{v}}^{N, g_0}(g) := |A_{g_0-g}^N(\bar{v})|, \quad g \leq g_0,$$

where $|A|$ denotes the cardinality of the set A .

The block-counting process thus tracks the number of *distinct* potential ancestors of a sample and forgets their labels. Note that the distribution of $Z_{\bar{v}}^{N, g_0}$ actually depends on \bar{v} only through the sample size $|\bar{v}|$. We omit the subscript \bar{v} when there is no danger of confusion. Conditionally on a realisation of the environment $\bar{Y} = \bar{y} \in [0, 1]^{\mathbb{Z}}$ both the ancestral process $A_{g_0-\bullet}^N(\bar{v})$ and its block-counting process $Z_{\bar{v}}^{N, g_0}$ are time-inhomogeneous Markov chains.

2.4. Sampling dualities. As expected from the construction, the processes introduced in Definition 2.2 and 2.5 have a duality relation determined by the graph. Indeed, we will identify two dualities (for the same processes): in the first case, the duality function is related to generating functions, in the second, the duality function is given by a hypergeometric function. Both are so-called *sampling-dualities*, where the duality function is given by a suitable sampling probability as will be explained in detail below. See [31] for sampling-dualities for the basic Wright-Fisher model.

Note that both processes X^{N, g_0} and Z^{N, g_0} need the additional time-parameter g_0 to indicate where they are anchored with respect to the absolute time, i.e. the time of the random environment $\bar{Y} = (Y_g)_{g \in \mathbb{Z}}$. Observe also, that the quantities $X^{N, g_0}(g)$ and $Z^{N, g_0}(g)$ actually depend on the environment process \bar{Y} only through its coordinates between starting and current time i.e. through the variables Y_{g_0+1}, \dots, Y_g and Y_{g+1}, \dots, Y_{g_0} respectively.

Recall that we denote by φ_y the probability generating function of the distribution $Q(y, \cdot)$, i.e. $\varphi_y(x) := \mathbb{E}[x^{K_y}]$ if $K_y \sim Q(y, \cdot)$. Let $V \sim \Lambda$ and for each $x \in [0, 1]$ let B_x be Bernoulli with success parameter x and let V and B_x be independent. With this, define the function $H : [0, 1] \times \mathbb{N}_0 \times \mathbb{Z} \times [0, 1]^{\mathbb{Z}} \rightarrow [0, 1]$ by

$$(9) \quad H(x, n; g, \bar{y}) := (1 - c)(\varphi_{y_g}(x))^n + c \mathbb{E} \left[\left(\varphi_{y_g}(V B_x + (1 - V)x) \right)^n \right],$$

which will be our (first sampling) duality function in the following

LEMMA 2.6 (Quenched sampling duality). Let X^N and Z^N be, respectively, the 0-allele frequency process and the block-counting process of a Wright-Fisher graph $\text{WF}(N, \bar{Y}, \mathbb{Q}, c, \Lambda)$.

Let H be the function given in (9). Then, for all $x \in [N]_0/N$, $n \in [N]_0$ and $r, s \in \mathbb{Z}$ with $r < s$, \mathbb{P} -a.s.

$$(10) \quad \mathbb{E} \left[H \left(X^{N, r}(s-1), n; s, \bar{Y} \right) \mid X^{N, r}(r) = x, \bar{Y} \right] \\ = \mathbb{E} \left[H \left(x, Z^{N, s}(r+1); r+1, \bar{Y} \right) \mid Z^{N, s}(s) = n, \bar{Y} \right].$$

REMARK 2.7. Let us give some context to this result.

1. The equality (10) establishes a quenched version of “sampling duality”, where by quenched we mean conditional on the realisation of the environment \bar{Y} . The name refers to the fact that the duality function $H(x, n, g; \bar{y})$ is precisely the conditional probability, given $\bar{Y} = \bar{y} = (y_g)_{g \in \mathbb{Z}}$, of all individuals in a sample of size n individuals from generation g are of the weak type 0, given that the 0-type frequency was x in the previous generation. This will be more apparent in the proof of Lemma 2.6 in Section 5.1. It is a direct consequence of the graph-construction.
2. For $c = 0$, i.e. with no skewed offspring distribution, and any choice of \mathbb{Q} and \bar{Y} resulting in $K_v = 1$ almost surely for every $v \in V_N$, the duality function in (9) is the moment-function $H(x, n, g; \bar{y}) = x^n$ and the equality (10) reduces to the well-known *moment duality* between the N -finite Wright-Fisher allele frequency process and the corresponding block-counting process of the genealogy [31].
3. For arbitrary c and \mathbb{Q} , but constant, deterministic environment, the *sampling duality* for Wright-Fisher graphs with selection was proved in Proposition 2.9 in [19].
4. In its general form, the duality function H is time-dependent, but only through the environment process. However some specific choices of environment process (e.g. *iid* environment) induce convenient forms of time-homogeneity. See also Proposition 2.8 below.
5. Since both processes X^N and Z^N are constructed on the same random graph, the duality stated in Lemma 2.6 is of strong type, i.e. can be made sense of “almost surely”. See also [19], Remark 2.10.

The next Proposition shows that a similar duality property holds in an annealed form (i.e. unconditionally on the environment \bar{Y}) when the environment is described by an *iid* process. With the choice of an *iid* environment, the distributions of X^N and Z^N do not depend on the specific choice of starting time g_0 , whence we will normally assume, without loss of generality, that $g_0 = 0$ and omit the corresponding superscript. Write

$$\mathbb{P}^n(\cdot) := \mathbb{P}(\cdot \mid Z^{N, g_0}(g_0) = n) \quad \text{and} \quad \mathbb{P}_x(\cdot) := \mathbb{P}(\cdot \mid X^{N, g_0}(g_0) = x)$$

for any $x \in [N]_0/N$ and $n \in [N]_0$.

Again, let $V \sim \Lambda$ and for each $x \in [0, 1]$ let B_x be Bernoulli with success parameter x , independent of V . If the random environment $\bar{Y} = (Y_g)_{g \in \mathbb{Z}}$ is a sequence of *iid* random variables with common distribution μ , define the function $H_\mu : [0, 1] \times \mathbb{N} \rightarrow [0, 1]$ by

$$\begin{aligned} H_\mu(x, n) &:= \mathbb{E}[H(x, n; 0, \bar{Y})] \\ (11) \quad &= (1 - c)\mathbb{E}[\varphi_{Y_0}(x)^n] + c \mathbb{E}[\varphi_{Y_0}(VB_x + (1 - V)x)^n], \end{aligned}$$

where H is the duality function defined in (9) and used in Lemma 2.6.

PROPOSITION 2.8. Assume that the random environment $\bar{Y} = (Y_g)_{g \in \mathbb{Z}}$ is a sequence of *iid* random variables with common distribution μ . Let H_μ be the function given in (11) Then for all $x \in [N]_0/N$, $n \in \mathbb{N}_0$ and $g \geq 0$

$$(12) \quad \mathbb{E}_x[H_\mu(X^{N, 0}(g), n)] = \mathbb{E}^n[H_\mu(x, Z^{N, 0}(-g))].$$

This proposition is interesting in its own right, but will also be helpful in proving the convergence of the scaling limits of the ancestral process in Theorem 3.10. Note that analogous observations to those in Remark 2.7 also hold for H_μ defined in (11) and Proposition 2.8.

As in [31] we can also observe a second duality from the graphical construction. For every $N \in \mathbb{N}$, define a function $\tilde{H}_N : [0, 1] \times \mathbb{N}_0 \rightarrow [0, 1]$ as

$$(13) \quad \tilde{H}_N(x, n) := \prod_{k=1}^n \frac{xN - k + 1}{N - k + 1}.$$

PROPOSITION 2.9 (Hypergeometric sampling duality). *Let X^N and Z^N be, respectively, the 0-allele frequency process and the block-counting process of a Wright-Fisher graph $\text{WF}(\tilde{N}, \tilde{Y}, \mathbb{Q}, c, \Lambda)$.*

Let \tilde{H}_N be the function given in (13). Then, for all $x \in [N]_0/N$, $n \in [N]_0$ and $r, s \in \mathbb{Z}$ with $r < s$,

$$(14) \quad \mathbb{E} \left[\tilde{H}_N \left(X^{N,r}(s), n \right) \mid X^{N,r}(r) = x \right] = \mathbb{E} \left[\tilde{H}_N \left(x, Z^{N,s}(r) \right) \mid Z^{N,s}(s) = n \right].$$

In this duality, the function does not depend on the random environment even without the assumption of an *iid* environment. The effect of the random environment is present solely through its effect on the distribution of $X^{N,r}(s)$ and $Z^{N,s}(r)$, respectively. In contrast, the duality function given in (9) does not depend on the population size N .

As described in [31], both dualities are to be expected when we have a Wright-Fisher graph and both are *sampling dualities* in the sense that the duality function can be expressed as the probability of sampling n weak individuals. In the case of (9), we consider this event conditionally on the frequency of weak alleles in the *previous* generation being x , thus giving rise to a “sampling with replacement” which yields the probability generating functions, while in the case of (13), the event is considered conditionally on the frequency of weak alleles in the *same* generation. This also explains the explicit appearance of the random environment in (9), but not in (13). For more details see the proofs of the statements in Section 5.1.

3. Scaling limit processes. From now on we will focus solely on models with *iid* random environment and always consider annealed results, i.e. we do not condition on the random environment, but rather average it out.

3.1. Forward scaling limit in iid environment: rare selection. We will study existence of a continuous-time Markov process $X = (X(t) : t \geq 0)$ defined as the solution to the SDE

$$(15) \quad \begin{aligned} dX(t) = & \int (\mathbb{E}[X(t^-)^{K_y} \mid X(t^-)] - X(t^-)) \mathcal{N}_s(dt, dy) - wX(t)(1 - X(t))dt \\ & + c \int \int z(\mathbb{1}_{\{u \leq X(t-)\}} - X(t-)) \mathcal{N}_c(dt, dz, du) \\ & + \sqrt{\sigma X(t)(1 - X(t))} dB(t) \end{aligned}$$

where $(B(t) : t \geq 0)$ is a Brownian motion, \mathcal{N}_c is a Poisson point measure on $[0, \infty[\times[0, 1]\times]0, 1[$ with intensity $ds \otimes z^{-2}\Lambda_c(dz) \otimes du$, where Λ_c is a distribution on $[0, 1]$ with no atom in 0, $c, w \geq 0$, \mathcal{N}_s is a Poisson point measure on $[0, \infty[\times[0, 1]$ with intensity $ds \otimes \mu(dy)$ and K_y is an $\mathbb{N} \cup \{\infty\}$ -valued random variable with distribution $Q(y)$ parametrised by $y \in [0, 1]$. Note that, if X exists, the compensator of \mathcal{N}_c is zero and Condition 1 guarantees that the compensator of \mathcal{N}_s is finite. As a result, X solves the martingale problem associated with the generator \mathcal{A} , given by

$$(16) \quad \begin{aligned} \mathcal{A}f(x) = & \int_{[0,1]} [f(\mathbb{E}[x^{K_y}]) - f(x)] \mu(dy) - wx(1-x)f'(x) \\ & + c \int_{[0,1]} \left\{ x[f(x(1-z)+z) - f(x)] + (1-x)[f(x(1-z)) - f(x)] \right\} \frac{1}{z^2} \Lambda_c(dz) \\ & + \sigma x(1-x) \frac{f''(x)}{2} \end{aligned}$$

for every C^2 function $f : [0, 1] \rightarrow \mathbb{R}$.

LEMMA 3.1. *Assume Condition 1. Then there exists a unique strong solution to (15).*

DEFINITION 3.2. For a distribution Λ_c on $[0, 1]$ with $\Lambda_c(\{0\}) = 0$, a measure μ on $[0, 1]$ and a kernel \mathbb{Q} satisfying Condition 1 and constants $w, c, \sigma \geq 0$, define the *two-type Fleming-Viot process with weak and rare selection (FVWRS)* parametrised by \mathbb{Q}, μ and w and Λ_c, c and σ as the unique strong solution $X = (X(t) : t \geq 0)$ to (15).

We have now all the elements to prove the convergence of the 0-allele frequency process in the Wright-Fisher graph to the solution $X = (X(t) : t \geq 0)$ of (15). For clarity, we will first focus on the case of pure *rare* selection, i.e. with $w = 0$. In the pre-limit model, the parameters μ, c and the random environment process \bar{Y} will now be indexed by the corresponding population size N : μ_N, c_N, \bar{Y}^N . We will choose an *iid* environment with a distribution μ_N assigning a large weight on the event of *no selection* (hence we have rare selection) but such that, when selection occurs, its strength does not scale with N . Recall that we consider annealed results, here, i.e. not conditional on the environment, but rather averaged over it.

THEOREM 3.3. *Let $(\rho_N)_{N \in \mathbb{N}}$ be a sequence of positive numbers converging to 0. For a constant $\sigma \in [0, \infty[$ assume*

$$(17) \quad \lim_{N \rightarrow \infty} \rho_N^{-1} N^{-1} = \sigma.$$

Let Λ_c be a distribution on $[0, 1]$ with $\Lambda_c(\{0\}) = 0$ and $(c_N)_{N \in \mathbb{N}}$ a sequence in $[0, 1]$ converging to 0. For an $\alpha \in (0, 1/2)$ define

$$\Lambda_N^\alpha(du) := u^{-2} \mathbb{1}_{[N^{-\alpha}, 1]}(u) \Lambda_c(du) \quad \text{and} \quad \bar{\Lambda}_N^\alpha := \frac{1}{\Lambda_N^\alpha([0, 1])} \Lambda_N^\alpha$$

and assume that for a constant $c \in [0, \infty[$,

$$(18) \quad \lim_{N \rightarrow \infty} \rho_N^{-1} \frac{c_N}{\Lambda_N^\alpha([0, 1])} = c.$$

Let \mathbb{Q} be a probability kernel from $[0, 1]$ to $\mathbb{N} \cup \{\infty\}$ and μ a measure in $[0, 1]$ such that Condition 1 holds. In particular, μ is σ -finite, so let $(I_N)_{N \in \mathbb{N}}$ be an increasing sequence of sets such that we can define finite measures

$$\gamma_N(\cdot) := \mu(\cdot \cap I_N) \quad \text{and} \quad \bar{\gamma}_N := \frac{1}{\gamma_N([0, 1])} \gamma_N$$

and such that $\gamma_N([0, 1])\rho_N$ converges to 0.

Lastly, assume

$$(19) \quad \lim_{N \rightarrow \infty} \gamma_N([0, 1])c_N = 0.$$

For each $N \in \mathbb{N}$, let X^N be the 0-allele frequency process started in 0 in a $WF(N, \bar{Y}^N, \mathbb{Q}, c_N, \bar{\Lambda}_N^\alpha)$ graph as introduced in Definition 2.2, where $\bar{Y}^N = (Y_g^N)_{g \in \mathbb{Z}}$ is an iid environment with common distribution

$$\bar{\mu}_N := (1 - \gamma_N([0, 1])\rho_N)\delta_0 + \gamma_N([0, 1])\rho_N\bar{\gamma}_N.$$

If $X_0^N \xrightarrow{w} x \in [0, 1]$, as $N \rightarrow \infty$,

$$(X^N(\lfloor \rho_N^{-1} t \rfloor) : t \geq 0) \implies (X(t) : t \geq 0),$$

where $X = (X(t) : t \geq 0)$ is the two-type Fleming-Viot process with rare selection, i.e. the unique strong solution to (15) (for $w = 0$) started in $X_0 = x$.

Recall that, due to the *iid* property of the environment, the distribution of X^N does not depend on the starting time-point $g_0 = 0$.

Observe also that the mechanisms for skewed offspring distribution and rare selection are very similar. The role of c_N in the skewed offspring distribution is played by $\gamma_N([0, 1])\rho_N$ for rare selection: they both describe the probability of an extreme, but rare event. In both cases we truncate a possibly infinite measure – $y^{-1}\Lambda_c(dy)$ and μ respectively – to obtain distributions for the finite population model. The analogue of assumption (18) for the case of rare selection also holds trivially as it is given by

$$\rho_N^{-1} \frac{\gamma_N([0, 1])\rho_N}{\gamma_N([0, 1])} = 1$$

The role of assumption (19) becomes more clear when rewritten as

$$\lim_{N \rightarrow \infty} \frac{\gamma_N([0, 1])\rho_N c_N}{\rho_N} = 0$$

which uncovers that (19) ensures that the probability of both rare events occurring simultaneously converges to 0 sufficiently quickly, in the appropriate time rescaling. Such an assumption simplifies the proof and it could be possibly weakened but this is beyond the scope of this paper.

REMARK 3.4 (Λ -selection). Recall from the introduction in Section 2.1 that the measure Λ_c , is a distribution on $[0, 1]$ with no atom in 0 which describes the extreme events resulting from a skewed offspring distribution through $y^{-2}\Lambda_c(dy)$ and recall also from the discussion after Condition 1, that Λ_s is likewise a finite measure on $[0, 1]$ with no atom in 0 describing the selection mechanism through $\mu(dy) = (\mathbb{E}[K_y - 1])^{-1} \mathbb{1}_{[0, 1]}(y)\Lambda_s(dy)$. The extent of the parallel between these two mechanisms becomes more transparent with the following observations. If for a sequence Λ_c^n , $n \in \mathbb{N}$, of distributions we have $\Lambda_c^n \rightarrow \sigma\delta_0$ weakly, as $n \rightarrow \infty$, (in the Prohorov-metric, for some $\sigma > 0$), then the component of the generator in (16) corresponding to these jumps converges

$$\begin{aligned} \int_{[0, 1]} \{x[f(x(1-z) + z) - f(x)] + (1-x)[f(x(1-z)) - f(x)]\} \frac{1}{z^2} \Lambda_c^n(dz) \\ \xrightarrow{n \rightarrow \infty} \sigma \frac{1}{2} x(1-x) f''(x), \end{aligned}$$

uniformly in $x \in [0, 1]$. The limit is the generator of the Wright-Fisher diffusion. For this reason it is common to find in the literature the interpretation that identifies the atom $\Lambda_c(\{0\})$ with the intensity σ of the “Wright-Fisher noise” or, as will be seen later, of the “Kingman component” of the dual coalescence-mechanism (see for example [36] or, in a spatial set-up, in [6]).

Naturally, the question arises whether an analogous interpretation holds for the selection mechanism, i.e. whether in this sense rare selection converges to weak selection when the measures governing selection converge to a Dirac-measure in 0 and we show here that this is indeed the case. Let Λ_s^n , $n \in \mathbb{N}$ be a sequence of finite measures on $[0, 1]$ such that $\Lambda_s^n \rightarrow w\delta_0$ weakly as $n \rightarrow \infty$ in the Prohorov-metric for a $w > 0$. Let \mathcal{A}_s^n be the summand in the generator (16) responsible for the rare selection mechanism induced by Λ_s^n . Consider the kernel given by $Q(y) = (1-y)\delta_1 + y\delta_2$, $y \in [0, 1]$, i.e. the mechanism with at most binary branching. Then Condition 1 holds automatically and implies that μ has density $1/y$ with respect to Λ_s^n on $(0, 1]$. Applying Taylor’s expansion around x reveals

$$A_s f(x) = \int_{[0, 1]} \{f(x) - x(1-x)yf'(x) + \mathcal{O}(y^2) - f(x)\} \frac{1}{y} \Lambda_s^n(dy)$$

$$\xrightarrow{n \rightarrow \infty} -wx(1-x)f'(x)$$

which is the component responsible for genic (*weak*) selection in (16). The same result holds true for the choice of *geometric* \mathbb{Q} (see Remark 2.4) and one might believe this to be true for every general kernel. Under some mild assumptions this is true, if one allows for a more general class of frequency dependent (weak) selection as studied in [19]. The following conditions on the kernel $\mathbb{Q} = (Q(y) : y \in [0, 1])$, ensure convergence of the generator for rare selection to the generator of genic (weak) selection

ASSUMPTION 3. Assume $y \mapsto Q(y, A)$ is continuous for every $A \in \mathfrak{N}$ and that

$$(20) \quad \sum_{k=2}^{\infty} \lim_{y \downarrow 0} \frac{\mathbb{P}(K_y = k)}{\mathbb{E}[K_y - 1]} = \lim_{y \downarrow 0} \frac{\mathbb{P}(K_y \geq 2)}{\mathbb{E}[K_y - 1]}.$$

Then, as we show below, for any $f \in C^2([0, 1])$

$$(21) \quad A_s f(x) = \int_{[0,1]} [f(\mathbb{E}[x^{K_y}]) - f(x)] \mu(dy) \xrightarrow{n \rightarrow \infty} -ws(x)x(1-x)f'(x)$$

uniformly in $x \in [0, 1]$, where $s(x) = \sum_{k=2}^{\infty} \eta(k) \sum_{l=0}^{k-2} x^l$ (for $\eta(k)$ defined in (24)), which are the selection coefficients studied in [19]. The observations below also immediately imply that if asymptotically when approaching 0 the kernels put most of the weight on binary branches, i.e.

$$\lim_{y \downarrow 0} \frac{\mathbb{P}(K_y = 2)}{\mathbb{E}[K_y - 1]} = 1$$

then $\eta(2) = 1$ and $\eta(k) = 0$, $k \geq 3$, i.e. and $s(x) \equiv 1$ and the limit is classis weak selection.

Let us now see why (21) holds. Using Taylor's expansion of order 1 around x we see

$$\begin{aligned} A_s f(x) &= \int_{[0,1]} [f(\mathbb{E}[x^{K_y}]) - f(x)] \mu(dy) = \int_{[0,1]} [f(\mathbb{E}[x^{K_y}]) - f(x)] \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s^n(dy) \\ &= \int_{[0,1]} (\mathbb{E}[x^{K_y}] - x) f'(x) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s^n(dy) \\ &\quad + \int_{[0,1]} \frac{1}{2} (\mathbb{E}[x^{K_y}] - x)^2 f''(\xi_x) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s^n(dy) \end{aligned}$$

for a suitable $\xi_x \in [\mathbb{E}[x^{K_y}], x]$. We can now consider the two integrals separately and show that the first converges to the desired limit, while the second vanishes. Note that in order to use weak convergence of distributions, we need functions that are continuous in $y \in [0, 1]$. The first part of Assumption 3 ensures that the expressions that we will encounter are continuous for $y \in (0, 1]$, but $y = 0$ needs particular care due to the term $1/\mathbb{E}[K_y - 1]$. (Recall that $Q(0) = \delta_1$.) The following two observations are useful:

$$(22) \quad \mathbb{P}(K_y \geq 2) \leq \sum_{k=1}^{\infty} (k-1) \mathbb{P}(K_y = k) \leq \mathbb{E}[K_y - 1]$$

$$(23) \quad |\mathbb{E}[x^{K_y} - x]| = |x(\mathbb{P}(K_y = 1) - 1) + \sum_{k=2}^{\infty} x^k \mathbb{P}(K_y = k)| \leq 2\mathbb{P}(K_y \geq 2).$$

With this we easily show that the remainder term vanishes:

$$\left| \int_{[0,1]} (\mathbb{E}[x^{K_y}] - x)^2 f''(\xi_x) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s^n(dy) \right| \stackrel{(23)}{\leq} 2 \|f''\|_{\infty} \int_{[0,1]} \mathbb{P}(K_y \geq 2) \Lambda_s^n(dy)$$

$$\xrightarrow{n \rightarrow \infty} 2 \|f''\|_\infty \mathbb{P}(K_0 \geq 2) w = 0,$$

since $y \mapsto \mathbb{P}(K_y \geq 2)$ is continuous in $y = 0$. Similarly,

$$\int_{[0,1]} (\mathbb{E}[x^{K_y}] - x) f'(x) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s^n(dy) \xrightarrow{n \rightarrow \infty} w f'(x) \lim_{y \downarrow 0} \frac{\mathbb{E}[x^{K_y}] - x}{\mathbb{E}[K_y - 1]}$$

which we know exists and is bounded by 1 by (22). To calculate this limit, we define for $y > 0$

$$\eta^y(1) := 1 - \frac{\mathbb{P}(K_y \geq 2)}{\mathbb{E}[K_y - 1]}, \quad \eta^y(k) := \frac{\mathbb{P}(K_y = k)}{\mathbb{E}[K_y - 1]}, k \geq 2$$

and likewise

$$(24) \quad \eta(1) := 1 - \lim_{y \downarrow 0} \frac{\mathbb{P}(K_y \geq 2)}{\mathbb{E}[K_y - 1]}, \quad \eta(k) := \lim_{y \downarrow 0} \frac{\mathbb{P}(K_y = k)}{\mathbb{E}[K_y - 1]}, k \geq 2.$$

Then $\eta^y := (\eta^y(k))_{k \in \mathbb{N}}$ and $\eta := (\eta(k))_{k \in \mathbb{N}}$ are distributions (the latter by Assumption (20)) such that $\lim_{y \downarrow 0} \eta^y(k) = \eta(k)$ for every $k \in \mathbb{N}$. Since thus η^y converge weakly to η ,

$$\lim_{y \downarrow 0} \frac{\mathbb{E}[x^{K_y}] - x}{\mathbb{E}[K_y - 1]} = \lim_{y \downarrow 0} \mathbb{E}_{\eta^y}[x^K - x] = \mathbb{E}_\eta[x^K - x] = \sum_{k \in \mathbb{N}} (x^k - x) \eta_k = -x(1-x)s(x)$$

proving (21).

We may, however, also obtain the diffusion with *weak* selection from the particle system directly. In order to obtain both weak and rare selection simultaneously, we have to carefully design our random environment. μ will still be responsible for the *rare* mechanism only. The kernel of distributions will now consist of distributions from two sets: a general one describing the rare selection and a set of geometric distributions describing weak selection. As before, in the particle system we will have to ensure to rarely make use of the former, while we weaken the effect of the latter as $N \rightarrow \infty$. In order to accomodate the different scalings of the mechanisms without overcomplicating notation, we exceptionally allow the environment to take values in $[-1, 1]$.

THEOREM 3.5. *Retain all the assumptions and notation of Theorem 3.3. For a sequence $(w_N)_{N \in \mathbb{N}}$ in $[0, 1]$ and a constant $w \in [0, \infty]$ assume*

$$\lim_{N \rightarrow \infty} \rho_N^{-1} w_N = w.$$

Let $\tilde{\mathbb{Q}} := (Q(y) : y \in]-1, 0])$ be such that $Q(y) = \text{Geo}_{\mathbb{N}}(1+y)$, for $y \in]-1, 0]$.

For each $N \in \mathbb{N}$, let X^N be the 0-allele frequency process in a $WF(N, \bar{Y}^N, \tilde{\mathbb{Q}} \cup \mathbb{Q}, c_N, \bar{\Lambda}_N^\alpha)$ graph and iid $[-1, 1]$ -valued environment \bar{Y}^N with common distribution

$$\mu_N := (1 - \gamma_N([0, 1])\rho_N)\delta_{-w_N} + \gamma_N([0, 1])\rho_N\bar{\gamma}_N.$$

Then, if $X_0^N \xrightarrow{w} x \in [0, 1]$, as $N \rightarrow \infty$,

$$(X^N(\lfloor \rho_N^{-1} t \rfloor) : t \geq 0) \implies (X(t) : t \geq 0),$$

where $X = (X(t) : t \geq 0)$ is the two-type Fleming-Viot process with weak and rare selection, i.e. the unique strong solution to (15) with $X_0 = x$.

3.2. *Moment duality and convergence to the branching coalescing process in random environment.* As foreshadowed in the introduction, there is a genealogical process arising as the moment dual to the scaling limit of the frequency process of a WF graph.

DEFINITION 3.6. For a distribution Λ_c on $[0, 1]$ with $\Lambda_c(\{0\}) = 0$ and for a measure μ on $[0, 1]$ and a kernel \mathbb{Q} satisfying Condition 1, define the *branching coalescing process in random environment* $Z = (Z(t) : t \geq 0)$ with branching intensity (Q, μ, w) and coalescing intensity (Λ_c, c, σ) as the continuous time $\mathbb{N} \cup \{\infty\}$ -valued Markov chain with positive jumps from $n \in \mathbb{N}$ to $n + k$ at rate

$$\int_{[0,1]} \mathbb{P} \left(\sum_{j=1}^n K_{y,j} = n + k \right) \mu(dy) + wn\delta_{1,k}$$

(where $K_{y,1}, \dots, K_{y,n}$ are iid with distribution $Q(y)$) and with negative jumps from n to $n - k \in \mathbb{N}$ occurring at rate

$$c \int_{[0,1]} \binom{n}{k+1} y^{k+1} (1-y)^{n-k-1} \frac{\Lambda_c(dy)}{y^2} + \sigma \binom{n}{2} \delta_{1,k},$$

and entrance law $\mathbb{E}^\infty[x^{Z(t)}] := \mathbb{P}_x(X(t) = 1)$, where $(X(t) : t \geq 0)$ is the two-type Fleming-Viot process with rare and weak selection parametrised by \mathbb{Q}, μ, w and Λ_c, c and σ .

The first important result about Z is its conservativeness and the Feller property with respect to the topology of the harmonic numbers. Denote by C the set of continuous functions from the metric space $\mathbb{N} \cup \{\infty\}$ with $d(n, m) = |\frac{1}{m} - \frac{1}{n}|$ to itself.

LEMMA 3.7. Under Condition 1, the BCPRE $Z = (Z(t) : t \geq 0)$ given in Definition 3.6 is Feller with respect to the topology of the harmonic numbers on $\mathbb{N} \cup \{\infty\}$ and conservative. More precisely,

$$\forall n \in \mathbb{N}, \quad \forall t \geq 0 : \quad \mathbb{E}^n[Z_t] < \infty.$$

The Feller property of Z is critically helpful in proving that Z is the moment dual of X .

LEMMA 3.8 (Moment duality). Under Condition 1 let $X = (X(t) : t \geq 0)$ be the two-type FV-process with rare and weak selection solution to (15) and let $Z = (Z(t) : t \geq 0)$ be the branching coalescing process in random environment from Definition 3.6.

For any $x \in [0, 1]$, $n \in \mathbb{N} \cup \{\infty\}$ and $t \geq 0$ we have

$$\mathbb{E}_x[X(t)^n] = \mathbb{E}^n[x^{Z(t)}].$$

As an important application of duality, the Feller property for the process X follows as a simple consequence of Z being conservative. This property is used in the proof of Theorem 3.5.

LEMMA 3.9. Under Condition 1 the two-type Fleming-Viot-process with weak and rare selection defined in Definition 3.2 is Feller.

A second and important consequence of the moment duality is, finally, the convergence of the ancestral process to the BCPRE from Definition 3.6.

THEOREM 3.10. *Assume the conditions of Theorem 3.5 and let Z^N be the ancestral process on the Wright-Fisher graph with selection in random environment and multiple mergers $WF(N, \bar{Y}^N, \mathbb{Q} \cup \tilde{\mathbb{Q}}, c_N, \Lambda_c)$ for an iid environment $\bar{Y}^N = (Y_g^N)_{g \in \mathbb{Z}}$ with common distribution*

$$\mu_N := (1 - \gamma_N([0, 1])\rho_N)\delta_{-w_N} + \gamma_N([0, 1])\rho_N\bar{\gamma}_N.$$

Then

$$(Z^N(-\lfloor \rho_N^{-1}t \rfloor)) : t \geq 0 \implies (Z(t) : t \geq 0),$$

where $Z = (Z(t) : t \geq 0)$ is the branching coalescing process in random environment given in Definition 3.6, when we equip \mathbb{N} with the topology of the harmonic numbers, by considering the distance $d(m, n) = |\frac{1}{n} - \frac{1}{m}|$ for $n, m \in \mathbb{N} \cup \{\infty\}$ with $\frac{1}{\infty} := 0$.

4. Long term behaviour. We finally come to consider the long-term behaviour or ergodicity properties of the two-type Fleming-Viot process with weak and/or rare selection $X = (X(t) : t \geq 0)$ (Def. 3.2) and the branching coalescing process in random environment $Z = (Z(t) : t \geq 0)$ (Def. 3.6).

Throughout this section assume \mathbb{Q} and μ to be such that Condition 1 holds. In order to avoid trivialities, assume also that $\mu([0, 1]) + w + c + \sigma > 0$.

Taking a closer look at (15), we see that the two-type FV-process for rare and/or weak selection $X = (X(t) : t \geq 0)$ is a (bounded) supermartingale, since the last two terms – corresponding to the neutral genetic drift – each yield martingales, while the first two terms – corresponding to selection – give a downward drift and downward jumps, respectively. Hence X converges \mathbb{P} -a.s. to a random variable, which we will name X_∞ . The distribution of this random variable is not only of mathematical interest, but also of biological relevance as it encodes the probabilities of fixation or extinction of the weak allele or, a priori, coexistence of the two types traced. As expected, coexistence can be ruled out almost surely, in this case as a direct consequence of the duality between X and the branching coalescing process in random environment Z : duality does indeed uncover how the weak allele's chance of survival (and thus fixation) depends on the ergodic properties of $Z = (Z(t) : t \geq 0)$, as we point out in the following remark.

REMARK 4.1. Applying the same arguments as in Lemma 4.7 of [19], the duality obtained in Lemma 3.8 implies:

1. If Z is *positive recurrent*, then it has a unique invariant distribution ν and

$$\forall x \in [0, 1] : \mathbb{P}_x(X_\infty \in \cdot) = (1 - \varphi_\nu(x))\delta_0 + \varphi_\nu(x)\delta_1,$$

where φ_ν is the probability generating function of ν .

2. If, on the other hand, Z is *not positive recurrent*, then

$$\forall x \in [0, 1[: \mathbb{P}_x(X_\infty = 0) = 1.$$

In other words, the duality implies that $X_\infty \in \{0, 1\}$ \mathbb{P}_x -a.s. for any $x \in [0, 1]$ and the weak allele only has a positive probability of fixation if Z is positive recurrent. In this case, the probability generating function of the invariant distribution of Z determines this probability.

Note also that the dichotomy in particular implies that if there exists an $x \in [0, 1[$ such that $\mathbb{P}_x(X_\infty = 0) < 1$, then this must indeed hold for all $x \in [0, 1[$.

Naturally, the question arises how the chances of survival of the weak allele depend on the strength of selection. We answer this with Theorem 4.2 below and can use the above observations to consequently deduce the ergodic behaviour of the branching coalescing process in random environment in Corollary 4.8.

4.1. *Probability of fixation of the weak allele.* The mechanisms of selection and genetic drift compete in their influence on the probability of fixation (or extinction) of the weak allele. The following quantities are at play:

(i) the *strength of genetic drift*, characterised by σ and the pair c and

$$(25) \quad \beta^* := \frac{1}{2} \mathbb{E} \left[\frac{1}{W(1-W)} \right],$$

respectively. Here, $W := Y^c U$, where $Y^c \sim \Lambda_c$, U has density $2u$ on $[0, 1]$ and they are independent;

(ii) the *strength of weak selection*, given by w , and that of *rare selection*, given by the pair $\alpha_s := \Lambda_s([0, 1])$ and

$$(26) \quad \alpha^* := \mathbb{E} \left[\frac{1}{1 + V \mathbb{E} [K_{Y^s} - 1 \mid Y^s]} \right],$$

where V is uniform on $[0, 1]$ and $Y^s \sim \alpha_s^{-1} \Lambda_s([0, 1])$ and they are independent.

As formalised in the following theorem, if the total strength of selection is sufficiently strong, it overcomes the genetic drift and the weak allele will become extinct almost surely. If not, it retains a positive probability of survival even in the presence of selection.

THEOREM 4.2. *Let $X = (X(t) : t \geq 0)$ be the two-type FV-process for weak and rare selection given by (15) with $\sigma = 0$. Define*

$$p(x) := \mathbb{P}_x(X_\infty = 0)$$

to be the probability of extinction of the weak allele 0, given that we start with a frequency $x \in [0, 1]$.

Assume $\beta^ < \infty$.*

1. *If $\alpha_s \alpha^* + w < c\beta^*$, then for all $x \in]0, 1]$: $p(x) < 1$.*
2. *If $\alpha_s \alpha^* + w > c\beta^*$, then for all $x \in [0, 1[$: $p(x) = 1$.*

$c\beta^*$ is the threshold identified for only *weak* selection as defined with this representation in Equation (26) in [20] (where $c = 1$) and coincides with the threshold in [13]. Note that, given Condition 1, the assumption of Theorem 4.2 is the assumption of Theorem 3 in [20], which, however, also treats the critical case when the strength of (weak) selection and genetic drift are equal. In particular under the condition that $\beta^* < \infty$, W cannot have an atom at zero (implying $\sigma = 0$ also in [13] and [20]). Let us stress that [13] succeeds to also cover the case $\beta^* = \infty$ for weak selection.

REMARK 4.3. In the case studied in [11] lizards with long fingers have a selective advantage whenever their habitat is hit by a hurricane, as their enhanced ability to hold on prevents them from being - literally - blown away. A generation under the influence of a hurricane can be modelled as a two-type Wright-Fisher model with selection in a random environment, taking $K_y \sim Q(y)$ to be geometric with parameter $1 - y$ and adapt the distribution of Y^s , to model the frequency and intensity of hurricanes. Theorem 4.2 now gives conditions for which the individuals with long fingers will go to fixation almost surely, and thus help us to understand how pulses of selection shape the evolution of lizards in particular and all forms of life in general.

REMARK 4.4. The strength of rare selection is determined by both the kernel $\mathbb{Q} = (Q(y) : y \in [0, 1])$ and the measure μ , respectively Λ_s . Two relevant examples are the case of *geometric* \mathbb{Q}

$$Q_{\text{geo}}(y) = \sum_{i=1}^{\infty} (1-y)y^{i-1}\delta_i$$

through its connection to *weak* selection (see Remark 2.4) and the case of *binary* \mathbb{Q}

$$Q_{\text{bin}}(y) = (1-y)\delta_1 + y\delta_2$$

as the simplest branching mechanism (in the dual branching-coalescing process Z). In these cases we obtain

$$\alpha_{\text{geo}}^* := \mathbb{E} \left[\frac{1}{1 + V \frac{Y_s}{1-Y_s}} \right] \quad \text{and} \quad \alpha_{\text{bin}}^* := \mathbb{E} \left[\frac{1}{1 + V Y_s} \right].$$

Observe that, given the same choice of Λ_s , i.e. the same distribution of Y_s , $\alpha_{\text{geo}}^* \leq \alpha_{\text{bin}}^*$ and we say that the *effective strength of rare selection* is larger in the binary case, than in the geometric case, since this, of course, implies $\alpha_s \alpha_{\text{geo}}^* \leq \alpha_s \alpha_{\text{bin}}^*$.

Theorem 4.2 is obtained as a simple combination of two observations: Lemma 4.5 explains that the frequency process converges to 0 almost surely, i.e. the weak allele dies out almost surely, if and only if it does so quickly, in the sense that its expectation is integrable over time. (Indeed, we will prove that in this case, the expectation decays exponentially.) Lemma 4.6 then establishes the connection between the threshold and the integrability in time of the expected frequency of the weak allele.

LEMMA 4.5. *Let $X = (X(t) : t \geq 0)$ be the two-type FV-process for weak and rare selection given by (15). Assume $\beta^* < \infty$. For any $x \in [0, 1]$:*

$$\int_{[0, \infty[} \mathbb{E}_x[X(s)] ds < \infty \quad \Longleftrightarrow \quad \mathbb{P}_x(X_\infty = 0) = 1.$$

The result of Lemma 4.5 can be explained with the following intuition: If one believes that the sample heterozygosity $\mathbb{E}_x[X(t)(1 - X(t))]$ decays exponentially over time for any $x \in [0, 1[$ as it does in the classical case, the assumption $\mathbb{P}_x(X_\infty = 0) = 1$ means that the term $1 - X(t)$ cannot contribute to this decay, hence $\mathbb{E}_x[X(t)]$ should decay exponentially (and thus be integrable), as we will indeed prove.

LEMMA 4.6. *Let $X = (X(t) : t \geq 0)$ be the two-type FV-process for weak and rare selection given by (15) with $\sigma = 0$. Assume $\beta^* < \infty$.*

1. *If $\alpha_s \alpha^* + w < c\beta^*$, then there exists an $\bar{x} \in]0, 1[$: $\int_{[0, \infty[} \mathbb{E}_{\bar{x}}[X(s)] ds = \infty$.*
2. *If $\alpha_s \alpha^* + w > c\beta^*$, then for all $x \in [0, 1[$: $\int_{[0, \infty[} \mathbb{E}_x[X(s)] ds < \infty$.*

The proof of Lemma 4.6 is quite delicate in that it requires analysis of a suitable Lyapunov function. It follows the strategy used in [20]. The proof, provided in Section 5.3, is also of independent interest because it provides a rigorous method to fill a technical gap often left open in the proof of similar statements found in the literature. Lemma 4.5 crucially helps circumventing such issues and guarantees our final result.

The proof of both lemmata can be found in Section 5.3.

Once the truth of Lemmata 4.6 and 4.5 is established, the proof of Theorem 4.2 is straightforward and thus we give it here.

PROOF OF THEOREM 4.2. Assuming $\alpha_s \alpha^* + w < c\beta^*$, Lemma 4.6, 1. and Lemma 4.5 imply the existence of an $\bar{x} \in]0, 1]$ such that $\mathbb{P}_{\bar{x}}(X_\infty = 0) < 1$. Therefore, we are in case 1. of Remark 4.1, i.e. the dual Z must be positive recurrent. Since this does not depend in the choice of starting point x we may conclude that $\mathbb{P}_x(X_\infty = 0) < 1$ for all $x \in [0, 1]$ which proves 1.. 2. is immediate from Lemmas 4.6, 2. and 4.5. \square

4.1.1. *Griffiths representation of the generator \mathcal{A} of $X = (X(t) : t \geq 0)$.* The proof of Lemma 4.6 follows the idea of the proof of Theorem 3 in [20], but we extend and formalise the arguments. Central to this argument is a representation of the generator \mathcal{A} (equation (16)) of X , in the spirit of Theorem 1 in [20]. This representation can be generalised and we believe this observation to be of independent interest, whence we present it here.

LEMMA 4.7 (Griffiths representation). *The generator \mathcal{A} of X can be written as*

$$\begin{aligned} \mathcal{A}f(x) = & \sigma \frac{1}{2} x(1-x) f''(x) + \frac{1}{2} cx(1-x) \mathbb{E} [f''(x(1-W) + VW)] \\ & - wx(1-x) f'(x) \\ & - \alpha_s x(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \times \right. \\ & \left. f' \left(x - x(1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \right) \right], \end{aligned}$$

for every $x \in [0, 1]$, and V , W and Y^s chosen as in (25) and (26).

4.2. *Ergodicity of the branching coalescing process in random environment.* As characterised in Remark 4.1, the chance of survival of the weak allele has a direct correspondence to the ergodic behaviour of the *branching coalescing process in random environment* $(Z(t) : t \geq 0)$. Hence, the following is a corollary of Theorem 4.2.

COROLLARY 4.8. *Let $Z = (Z(t) : t \geq 0)$ be the coordinated branching coalescing process from Definition 3.6 with $\sigma = 0$.*

Assume $\beta^ < \infty$. Then*

1. *If $\alpha_s \alpha^* + w < c\beta^*$, then Z is positive recurrent.*
2. *If $\alpha_s \alpha^* + w > c\beta^*$, then $\mathbb{P}(Z(t) \leq M) \rightarrow 0$, as $t \rightarrow \infty$, for all $M \in \mathbb{N}$ i.e. Z is null-recurrent or transient.*

In the case of 1., the generating function of ν , the invariant distribution of the positive recurrent Z , is given by $\varphi_\nu(x) = \mathbb{P}_x(X_\infty = 1)$.

5. Proofs. We collect in this section mainly the proofs of the results presented in the previous sections. Each subsection corresponds to a section above.

5.1. *Individual based models for a finite population.* The first observation in this section was the precise transition probabilities of the discrete allele-frequency process.

PROOF OF PROPOSITION 2.3. Conditionally on the environment $\bar{Y} = \bar{y}$, the proof follows the same steps as in Corollary 2.6 of [19], the main idea of which we recall briefly here. The individual (g, i) inherits the weak type 0 if and only if it chooses all its $K_{(g,i)}$ potential parents to be of type 0. Hence, the probability of this event is given by $\mathbb{E}[x^{K_{(g,i)}} \mid \bar{y}]$.

If generation g does not see an extreme reproductive event (which happens with probability $1 - c$) all individuals choose all their potential parents independently and the total number of individuals obtaining all type 0 potential parents is binomial with parameters N and

$$\mathbb{E}[x^{K_{(g,i)}} | \bar{y}] = \varphi_{y_g}(x).$$

For generations with extreme reproductive events (occurring with probability c), first $V \sim \Lambda$ and $I_g \sim \mathcal{U}_{[N]}$ are sampled. I_g will be the label of the preferred individual. For each of its choices of potential parents, each individual then flips a coin with (the same) random bias V . In the case of success, it will choose the preferred individual as a potential parent, i.e. label I_g . The random variable B_x is precisely the indicator of the event that the special parent I_g carries type 0, conditionally on the frequency x of 0-type individuals in the previous generation $g - 1$. If the V -coin returns a failure, the individual chooses this potential parent uniformly in $[N]$ and hence the probability of choosing a weak one is x . Therefore, the probability of choosing *one* weak potential parent is given by

$$VB_x + (1 - V)x.$$

Consequently, since an individual is weak only if *all* his potential parents are of the weak type 0, the probability for this even is

$$\varphi_{y_g}(VB_x + (1 - V)x).$$

Since the individuals act independently conditioned on I_g and V , the number of weak individuals in the next generation, conditioned on I_g and V is a binomial with parameters N and $\varphi_{y_g}(VB_x + (1 - V)x)$. \square

Next we move to the proof of the sampling dualities between the 0-allele-frequency process given in Definition 2.2 and the genealogy given in Definition 2.5. They are a direct consequence of the graph-construction, as observed in [31]. We begin with the proof of Proposition 2.9, as it uses the same idea as the proofs of Lemma 2.6 and Proposition 2.8, but in a simpler set-up.

The proofs of Lemma 2.6 and Proposition 2.8 follow the same idea, except that now we consider sampling probabilities when the frequency is given in the *previous* generation. Since the random environment acts between generations, it will appear explicitly in our calculations here. Also, since the probability of an individual being of the 0 type is then determined by the types of its potential parents and these are sampled with replacement, the resulting structure is that of sampling *with* replacement and thus obtains the form of a generating function.

PROOF OF PROPOSITION 2.9. Note that \tilde{H}_N defined in (13) gives the probability that $n \leq N$ individuals sampled uniformly without replacement in a given generation are all of type 0 if the frequency of type 0 is $x \in [N]_0/N$ in *that same generation*. In order to prove (14), we show that both sides express the same quantity, namely the probability, that $n \leq N$ individuals sampled uniformly without replacement from a generation $s \in \mathbb{Z}$ are all of type 0, given that the frequency of 0 was $x \in [N]_0/N$ in generation $r < s$, which we denote by $p(x, n)$.

First, we calculate this probability using the frequency process: We know that the frequency of the 0-allele in generation s is given by $X^{N,r}(s)$, i.e. that there are $X^{N,r}(s)N$ many weak individuals and thus

$$p(x, n) = \mathbb{E} \left[\prod_{k=1}^n \frac{X^{N,r}(s)N - k}{N - k} \mid X^{N,r}(r) = x \right] = \mathbb{E}[\tilde{H}_N(X^{N,r}(s), n) \mid X^{N,r}(r) = x].$$

On the other hand, we can also calculate it using the fact that the individuals in generation s will only be of type 0 if all their ancestors in generation r are of type 0, i.e. using the ancestral process. The frequency of type 0 individuals in generation r is x and thus

$$p(x, n) = \mathbb{E} \left[\prod_{k=1}^{Z^{N,s}(r)} \frac{xN - k}{N - k} \mid Z^{N,s}(r) = n \right] = \mathbb{E}[\tilde{H}_N(x, Z^{N,s}(r)) \mid Z^{N,s}(r) = n],$$

which completes the proof. \square

PROOF OF LEMMA 2.6. To prove equality (10), we show that both sides express the same quantity. Denote by $p(\bar{Y}; x, n)$ the conditional probability, given the random environment \bar{Y} , that $n \leq N$ individuals, sampled uniformly at random without replacement from generation $s \in \mathbb{Z}$, are all of type 0, given that the frequency of 0-type individuals was $x \in [N]_0/N$ in generation $r < s$. Without loss of generality, we may assume that the sample in generation s consists of the individuals with the labels $1, \dots, n$ and therefore

$$p(\bar{Y}; x, n) := \mathbb{P}((s, 1), \dots, (s, n) \text{ are all of type 0} \mid X^{N,r}(r) = x, \bar{Y}).$$

We will calculate this probability in two ways: first, by considering the choices of potential parents of the individuals in our sample $(s, 1), \dots, (s, n)$ in the previous generation. Second, by tracing their potential ancestors further back in time up until generation $r + 1$ and then considering the choices of these potential ancestors. We will, thus, be working both with the frequency process $X^{N,r}$ as well as the genealogical process $Z^{N,s}$. Before we come to these calculations let us make the following two observations similar to the calculations done in the proof of Proposition 2.3.

For any generation $g \in \mathbb{Z}$, conditionally on Y_g the random variables $K_{(g,1)}, \dots, K_{(g,n)}$ are independent and identically distributed with distribution $Q(Y_g)$. Hence, for any $\tilde{x} \in [0, 1]$

$$(27) \quad \mathbb{E}[\tilde{x}^{\sum_{i=1}^n K_{(g,i)}} \mid \bar{Y}] = (\mathbb{E}[\tilde{x}^{K_{(g,1)}} \mid \bar{Y}])^n = (\mathbb{E}[\varphi_{Y_g}(\tilde{x}) \mid \bar{Y}])^n = (\varphi_{Y_g}(\tilde{x}))^n.$$

Recall the independent variables $V \sim \Lambda$ and $B_{\tilde{x}} \sim \text{Ber}_{\tilde{x}}$ introduced for (9) and let $(V_g, g \in \mathbb{Z})$, be iid copies of V and $(B_{\tilde{x},g}, g \in \mathbb{Z})$ be iid copies of $B_{\tilde{x}}$, independent of everything else. Then, similarly to the last observation also

$$(28) \quad \mathbb{E} \left[((1 - V_g)\tilde{x} + V_g(B_{\tilde{x},g}))^{\sum_{i=1}^n K_{(g,i)}} \mid \bar{Y} \right] = \mathbb{E} \left[(\varphi_{Y_g}((1 - V_g)\tilde{x} + V_g(B_{\tilde{x},g})))^n \mid \bar{Y} \right].$$

All equalities between random variables in the following hold \mathbb{P} -a.s.

We now first calculate $p(\bar{Y}; x, n)$ with the following idea: all individuals in our sample are of type 0, if and only if all the potential parents they choose in the previous generation are of type 0 and the frequency of type 0 in that generation is given by $X^{N,r}(s - 1)$. Observe that since $K_{(s,1)}, \dots, K_{(s,n)}$ are also independent of $X^{N,r}(r)$, $X^{N,r}(s - 1)$, V_s and $\xi(I_s)$, this implies :

- (i) In a generation s without skewed offspring distribution, the probability for all (s, i) , $i = 1, \dots, n$, to be weak, conditional on the frequency of the weak type in the previous generation $X^{N,r}(s - 1)$, using (27) is

$$\begin{aligned} & \mathbb{E} \left[[X^{N,r}(s - 1)]^{\sum_{i=1}^n K_{(s,i)}} \mid X^{N,r}(s - 1), X^{N,r}(r) = x, \bar{Y} \right] \\ &= \mathbb{E} \left[(\varphi_{Y_s}(X^{N,r}(s - 1)))^n \mid X^{N,r}(s - 1), X^{N,r}(r) = x, \bar{Y} \right] \end{aligned}$$

Denote this probability by $p_1(X^{N,r}(s - 1), \bar{Y}; x, n)$.

(ii) In a generation with skewed offspring distribution, recall that I_s denotes the favoured individual (special marked label) collectively chosen at time s and $\xi(I_s)$ its type. By construction, the distribution of $1 - \xi(I_s)$ conditional on the frequency of weak in the previous generation is Bernoulli with parameter $X^{N,r}(s-1)$. Hence, in such a generation, the conditional probability for all (s, i) , $i = 1, \dots, n$, to be weak, conditional on the frequency of the weak type in the previous generation is, by (7),

$$\begin{aligned} & \mathbb{E} \left[[(1 - V_s)X^{N,r}(s-1) + V_s(1 - \xi(I_s))]^{K_{(s,i)}} \mid X^{N,r}(s-1), X^{N,r}(r) = x, \bar{Y} \right] \\ &= \mathbb{E} \left[\left(\varphi_{Y_s^N}((1 - V_s)X^{N,r}(s-1) + V_s(B_{X^{N,r}(s-1)})) \right)^n \mid X^{N,r}(s-1), X^{N,r}(r) = x, \bar{Y} \right]. \end{aligned}$$

Denote this probability by $p_2(X^{N,r}(s-1), \bar{Y}, x)$.

Combining this, we obtain

$$\begin{aligned} p(\bar{Y}; x, n) &= (1 - c) \mathbb{E} \left[p_1(X^{N,r}(s-1), \bar{Y}; x) \mid X^{N,r}(r) = x, \bar{Y} \right] \\ &\quad + c \mathbb{E} \left[p_2(X^{N,r}(s-1), \bar{Y}; x) \mid X^{N,r}(r) = x, \bar{Y} \right] \\ &= (1 - c) \mathbb{E} \left[(\varphi_{Y_s^N}(X^{N,r}(s-1)))^n \mid X^{N,r}(r) = x, \bar{Y} \right] \\ &\quad + c \mathbb{E} \left[\left(\varphi_{Y_s^N}((1 - V_s)X^{N,r}(s-1) + V_s(B_{X^{N,r}(s-1)})) \right)^n \mid X^{N,r}(r) = x, \bar{Y} \right] \\ &= \mathbb{E} [H(X^{N,r}(s-1), n; s, \bar{Y}) \mid X^{N,r}(r) = x, \bar{Y}], \end{aligned}$$

which is the left-hand-side of the statement of Lemma 2.6.

On the other hand, we can calculate the $p(\bar{Y}^{N,s}; x, n)$ by looking at the number of potential ancestors of the sample in generation $r+1$: these, too, need to be all of the weak type 0 in order for our sample to be of type 0. This in turn is only possible if all *their* potential ancestors in the previous generation r are of the weak type, the frequency of which is given by x . Note that we can add a condition on the event $Z^{N,s}(s) = n$ to $p(\bar{Y}; x, n)$ without altering the probability. These two observations mean

$$\begin{aligned} p(\bar{Y}; x, n) &= \mathbb{P}((s, 1), \dots, (s, n) \text{ are all of type 0} \mid Z^{N,s}(s) = n, X^{N,r}(r) = x, \bar{Y}) \\ &= \mathbb{P}(A_{s-r-1}^N(\{(s, 1), \dots, (s, n)\}) \text{ are all of type 0} \mid X^{N,r}(r) = x, \bar{Y}) \end{aligned}$$

The number of potential ancestors of $(s, 1), \dots, (s, n)$ in generation $r+1$ is

$$|A_{s-r-1}^N(\{(s, 1), \dots, (s, n)\})| = Z^{N,s}(r+1)$$

and the frequency of the weak type 0 in generation r is, by assumption, equal to x . Since each of the individuals in $A_{s-r-1}^N(\{(s, 1), \dots, (s, n)\})$ chooses its potential parents independently, using the same observations as in the first step the sought probability can be expressed as

$$\begin{aligned} p(\bar{Y}; x) &= (1 - c) \mathbb{E} \left[x^{\sum_{i=1}^{Z^{N,s}(r+1)} K_{(r+1,i)}} \mid Z^{N,s}(s) = n, X^{N,r}(r) = x, \bar{Y} \right] \\ &\quad + c \mathbb{E} \left[[(1 - V_{r+1})x + V_{r+1}(1 - \xi(I_{r+1}))]^{\sum_{i=1}^{Z^{N,s}(r+1)} K_{(s,i)}} \mid Z^{N,s}(s) = n, X^{N,r}(r) = x, \bar{Y} \right] \\ &= (1 - c) \mathbb{E} \left[\varphi_{Y_g}(x)^{Z^{N,s}(r+1)} \mid Z^{N,s}(s) = n, \bar{Y} \right] \end{aligned}$$

$$\begin{aligned}
& + c \mathbb{E} \left[\varphi_{Y_g} \left((1 - V_{r+1})x + V_{r+1}(B_x) \right)^{Z^{N,s}(r+1)} \mid Z^{N,s}(s) = n, \bar{Y} \right] \\
& = \mathbb{E} \left[H(x, Z^{N,s}(r+1); r+1, \bar{Y}) \mid Z^{N,s}(s) = n, \bar{Y} \right]
\end{aligned}$$

where, again, we have profited from the fact that, conditionally on \bar{Y} , the vector $(K_{(r+1,i)} : i \in [N])$ has *iid* coordinates and is independent of $Z^{N,s}(r+1)$ and $Z^{N,s}(s)$, along with the observations (27) and (28). \square

We now turn to the proof of the unconditioned duality.

PROOF OF PROPOSITION 2.8. The idea of the proof is the same as for Lemma 2.6, but we have to take the *iid* environment carefully into account. Let $g \geq 0$. First note that, by identity in distribution,

$$\begin{aligned}
(29) \quad H_\mu(x, n) &= \mathbb{E}[H(x, n; 1, \bar{Y})] \\
&= (1 - c)\mathbb{E}[\varphi_{Y_1}(x)^n] + c \mathbb{E}[\varphi_{Y_1}((1 - V)x + VB_x)^n] \\
&= \mathbb{E}[H(x, n; g+1, \bar{Y})]
\end{aligned}$$

$$(30) \quad = (1 - c)\mathbb{E}[\varphi_{Y_{g+1}}(x)^n] + c \mathbb{E}[\varphi_{Y_{g+1}}((1 - V)x + VB_x)^n]$$

and with the new notation

$$(31) \quad \mathbb{E}^n [H_\mu(x, Z^{N,0}(-g))] = \mathbb{E}^n [H_\mu(x, Z^{N,g+1}(1))] .$$

We again seek to determine the same probability, except that this time we do not condition on the random environment, i.e. we search

$$p(x, n) := \mathbb{P}_x((g+1, 1), \dots, (g+1, n) \text{ are all of type } 0)$$

As before, this probability can, on one hand, be expressed using $X^{N,0}(g)$ as

$$\begin{aligned}
p(x, n) &= (1 - c)\mathbb{E}_x[(X^{N,0}(g))^{\sum_{i=1}^n K_{(g+1,i)}}] \\
&\quad + c \mathbb{E}_x \left[\left((1 - V_{g+1})X^{N,0}(g) + V_{g+1}(1 - \xi(I_{g+1})) \right)^{\sum_{i=1}^n K_{(g+1,i)}} \right] \\
&= (1 - c)\mathbb{E}_x \left[\varphi_{Y_{g+1}}(X^{N,0}(g))^n \right] \\
&\quad + c \mathbb{E}_x \left[\varphi_{Y_{g+1}} \left((1 - V_{g+1})X^{N,0}(g) + V_{g+1}B_{X^{N,0}(g)} \right)^n \right] \\
&= \mathbb{E}_x [H_\mu(X^{N,0}(g), n)] .
\end{aligned}$$

For the second equality we used that $(K_{(g+1,1)}, \dots, K_{(g+1,n)}, Y_{g+1})$ and $X^{N,0}(g)$ are independent, and that conditional on Y_{g+1} the $K_{(g+1,1)}, \dots, K_{(g+1,n)}$ are *iid*, which in particular implies for any $\tilde{x} \in [0, 1]$:

$$(32) \quad \mathbb{E}[\tilde{x}^{\sum_{i=1}^n K_{(g+1,i)}}] = \mathbb{E}[\mathbb{E}[\tilde{x}^{K_{(g+1,1)}} \mid \bar{Y}]^n] = \mathbb{E}[\varphi_{Y_{g+1}}(\tilde{x})^n]$$

and likewise

$$\begin{aligned}
(33) \quad & \mathbb{E}[(1 - V_g)\tilde{x} + V_g B_{\tilde{x},g}]^{\sum_{i=1}^n K_{(g+1,i)}} \\
&= \mathbb{E} \left[\mathbb{E}[(1 - V_g)\tilde{x} + V_g B_{\tilde{x},g}]^{K_{(g+1,1)}} \mid V_g, B_{\tilde{x},g}, \bar{Y} \right]^n \\
&= \mathbb{E}[\varphi_{Y_{g+1}}((1 - V_g)\tilde{x} + V_g B_{\tilde{x},g})^n]
\end{aligned}$$

Note that the random variables $K_{(g+1,1)}, \dots, K_{(g+1,n)}$ are normally only *conditionally* independent, given a realisation of Y_{g+1} , which is why we cannot, in the unconditioned case, write the duality function as a product of expectations.

On the other hand, we can calculate the same probability $p(x, n)$ considering the ancestry with $Z^{N,g+1}(1)$ through

$$\begin{aligned} p(x) &= (1-c)\mathbb{E}[x^{\sum_{i=1}^{Z^{N,g+1}(1)} K_{(1,i)}}] + c\mathbb{E}_x\left[\left((1-V_1)x + V_1(1-\xi(I_1))\right)^{\sum_{i=1}^{Z^{N,g+1}(1)} K_{(1,i)}}\right] \\ &= (1-c)\mathbb{E}_x\left[\varphi_{Y_1}(x)^{Z^{N,g+1}(1)}\right] + c\mathbb{E}_x\left[\varphi_{Y_1}\left((1-V_1)x + V_1B_x\right)^{Z^{N,g+1}(1)}\right] \\ &= \mathbb{E}[H_\mu(x, Z^{N,g+1}(1))] \end{aligned}$$

where we again profited from the independence of $Z^{N,g+1}(1)$ and $(K_{(1,1)}, \dots, K_{(1,N)}, \bar{Y})$ and the observations (32) and (33) with $g+1$ replaced by 1. Hence, we have proven

$$\mathbb{E}_x[H_\mu(X^{N,0}(g), n)] = \mathbb{E}^n[H_\mu(x, Z^{N,g+1}(1))] \quad \square$$

which together with (31) completes the proof. \square

5.2. Scaling limits. We begin with the proof of existence and uniqueness of a solution to the SDE.

PROOF OF LEMMA 3.1. For the case without the third summand in the generator (15) the statement of Lemma 3.1 was already proved in [4, Cor. 3.3]. For the general case, the proof is an application of Theorem 5.1 in [30], of which we only need to verify conditions 3a), 3b) and 5a). Using the notation of [30], in our case the relevant functions are $\sigma(x) = \sqrt{\sigma x(1-x)}\mathbb{1}_{[0,1]}(x)$, $b(x) = wx(1-x)$, $g_0(x, (z, u)) = z(\mathbb{1}_{[0,x]}(u) - x)\mathbb{1}_{[0,1]}(x)$ for $U_0 = [0, 1]^2$ and $g_1(x, y) = \mathbb{E}[x^{K_y} - x]\mathbb{1}_{[0,1]}(x)$ for $U_1 = [0, 1]$. All conditions are easy to check for g_0 and were shown for b and σ in Equation (3.a) and (26) of [19], respectively, so we only need to concern ourselves with the selection component g_1 .

Note that $|\frac{\partial g_1}{\partial x}(x, y)| = |\mathbb{E}[K_y x^{K_y-1} - 1]| \leq \max\{\mathbb{P}(K_y \geq 2), \mathbb{E}[K_y - 1]\} \leq \mathbb{E}[K_y - 1]$, for all $x \in [0, 1]$. Then, by the mean value theorem, for any $x_1, x_2 \in [0, 1]$

$$\int_{[0,1]} |g_1(x_1, y) - g_1(x_2, y)| \mu(dy) \leq |x_1 - x_2| \int_{[0,1]} \mathbb{E}[K_y - 1] \mu(dy)$$

Since the integral is finite by Condition 1, this gives us condition 3a) of [30]. The observation above also shows that $\frac{dg_1}{dx}$ is $(\mu(dy) \otimes dx)$ -integrable on $[0, 1] \times [0, 1]$, because

$$\int_{[0,1] \times [0,1]} \left| \frac{\partial g_1}{\partial x}(x, y) \right| \mu(dy) \otimes dx \leq \int_{[0,1]} \mathbb{E}[K_y - 1] \mu(dy) < \infty,$$

whence we can use Fubini's Theorem and, using the fact that $g_1(x, y)^2 \leq |g_1(x, y)|$, estimate

$$\begin{aligned} \int_{[0,1]} (g_1(x, y))^2 \mu(dy) &\leq \int_{[0,1]} \int_{[0,x]} \left| \frac{dg_1}{ds}(s, y) \right| ds \mu(dy) \\ &\leq x \int_{[0,1]} \mathbb{E}[K_y - 1] \mu(dy) \end{aligned}$$

for $x \in [0, 1]$, which yields the estimates for condition 5a) of [30]. \square

We now move to the proof of convergence of the discrete frequency process defined in Definition 2.2 to the scaling limit (15) for $w = 0$, i.e with only rare selection, started in $X_0 = x$

PROOF OF THEOREM 3.3. We will prove strong convergence of the sequence of generators corresponding to X^N , $N \in \mathbb{N}$ to the generator of X on $C^2([0, 1])$. Our claim then follows from Theorem 19.28 in [24] (Theorem 17.28 in the other editions) which however requires the limit process X to be also Feller. The Feller property is established in Section 3.1, Lemma 3.9 and will be assumed true in this proof.

Denote with \mathcal{A}^N the discrete generator of $(X^N(\lfloor \rho_N^{-1}t \rfloor) : t \geq 0)$. We calculate \mathcal{A}^N conditioning on the four possible cases reflecting occurrence vs non-occurrence of selection or multiple mergers. To this end, starting without loss of generality from generation $g_0 = 0$, let B_c and B_s be the random variables indicating whether selection, resp. multiple mergers occur in the first time-step. By construction of the reproduction mechanism, they are independent Bernoulli random variables with success parameter c_N and $\rho_N \gamma_N([0, 1])$ respectively. For C_2 functions $f : [0, 1] \mapsto \mathbb{R}$ the generator then acts according to the rule

$$(34) \quad \begin{aligned} \mathcal{A}^N f(x) &:= \rho_N^{-1} \mathbb{E}_x [f(X_1^N) - f(x)] \\ &= \rho_N^{-1} \left\{ (1 - \rho_N \gamma_N([0, 1]))(1 - c_N) \left(\frac{1}{2N} x(1-x)f''(x) + O\left(\frac{1}{N^2}\right) \right) \right. \\ &\quad + (1 - \rho_N \gamma_N([0, 1]))c_N \mathbb{E}_x [f(X_1^N) - f(x) | B_c = 1, B_s = 0] \\ &\quad + \rho_N \gamma_N([0, 1])(1 - c_N) \mathbb{E}_x [f(X_1^N) - f(x) | B_c = 0, B_s = 1] \\ &\quad \left. + \rho_N \gamma_N([0, 1])c_N \mathbb{E}_x [f(X_1^N) - f(x) | B_c = 1, B_s = 1] \right\}. \end{aligned}$$

The first summand corresponds to the standard neutral Wright-Fisher model without multiple mergers, hence assumption (17) implies the convergence to the first component of the generator given in (16).

Since the last expectation is bounded, assumption (19) implies that the last summand vanishes as $N \rightarrow \infty$, so we only need to calculate carefully the remaining two summands:

For the first of these two – the case of only skewed offspring distribution – observe that, since $B_c = 1, B_s = 0$, by virtue of the reproduction mechanism (7), for a given initial frequency $X_0^N = x$, strength of correlation (multiple mergers) $C_1^N = u \in [0, 1]$ and type $B = b \in \{0, 1\}$ of the favoured parent, NX_1^N is simply a Binomial random variable with parameters N and $(1-u)x + u(1-b)$. Also note that given $X_0^N = x$, B is a Bernoulli random variable with success parameter $1-x$.

Apply Taylor's expansion to f around $(1 - C_1^N)x + C_1^N(1 - B)$. We obtain, using the Lagrange remainder term, for an η depending on X_1^N, C_1^N, B and x ,

$$\begin{aligned} &\mathbb{E}_x [f(X_1^N) - f(x) | B_c = 1, B_s = 0] \\ &= \mathbb{E}_x \left[f((1 - C_1^N)x + C_1^N(1 - B)) - f(x) \middle| B_c = 1, B_s = 0 \right] \\ &\quad + \mathbb{E}_x \left[\frac{1}{2} f''(\eta) (X_1^N - (1 - C_1^N)x - C_1^N(1 - B))^2 \middle| B_c = 1, B_s = 0 \right] \end{aligned}$$

Using the independence of U_1^N, B, B_c , and B_s , we can rewrite the first term:

$$\begin{aligned} &\rho_N^{-1} c_N \mathbb{E}_x \left[f((1 - C_1^N)x + C_1^N(1 - B)) - f(x) \middle| B_c = 1, B_s = 0 \right] \\ &= \frac{\rho_N^{-1} c_N}{\Lambda_N^\alpha([0, 1])} \int_{[0, 1]} (xf((1-u)x+u) + (1-x)f((1-u)x) - f(x)) \frac{1}{u^2} \mathbb{1}_{[N^{-\alpha}, 1]}(u) \Lambda_c(du) \\ &\xrightarrow{N \rightarrow \infty} c \int_{[0, 1]} (xf((1-u)x+u) + (1-x)f((1-u)x) - f(x)) \frac{1}{u^2} \Lambda_c(du). \end{aligned}$$

The convergence in the last step follows from (18) and is uniform in $x \in [0, 1]$.

Using the conditional variance of NX_1^N given C_1^N and B allows us to estimate

$$\begin{aligned}
& \left| \mathbb{E}_x \left[\frac{1}{2} f''(\eta) (X_1^N - (1 - C_1^N)x - C_1^N(1 - B))^2 \middle| B_c = 1, B_s = 0 \right] \right| \\
& \leq \frac{1}{2} \sup_{z \in [0, 1]} |f''(z)| \left| \mathbb{E}_x \left[\frac{N(1 - (1 - C_1^N)x - C_1^N(1 - B))((1 - C_1^N)x + C_1^N(1 - B))}{N^2} \right] \right| \\
& = \frac{1}{2} \sup_{z \in [0, 1]} |f''(z)| \frac{1}{N \bar{\Lambda}_N^\alpha([0, 1])} \\
& \quad \times \left| \int_{[N^{-\alpha}, 1]} \mathbb{E}_x \left[(1 - (1 - u)x - u(1 - B))((1 - u)x + u(1 - B)) \right] \frac{1}{u^2} \Lambda_c(du) \right| \\
& \leq \frac{1}{2} \sup_{z \in [0, 1]} |f''(z)| N^{-1}.
\end{aligned}$$

In the second step we simply used the independence of C_1^N and B as well as the definition of $\bar{\Lambda}_N^\alpha$ and in the last that the expectation can be bounded by 1. Since $(C_N)_{N \in \mathbb{N}}$ converges to 0, assumption (17) implies that this term vanishes. Note that also this bound is uniform in $x \in [0, 1]$.

In the same spirit as for the skewed offspring distribution, given that we only observe rare selection, i.e. $B_c = 0, B_s = 1$, the distribution of NX_1^N conditioned on a strength of selection $Y_1^N = y \in [0, 1]$ is again binomial with parameters N and $\mathbb{E}[x^{K_y}]$. Applying Taylor's formula in this point one obtains for an η depending on X_1^N, Y_1^N and x :

$$\begin{aligned}
& \mathbb{E}_x [f(X_1^N) - f(x) | B_c = 0, B_s = 1] \\
& = \mathbb{E}_x \left[f(\mathbb{E}[x^{K_{Y_1^N}} | Y_1^N]) - f(x) | B_c = 0, B_s = 1 \right] \\
& \quad + \mathbb{E}_x \left[\frac{1}{2} f''(\eta) (X_1^N - \mathbb{E}[x^{K_{Y_1^N}} | Y_1^N])^2 | B_c = 0, B_s = 1 \right].
\end{aligned}$$

Again, using the conditional variance of NX_1^N given Y_1^N , we estimate

$$\begin{aligned}
& \mathbb{E}_x \left[\frac{1}{2} f''(\eta) (X_1^N - \mathbb{E}[x^{K_{Y_1^N}} | Y_1^N])^2 | B_c = 0, B_s = 1 \right] \\
& \leq \frac{1}{2} \max_{z \in [0, 1]} f''(z) \frac{1}{N} \mathbb{E}_x \left[\mathbb{E}[x^{K_{Y_1^N}} | Y_1^N] (1 - \mathbb{E}[x^{K_{Y_1^N}} | Y_1^N]) | B_c = 0, B_s = 1 \right] \\
& \leq \frac{1}{2} \max_{z \in [0, 1]} f''(z) \frac{1}{N}
\end{aligned}$$

since the last expectation is bounded by 1. Since $\rho_N \gamma_N([0, 1])$ vanishes, assumption (17) proves that this term also vanishes in the limit.

To calculate the first expectation on the other hand, note that conditioned on $B_s = 1, Y_1^N$ is distributed according to $\bar{\gamma}_N$. Hence,

$$\begin{aligned}
& \gamma_N([0, 1]) \mathbb{E}_x \left[f(\mathbb{E}[x^{K_{Y_1^N}} | Y_1^N]) - f(x) | B_c = 0, B_s = 1 \right] \\
& = \gamma_N([0, 1]) \int_{[0, 1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \bar{\gamma}_N(dy) \\
& = \int_{[0, 1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \mathbf{1}_{I_N}(y) \mu(dy) \xrightarrow{N \rightarrow \infty} \int_{[0, 1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \mu(dy).
\end{aligned}$$

To see that this convergence is uniform in $x \in [0, 1]$, recall Condition 1 and observe that using the mean-value-theorem

$$\begin{aligned} & \left| \int_{[0,1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \mathbb{1}_{I_N^c}(y) \mu(dy) \right| \\ &= \left| \int_{[0,1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \mathbb{1}_{I_N^c}(y) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s(dy) \right| \\ &\leq \max_{z \in [0,1]} |f'(z)| \int_{[0,1]} |\mathbb{E}[x^{K_y}] - x| \mathbb{1}_{I_N^c}(y) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s(dy). \end{aligned}$$

Since K_y only takes integer values, as observed in the proof of Lemma 3.1, $|\mathbb{E}[x^{K_y}] - x| \leq \mathbb{E}[K_y - 1]$ and the claim follows since Λ_s is a distribution.

Therefore we have proven that the assumptions of the theorem ensure $\mathcal{A}^N f \rightarrow \mathcal{A}f$ uniformly on the compact set $[0, 1]$. Hence, using Theorem 19.28 in [24] (Theorem 17.28 in the other editions) we conclude the desired weak convergence. \square

We now move to the proof of convergence of the discrete frequency process defined in Definition 2.2 to the scaling limit (15) with weak and rare selection.

PROOF OF THEOREM 3.5. The proof is exactly analogous to that of Theorem 3.3. As done there, one can condition, with the help of suitable Bernoulli random variables, on whether or not now *rare* selection or multiple mergers occur. For the first summand in (34), however, one now calculates

$$\begin{aligned} & \mathbb{E}_x[f(X_1^N) - f(x) | B_c = 0, B_s = 0] \\ &= -w_N \frac{x(1-x)}{1-xw_N} f'(x) + \frac{1}{2N} x(1-x) f''(x) \frac{1-w_N}{(1-xw_N)^2} + O(w_N^2) + O(N^{-2}) \end{aligned}$$

(doing a Taylor expansion around x). The three other summands coincide with those in the proof of Theorem 3.3 and the claim follows. \square

After having proved the results for the forward frequency process, we now turn to the proofs on the branching coalescing process. We begin by showing that it is Feller and conservative under our main Condition 1.

PROOF OF LEMMA 3.7. To prove that Z is Feller we have to show that its semigroup given by $P_t f(n) = \mathbb{E}_n[f(Z_t)]$ sends C to C . (Recall that we consider $\mathbb{N} \cup \{\infty\}$ with the metric of the harmonic numbers and are thus on a compact, where C , C_b and C_0 coincide.) Note that, for such a continuous function f , only the continuity of $p_t f$ in ∞ needs to be checked.

In order to do that we will construct a coupling of Z started at any possible number of initial particles/lineages. To this end, for any finite number of initial particles $l \in \mathbb{N}$ we define the process $(B^{(l)}(t) : t \geq 0)$ with $B^{(l)}(t) := (B_t^{(1,l)}, B_t^{(2,l)}, \dots, B_t^{(l,l)})$ with starting condition $(B_0^{(1,l)}, B_0^{(2,l)}, \dots, B_0^{(l,l)}) = (1, 1, \dots, 1)$. Each of the $B^{(i,l)}$ describes the individuals resulting from one initial particle in Z , i.e. it traces the offspring numbers of that one individual resulting from (spontaneous or coordinated) branching events and (pairwise or coordinated) coalescence events. While the former can be assigned to each individual, the coalescence events require interaction between the $B^{(i,l)}$, $i, j = 1, \dots, l$, as is reflected in the following

rates. $(B^{(l)}(t) : t \geq 0)$ branches from $(n_1, \dots, n_l) \in \mathbb{N} \times \mathbb{N}_0 \times \dots \times \mathbb{N}_0$ to $(n_1 + k_1, \dots, n_l + k_l)$ given by

$$\int_{[0,1]} \prod_{i=1}^l \mathbb{P} \left(\sum_{j=1}^{n_i} K_{y,j} = n_i + k_i \right) \mu(dy) + \sum_{i=1}^l w n_i \delta_{e_i, (k_1, \dots, k_l)}$$

which corresponds to each individual independently following its own (coordinated and spontaneous) branching mechanism. Coalescence events allow the process $(B^{(l)}(t) : t \geq 0)$ to go from a state (n_1, \dots, n_l) to $(n_1 - k_1, \dots, n_l - k_l) + e(k_1, \dots, k_l)$, where $e(k_1, \dots, k_l)$ is the unit vector in $\min\{j \mid k_j > 0\}$. A coalescence can thus reduce the number of lineages in each $B^{(i,l)}$ and will assign the surviving lineage to the offspring group with the lowest label. The rate for this event is

$$\begin{aligned} c \int_{[0,1]} \prod_{i=1}^l \binom{n_i}{k_i} y^{k_i} (1-y)^{n_i-k_i} \frac{\Lambda_c(dy)}{y^2} \\ + \sum_{i=1}^l \sigma \binom{n_i}{2} \delta_{2e_i, (k_1, \dots, k_l)} + \sigma \sum_{i=1}^l \sum_{j=1}^{i-1} n_i n_j \delta_{e_i + e_j, (k_1, \dots, k_l)}. \end{aligned}$$

The first term accounts for the coordinated coalescences across all offspring populations; the second accounts for pairwise coalescences within the same offspring group and the last term for pairwise coalescences across offspring populations.

Note that the processes are constructed such that a $B^{(i,l)}$ will always only notice the action of or interaction with a $B^{(j,l)}$ of a lower label $j \leq i$. Thus the laws of the processes $(B^{(l)}(t) : t \geq 0)$ form consistent family in $l \in \mathbb{N}$ and by Kolmogorov's extension theorem can be extended to a process $(B^{(\infty)}(t) : t \geq 0)$ with $B^{(\infty)}(t) := (B^{(1)}(t), B^{(2)}(t), \dots)$.

This gives the desired coupling of Z started with different initial conditions in the following way: If we denote by $Z^{(n)}$ the process Z starting in n , then, in distribution,

$$Z_t^{(n)} \stackrel{\mathcal{L}}{=} \sum_{i=1}^n B_t^{(i)} \quad \text{and} \quad Z_t^{(\infty)} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{\infty} B_t^{(i)}.$$

Note that, by construction $\lim_{n \rightarrow \infty} \sum_{i=1}^n B_t^{(i)} = \sum_{i=1}^{\infty} B_t^{(i)}$, almost surely. Hence, for f continuous on $N \cup \{\infty\}$, we get

$$\begin{aligned} |P_t f(n) - P_t f(\infty)| &= |\mathbb{E}^n[f(Z(t))] - \mathbb{E}^\infty[f(Z(t))]| = |\mathbb{E}[f(Z^{(n)}(t))] - \mathbb{E}[f(Z^{(\infty)}(t))]| \\ &= \left| \mathbb{E} \left[f \left(\sum_{i=1}^n B^{(i)}(t) \right) - f \left(\sum_{i=1}^{\infty} B^{(i)}(t) \right) \right] \right| \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

by bounded convergence, which implies that $P_t f \in C$ and thus that Z is indeed Feller.

We now turn to the proof of conservativeness of Z . We will show that the branching mechanism itself is non-explosive, even without the “help” of the fast coalescence rate. To this end, let $W = (W(t) : t \geq 0)$ be the process with the same branching mechanism as $Z = (Z(t) : t \geq 0)$ but without coalescence. It is clear that then stochastically $W(t) \geq Z(t)$ for any $t \geq 0$. For any $m \in \mathbb{N}$ define $W^m = (W^m(t) : t \geq 0)$ with $W^m(t) := \min\{W(t), m\}$, the branching process truncated at m . Since W is non-decreasing, W^m is also Markovian with generator

$$Gf(k) = \sum_{l=k}^{m-1} \left\{ \int_{[0,1]} \mathbb{P} \left(\sum_{j=1}^k K_{y,j} = l \right) \mu(dy) + w k \delta_{l-k,1} \right\} [f(l) - f(k)]$$

$$+ \int_{[0,1]} \mathbb{P} \left(\sum_{j=1}^k K_{y,j} \geq m \right) \mu(dy) [f(m) - f(k)]$$

where the $K_{y,1}, \dots, K_{y,k}$ are *iid* with common distribution $Q(y)$. For the identity on $\{1, \dots, m\}$, this simplifies to

$$\begin{aligned} \text{GIId}(k) &= \sum_{l=k}^{m-1} \left\{ \int_{[0,1]} \mathbb{P} \left(\sum_{j=1}^k K_{y,j} = l \right) \mu(dy) + wk\delta_{l-k,1} \right\} [l - k] \\ &\quad + \int_{[0,1]} \mathbb{P} \left(\sum_{j=1}^k K_{y,j} \geq m \right) \mu(dy) [m - k] \\ &= wk + \int_{[0,1]} \left\{ \sum_{l=k}^{m-1} \mathbb{P} \left(\sum_{j=1}^k K_{y,j} = l \right) [l - k] + \mathbb{P} \left(\sum_{j=1}^k K_{y,j} \geq m \right) [m - k] \right\} \mu(dy) \\ &= wk + \int_{[0,1]} \mathbb{E} \left[\left(\sum_{j=1}^k K_{y,j} \right) \wedge m - k \right] \mu(dy) \\ &\leq wk + \int_{[0,1]} \mathbb{E} \left[\sum_{j=1}^k K_{y,j} - k \right] \mu(dy) \\ &= \left(w + \int_{[0,1]} \mathbb{E} [K_{y,1} - 1] \mu(dy) \right) k \end{aligned}$$

which is finite by Condition 1. Using Dynkin's formula we observe that

$$\begin{aligned} \mathbb{E}^n[W^m(t)] - n &= \int_0^t \mathbb{E}^n[\text{GIId}(W^m(s))] ds \\ &\leq \int_0^t \left(w + \int_{[0,1]} \mathbb{E} [K_{y,1} - 1] \mu(dy) \right) \mathbb{E}^n[W^m(s)] ds \end{aligned}$$

which with Grönwall's inequality implies

$$\mathbb{E}^n[W^m(t)] \leq n \exp \left(t \left\{ w + \int_{[0,1]} \mathbb{E} [K_{y,1} - 1] \mu(dy) \right\} \right) < \infty.$$

Since the bound is uniform in $m \in \mathbb{N}$ and $W^m(t) \uparrow W(t)$ a.s. as $m \rightarrow \infty$ for any $t \geq 0$, we have completed the proof. \square

Given these properties of the branching coalescing process $Z = (Z(t) : t \geq 0)$, we now prove the moment duality between Z and $X = (X(t) : t \geq 0)$.

PROOF OF LEMMA 3.8. The case $n = \infty$ holds by the choice of the entrance law. By the Markov property, for every other n , the proof can be done by calculating the generator applied to the function(s) $f_x(n) := f^n(x) := x^n$. Recall the generator \mathcal{A} of X from (16) and let \mathcal{B} be the generator of Z .

Since the moment duality relations between the Wright-Fisher diffusion and the Kingman coalescent, between the Λ -jump diffusion and the Λ -coalescents and between binary branching and the logistic ODE are well-known [7], the additive structure of the generator allows us to only consider the component of the generators responsible for *rare* selection, which we denote by \mathcal{A}_s and \mathcal{B}_s respectively. For any $y \in [0, 1]$ and $n \in \mathbb{N}$ let $K_{y,j}, j = 1, \dots, n$ be *iid* with distribution $Q(y)$. Then, for any $x \in [0, 1]$ and $n \in \mathbb{N}_0$

$$\begin{aligned} \mathcal{A}_s f^n(x) &= \int_{[0,1]} \{(\mathbb{E}[x^{K_y}])^n - x^n\} \mu(dy) \\ &= \int_{[0,1]} \{\mathbb{E}[x^{\sum_{j=1}^n K_{y,j}}] - x^n\} \mu(dy) \\ &= \int_{[0,1]} \left[\sum_{k=n}^{\infty} x^k \mathbb{P} \left(\sum_{j=1}^n K_{y,j} = k \right) - x^n \right] \mu(dy) \\ &= \int_{[0,1]} \left[\sum_{k=n}^{\infty} \mathbb{P} \left(\sum_{j=1}^n K_{y,j} = k \right) (x^k - x^n) \right] \mu(dy) = \mathcal{B}_s f_x(n) \end{aligned}$$

As mentioned above, by linearity of the operators this implies $\mathcal{A}f^n(x) = \mathcal{B}f_x(n)$ for any $x \in [0, 1]$ and $n \in \mathbb{N}_0$. Hence, since X is a Markovian solution to the martingale problem associated to \mathcal{A} , and Lemma 3.7 implies that Z is Feller, the moment duality between X and Z follows immediately (see e.g. Proposition 6.1 in [18]). \square

PROOF OF LEMMA 3.9. Let $f(x) := x^n$ for fixed $n \in \mathbb{N}_0$. Let $(\tilde{P}_t)_{t \geq 0}$ be the semigroup generated by \mathcal{A} given in (16). Then, using the moment duality from Lemma 3.8

$$\tilde{P}_t f(x) = \mathbb{E}_x[f(X(t))] = \mathbb{E}_x[X(t)^n] = \mathbb{E}^n[x^{Z(t)}].$$

Note that, since Z is conservative, $x \mapsto x^{Z(t)(\omega)}$ is continuous for all $x \in [0, 1]$ for \mathbb{P} -almost all $\omega \in \Omega$. Hence, by bounded convergence and using duality again

$$\lim_{x \rightarrow \bar{x}} \tilde{P}_t f(x) = \lim_{x \rightarrow \bar{x}} \mathbb{E}^n[x^{Z(t)}] = \mathbb{E}^n[\bar{x}^{Z(t)}] = \mathbb{E}_{\bar{x}}[X(t)^n] = \tilde{P}_t f(\bar{x})$$

for all $\bar{x} \in [0, 1]$. Therefore, for every $t \geq 0$, \tilde{P}_t maps monomials to continuous functions. By the linearity of the expectation this also holds for polynomials and by the Stone-Weierstrass-Theorem we have proven the claim. \square

Finally, we use the duality and the previous convergence result for the frequency process, to prove the analogous result for the genealogy.

PROOF OF THEOREM 3.10. The key in this proof is to show that the discrete semigroup P^N of $(Z^N(\lfloor -\rho_N^{-1}t \rfloor) : t \geq 0)$ converges for any $t > 0$ strongly to the semigroup P of $Z := (Z(t) : t \geq 0)$, when applied to functions of the form $f_\lambda(n) = e^{-\lambda n}$, $\lambda > 0$. Since Z is Feller by Lemma 3.7 and the functions f_λ are dense in the set of functions that are continuous and vanish at infinity in this topology, this implies weak convergence by Theorem 19.28, (ii) of [24] (Theorem 17.28 in the other editions).

The convergence of the semigroups will prove to be a consequence of the convergence in Theorem 3.3 using the moment duality from Lemma 3.8 for X and Z and the sampling duality from Proposition 2.8 for X^N and Z^N . First, observe that the sampling duality approximates the moment duality for large N for our choice of μ_N : conditioning on the choice

of distribution for Y_0^N , we can rewrite the sampling duality function in the notation of Proposition 2.8 as

$$\begin{aligned}
H_{\mu_N}(x, n) &= (1 - c_N) \mathbb{E}[\varphi_{Y_0^N}(x)^n] + c_N \mathbb{E}[\varphi_{Y_0^N}((1 - U)x + U(1 - B_x))^n] \\
&= (1 - c_N)(1 - \rho_N \gamma_N([0, 1])) x^n \left(\frac{1 - w_N}{1 - x w_N} \right)^n \\
&\quad + \mathbb{E}[\varphi_{Y_0^N}(x)^n \mid Y_0^N \sim \bar{\gamma}_N] \rho_N \gamma_N([0, 1]) + O(c_N) \\
&= (1 - c_N)(1 - \rho_N \gamma_N([0, 1])) x^n \left(\frac{1 - w_N}{1 - x w_N} \right)^n + O(\rho_N \gamma_N([0, 1])) + O(c_N) \\
&= x^n \left(\frac{1 - w_N}{1 - x w_N} \right)^n + O(\rho_N \gamma_N([0, 1])) + O(c_N)
\end{aligned}$$

where we crudely estimated the expectations and the expression in the first summand by 1. Note that this way the convergence of the remainder terms is uniform in n and x . Then the sampling duality can be written as

$$\begin{aligned}
&\mathbb{E}^n \left[x^{Z^N(-g)} \left(\frac{1 - w_N}{1 - x w_N} \right)^{Z^N(-g)} \right] \\
&= \mathbb{E}_x \left[(X^N(g))^n \left(\frac{1 - w_N}{1 - X^N(g) w_N} \right)^n \right] + O(\rho_N \gamma_N([0, 1])) + O(c_N)
\end{aligned}$$

which we can rearrange to

$$\begin{aligned}
(35) \quad \mathbb{E}^n \left[x^{Z^N(-g)} \right] &= \mathbb{E}_x \left[(X^N(g))^n \left(\frac{1 - w_N}{1 - X^N(g) w_N} \right)^n \right] + O(\rho_N \gamma_N([0, 1])) + O(c_N) \\
&\quad + \mathbb{E}^n \left[x^{Z^N(-g)} \left(1 - \left(\frac{1 - w_N}{1 - x w_N} \right)^{Z^N(-g)} \right) \right].
\end{aligned}$$

Note that the function $h_N : [0, 1] \rightarrow [0, 1]$, $h_N(x) := x^n \left(\frac{1 - w_N}{1 - x w_N} \right)^n$, is continuous and bounded and converges uniformly to $h(x) := x^n$. In addition, the sequence is decreasing in n . Hence, the weak convergence of $X^N(\lfloor \rho_N^{-1} t \rfloor) \rightarrow X(t)$ from Theorem 3.3, implies

$$\mathbb{E}_x \left[(X^N(\lfloor \rho_N^{-1} t \rfloor))^n \left(\frac{1 - w_N}{1 - X^N(\lfloor \rho_N^{-1} t \rfloor) w_N} \right)^n \right] \xrightarrow{N \rightarrow \infty} \mathbb{E}_x[X(t)^n]$$

for any $t \geq 0$ and any $x \in [0, 1]$, uniformly in n , which takes care of the first summand. For the second, consider the functions $\tilde{h}_{N,x} : [0, 1] \rightarrow [0, 1]$, $\tilde{h}_{N,x}(u) := \exp \left(u^{-1} \log \left(x \frac{1 - w_N}{1 - x w_N} \right) \right) \mathbb{1}_{\{0\}}(u)$ and note that by Dini's Theorem this converges to $\tilde{h}_x(u) := \exp(u^{-1} \log(x)) \mathbb{1}_{\{0\}}(u) = x^{(u^{-1})} \mathbb{1}_{\{0\}}(u)$ uniformly over all $u \in [0, 1]$. This allows us to estimate

$$\begin{aligned}
0 &\leq \mathbb{E}^n \left[x^{Z^N(-g)} \left(1 - \left(\frac{1 - w_N}{1 - x w_N} \right)^{Z^N(-g)} \right) \right] \\
&\leq \sup_{k \in \mathbb{N}} x^k \underbrace{\left(1 - \left(\frac{1 - w_N}{1 - x w_N} \right)^k \right)}_{\tilde{h}_x(k^{-1}) - \tilde{h}_{N,x}(k^{-1})}
\end{aligned}$$

$$\leq \sup_{u \in [0,1]} (\tilde{h}_x(u) - \tilde{h}_{N,x}(u)) \xrightarrow{N \rightarrow \infty} 0$$

for any $g \in \mathbb{Z}$ and $x \in [0, 1]$. Note that this convergence is again uniform in n . Therefore, rescaling time and taking the limits in (35) we obtain

$$\lim_{N \rightarrow \infty} \mathbb{E}^n \left[x^{Z^N(-\lfloor \rho_N^{-1} t \rfloor)} \right] = \mathbb{E}_x[X(t)^n]$$

for every $t \geq 0$ and $x \in [0, 1]$, uniformly in $n \in \mathbb{N}_0$. Now to prove convergence of the semi-groups fix $t > 0$ and $\lambda > 0$. Define $x := e^{-\lambda}$ and observe that, using the observations above and the moment duality from Lemma 3.8

$$\begin{aligned} P_t^N f_\lambda(n) &= \mathbb{E}_n[f_\lambda(Z^N(-\lfloor \rho_N^{-1} t \rfloor))] = \mathbb{E}_n[x^{Z^N(-\lfloor \rho_N^{-1} t \rfloor)}] \\ &\xrightarrow{N \rightarrow \infty} \mathbb{E}_x[X(t)^n] = \mathbb{E}_n[x^{Z(t)}] = P_t f_\lambda(n), \end{aligned}$$

uniformly in n , which completes the proof. \square

5.3. Long-term behaviour. Given its simplicity the proof of Theorem 4.2 was given in Section 4.1, hence we only have to give the proofs of Lemma 4.5 (Section 5.3.1) and Lemma 4.6 (Section 5.3.2). The proof of the latter requires the representation of the generator of the Λ -Fleming-Viot process with weak and rare selection given in Lemma 4.7 as well as two technical estimates summarized in Proposition 5.1. The proposition is stated in Section 5.3.2 together with the proof of Lemma 4.6, but the proofs of the proposition and Lemma are in turn postponed to Sections 5.3.3 and 5.3.4.

5.3.1. Proof of Lemma 4.5. The proof of this Lemma makes use of the approximation of the Λ -Fleming-Viot process with the individual-based-model defined in Section 2.1.

PROOF OF LEMMA 4.5. As mentioned, the assumption $\mathbb{P}_x(X_\infty = 0) = 1$ should imply that the term $\mathbb{E}_x[X(t)]$ decays exponentially. The idea of the proof is to obtain a geometric bound for the analogous quantity in the finite population size model. This will be done by estimating the probability that a uniformly chosen individual and an individual of the strong type picked from the same generation have a disjoint ancestry. A crucial observation for this estimate is that the assumption $\mathbb{P}_x(X_\infty = 0) = 1$ implies an infinite genealogy of strong individuals, which will allow us to sample the strong individual uniformly and independently of the first individual.

Since we know that the limit for $t \rightarrow \infty$ of $X(t)$ exists and takes values in $\{0, 1\}$ \mathbb{P}_x -a.s. for any $x \in [0, 1]$, the implication “ \Rightarrow ” is immediate.

The converse “ \Leftarrow ”, however, requires more care. Let us first consider the case where $\sigma + c = 0$, i.e. there is no genetic drift. Since we assumed the process not to be constant, we know $\alpha_s \alpha^* + w > 0$, and by Lemma 4.6.2 $\int_{[0, \infty[} \mathbb{E}_x[X(s)] ds < \infty$ (for all $x \in [0, 1[$), which suffices for “ \Leftarrow ”.

Assume now that $\sigma + c > 0$. The proof is split in three steps and uses that $(X(t))_{t \geq 0}$ is the (appropriately rescaled) weak limit of a finite particle system $(X^N(g))_{g \in \mathbb{N}_0}$, as observed in Theorem 3.5. (Without loss of generality we let the time-index of X^N start in $g_0 = 0$ and hence omit the additional superscript.) Since $(X^N(g))_{g \in \mathbb{N}_0}$ is a Markov chain on a finite state-space with two absorbing states 0 and 1, it will be absorbed in finite time, hence $X^N(\infty) := \lim_{g \rightarrow \infty} X^N(g) \in \{0, 1\}$ exists \mathbb{P}_x -a.s. for any $x \in [N]_0/N$.

Fix $x \in [0, 1[$ rational. All N appearing henceforth are assumed to be such that $x \in [N]_0/N$.

Step 1: We begin by proving that the assumption $\mathbb{P}_x(X_\infty = 0) = 1$ in particular implies

$$(36) \quad \lim_{N \rightarrow \infty} \mathbb{P}_x(X^N(\infty) = 0) = 1.$$

Recall the duality from Proposition 2.8:

$$\mathbb{E}^n[H_{\mu^N}(x, Z^N(-g))] = \mathbb{E}_x[H_{\mu^N}(X^N(g), n)]$$

for any $n \in \mathbb{N}$. The map $\bar{x} \mapsto H_{\mu^N}(\bar{x}, n)$ is continuous and bounded (for any $n \in \mathbb{N}_0$) with $H_{\mu^N}(0, n) = 0$ and $H_{\mu^N}(1, n) = 1$ therefore

$$(37) \quad \lim_{g \rightarrow \infty} \mathbb{E}_x[H_{\mu^N}(X^N(g), n)] = \mathbb{P}_x(X^N(\infty) = 1).$$

Since for fixed $N \in \mathbb{N}$ the state space of $(Z^N(-g))_{g \in \mathbb{N}_0}$ is finite, it has an invariant distribution which we will denote by ν^N (and assume to be defined on \mathbb{N}_0). Let \mathbb{E}^{ν^N} denote the expectation under the law of the process with initial distribution ν^N . Then, using again the duality

$$\begin{aligned} \mathbb{E}^{\nu^N}[H_{\mu^N}(x, Z^N(-g))] &= \sum_{n \in \mathbb{N}_0} \nu^N(n) \mathbb{E}^n[H_{\mu^N}(x, Z^N(-g))] \\ &= \sum_{n \in \mathbb{N}_0} \nu^N(n) \mathbb{E}_x[H_{\mu^N}(X^N(g), n)] \\ &\xrightarrow{g \rightarrow \infty} \mathbb{P}_x(X^N(\infty) = 1), \end{aligned}$$

by (37) as the sum is actually finite and ν^N a distribution. However, the left-hand-side does actually not depend on g , since ν^N is precisely the invariant distribution. Therefore we can substitute the limit by an equality and obtain the representation

$$(38) \quad \sum_{n \in \mathbb{N}_0} \nu^N(n) \mathbb{E}_x[H_{\mu^N}(X^N(g), n)] = \mathbb{P}_x(X^N(\infty) = 1)$$

for any $g \in \mathbb{N}_0$. One can check that for any $N \in \mathbb{N}$, $\bar{x} \in [0, 1]$ and $n \in \mathbb{N}_0$

$$H_{\mu^N}(\bar{x}, n) \leq H_{\mu^N}(\bar{x}, 1)$$

and that $\bar{x} \mapsto H_{\mu^N}(\bar{x}, 1)$ converges to the identity map $\bar{x} \mapsto \bar{x}$ uniformly in $[0, 1]$, as $N \rightarrow \infty$. By the latter, the weak convergence of $X^N(\lfloor \rho_N^{-1} t \rfloor)$ to $X(t)$ implies

$$\mathbb{E}_x[H_{\mu^N}(X^N(\lfloor \rho_N^{-1} t \rfloor), 1)] \xrightarrow{N \rightarrow \infty} \mathbb{E}_x[X(t)]$$

for every $t \geq 0$. Thus, using (38) with $g = \lfloor \rho_N^{-1} t \rfloor$ and the observations above

$$\begin{aligned} 0 &\leq \limsup_{N \rightarrow \infty} \mathbb{P}_x(X^N(\infty) = 1) = \limsup_{N \rightarrow \infty} \sum_{n \in \mathbb{N}_0} \nu^N(n) \mathbb{E}_x[H_{\mu^N}(X^N(\lfloor \rho_N^{-1} t \rfloor), n)] \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_x[H_{\mu^N}(X^N(\lfloor \rho_N^{-1} t \rfloor), 1)] \sum_{n \in \mathbb{N}_0} \nu^N(n) \\ &= \mathbb{E}_x[X(t)] \end{aligned}$$

for any $t \geq 0$, where the last equality follows from the fact that $\sum_{n \in \mathbb{N}_0} \nu^N(n) = 1$. Under the assumption $\mathbb{P}_x(X_\infty = 0) = 1$, however, the right-hand-side converges to 0, as $t \rightarrow \infty$, and since the left-hand-side does not depend on t , we have proven (36).

Step 2: Step 1 now allows us to condition on the event of extinction of the finite-population process (for N sufficiently large). Denote by $\bar{\mathbb{P}}_x^N$ the probability measure obtained from \mathbb{P}_x by conditioning on $\{X^N(\infty) = 0\}$. Let T_0^N be the time to extinction of X^N , i.e.

$$T_0^N := \inf\{g \geq 0 \mid X^N(g) = 0\}.$$

Hence, we are precisely conditioning on T_0^N being finite. By definition of T_0^N , any individual in a generation after T_0^N is of the strong type 1. Recall from Definition 2.5, that we denote by $A_l^N(w)$ the potential ancestors of an individual w alive l generations back in time (counting from $g(w) \in \mathbb{N}_0$, the generation of w). Analogously, define the *strong potential ancestors* of w in the typed process alive l generations back

$$SA_l^N(w) := \{v \in A_l^N(w) \mid v \text{ is of type 1}\}, \quad l = 0, \dots, g(w).$$

Note that $SA_l^N(w) = \emptyset$ if, and only if, w is of the weak type 0.

Fix a generation $g \geq 0$. We want to prove a geometric bound on $\bar{\mathbb{E}}_x^N[X^N(g)]$. First, note that for I uniform on $[N]$ and independent of all other randomness in the model, we can express

$$\bar{\mathbb{E}}_x^N[X^N(g)] = \bar{\mathbb{P}}_x^N((g, I) \text{ is of type 0}).$$

We will use the fact that, for an individual from generation g to be of the weak type 0, its ancestry must be disjoint from the strong ancestry of any strong individual in the same generation g (which we know to exist \mathbb{P}_x^N -a.s.). However, in order to estimate the probability of this event, said individual must be chosen with care.

Let J be another random variable uniform on $[N]$, independent of any other randomness and use it to choose uniformly at random an individual from the first generation not before g in which the strong type has fixated:

$$W := (\max\{g, T_0^N\}, J).$$

Like all individuals in its generation, W is of the strong type 1. $SA_{\max\{g, T_0^N\}-g}^N(W)$ are its strong potential ancestors alive in generation g . Consequently, choose W' uniformly at random (independently of all other randomness) among those $SA_{\max\{g, T_0^N\}-g}^N(W)$. Then W' is by definition a strong individual in generation g . Due to exchangeability of the underlying ancestral process, however, its label $i(W')$ is uniform on $[N]$ and independent of I , the label of (g, I) , the first individual we chose.

As remarked, for (g, I) to be of the weak type 0, its ancestry $A_l^N((g, I))$ must be, in particular, disjoint from the (non-empty) strong ancestry $SA_l^N(W')$ of W' , for every previous generation, allowing us to estimate

$$\bar{\mathbb{P}}_x^N((g, I) \text{ is of type 0}) \leq \bar{\mathbb{P}}_x^N(\forall l = 0, \dots, g: A_l^N((g, I)) \cap SA_l^N(W') = \emptyset).$$

The probability for two given individuals to find a common ancestor of any type in the previous generation can be bounded from above by the probability p_N for this same event with the selection mechanism “turned off”, whereby each individual always samples one and only one parent. A close look at the parent-choosing mechanism described in Section 2.1 reveals that

$$(39) \quad p_N = (1 - c_N)/N + m_1 c_N + m_2 c_N / N$$

for positive, finite constants m_1 and m_2 . Indeed, recall that in this set-up we have a generation with no skewed offspring distribution with probability $1 - c_N$, in which case each individual chooses its parent uniformly, independently of the others and hence the probability of two given individuals choosing the same parent is $1/N$. In a generation with skewed offspring

distribution (occurring with probability c_N) the strength of the correlation, i.e. the probability of choosing the specially marked parent was given by $V \sim \Lambda_c$ which was sampled once for the whole generation. In this case we have three possibilities of choosing the same parent: Either both individuals choose the specially marked one, which happens with probability V^2 . Or one chooses the specially marked and the other chooses uniformly at random, but still ends up with the same individual, which has probability $V(1 - V)/N$. Or both choose uniformly and happen to choose the same, which has probability $(1 - V)^2/N$. Together this means that

$$\begin{aligned} p_N &= (1 - c_N) \frac{1}{N} + c_N \int_{[0,1]} \left\{ v^2 + v(1 - v) \frac{1}{N} + (1 - v)^2 \frac{1}{N} \right\} \Lambda_c(dv) \\ &= (1 - c_N) \frac{1}{N} + c_N \int_{[0,1]} v^2 \Lambda_c(dv) + \frac{c_N}{N} \int_{[0,1]} (1 - v) \Lambda_c(dv). \end{aligned}$$

Since we constructed W' and (g, I) such that they are both uniform on $[N]$ and independent, by exchangeability and independence of the generations we can use this estimate for every generation and obtain

$$\begin{aligned} \bar{\mathbb{E}}_x^N[X^N(g)] &\leq \bar{\mathbb{P}}_x^N(\forall l = 0, \dots, g : A_l^N((g, I)) \cap SA_l^N(W') = \emptyset) \\ &\leq (1 - p_N)^g \left(1 - \frac{1}{N}\right), \end{aligned}$$

where the last factor is the probability of the two individuals (g, I) and W' not being identical.

Step 3: All that remains to be done is to translate this bound into a bound for the scaling limit. Since the identity map is bounded and continuous on $[0, 1]$ Theorem 3.5 implies

$$\mathbb{E}_x[X^N(\lfloor \rho_N^{-1} t \rfloor)] \xrightarrow{N \rightarrow \infty} \mathbb{E}_x[X(t)]$$

for any $t \geq 0$. At the same time, since $\lim_{N \rightarrow \infty} \mathbb{P}_x(X^N(\infty) = 0) = 1$ by step 1, using bounded convergence, we see

$$\begin{aligned} &|\mathbb{E}_x[X(t)] - \bar{\mathbb{E}}_x^N[X^N(\lfloor \rho_N^{-1} t \rfloor)]| \\ &\leq |\mathbb{E}_x[X(t)] - \mathbb{E}_x[X^N(\lfloor \rho_N^{-1} t \rfloor)]| \\ &\quad + \left| \mathbb{E}_x \left[X^N(\lfloor \rho_N^{-1} t \rfloor) \left(1 - \frac{\mathbb{1}_{\{X^N(\infty)=0\}}}{\mathbb{P}_x(X^N(\infty)=0)} \right) \right] \right| \\ &\xrightarrow{N \rightarrow \infty} 0. \end{aligned}$$

Therefore the estimate from step 2 yields

$$\mathbb{E}_x[X(t)] = \lim_{N \rightarrow \infty} \bar{\mathbb{E}}_x^N[X^N(\lfloor \rho_N^{-1} t \rfloor)] \leq \lim_{N \rightarrow \infty} (1 - p_N)^{\lfloor \rho_N^{-1} t \rfloor}$$

for every $t \geq 0$. Under the assumptions of Lemma 3.3, $\lim_{N \rightarrow \infty} p_N \rho_N^{-1} = \sigma + m_1 c > 0$, and therefore

$$\mathbb{E}_x[X(t)] \leq \lim_{N \rightarrow \infty} (1 - p_N)^{\lfloor \rho_N^{-1} t \rfloor} = \exp(-(\sigma + m_1 c)t)$$

which in turn implies $\int_0^\infty \mathbb{E}_x[X(s)] ds < \infty$.

Finally we remove the assumption of rationality on x . Thus, let $x \in [0, 1[$ and assume $\mathbb{P}_x(X_\infty = 0) = 1$. By Remark 4.1, we may therefore choose an $\tilde{x} \in [x, 1[$ rational, such that $\mathbb{P}_{\tilde{x}}(X_\infty = 0) = 1$. For such \tilde{x} we have just proven that $\int_0^\infty \mathbb{E}_{\tilde{x}}[X(s)] ds < \infty$. The claim follows since $\mathbb{E}_x[X(t)] \leq \mathbb{E}_{\tilde{x}}[X(t)]$ for every $t \geq 0$. \square

5.3.2. *Proof of Lemma 4.6 via Lyapunov functions.* For the proof of Lemma 4.6 we use the same Lyapunov approach as in [20] to attain slightly weaker conclusions than [20] in that extra care was needed to avoid the risk of some illegal exchanges of expectation and integration. To this end, let $\kappa \in \mathbb{R}$ and $N \in \mathbb{N}$ and consider the functions $f_{N,\kappa} : [0, 1] \rightarrow \mathbb{R}$ defined as

$$(40) \quad f_{N,\kappa}(x) := \sum_{n=1}^N \frac{1}{n} x^n - \kappa x.$$

Note that for $N \rightarrow \infty$ the sum converges to $-\log(1-x)$, but for each finite N they have convenient integrability properties which will allow us to use these functions for a Lyapunov-type argument to prove both parts of the lemma.

Before we come to the proof we first summarize some convenient observations on these functions and their interplay with the generator of X given in (16) in the following proposition. Its proof is postponed to Section 5.3.3.

Recall from Section 4.1 the definitions of the strength of selection w , $\alpha_{\mathfrak{s}}$ and α^* and the strength of genetic drift β^* and the quantities involved in it: $W := Y^c U$, where $Y^c \sim \Lambda_c$, U has density $2u$ on $[0, 1]$ and they are independent; $\alpha_{\mathfrak{s}} := \Lambda_{\mathfrak{s}}([0, 1])$ and V is uniform on $[0, 1]$ and $Y^{\mathfrak{s}} \sim \alpha_{\mathfrak{s}}^{-1} \Lambda_{\mathfrak{s}}([0, 1])$ and they are independent.

PROPOSITION 5.1. *For $\kappa \in \mathbb{R}$ and $N \in \mathbb{N}$ let $f_{N,\kappa} : [0, 1] \rightarrow \mathbb{R}$ be defined as in (40) above and recall the generator \mathcal{A} of the Λ -Fleming-Viot process with weak and rare selection given in (16). Then the following holds:*

1. *We can define an operator $\tilde{\mathcal{A}}$ such that for all $\kappa \in \mathbb{R}$, $N \in \mathbb{N}$ and $x \in [0, 1]$*

$$\mathcal{A}f_{N,\kappa}(x) = x\tilde{\mathcal{A}}f_{N,\kappa}(x).$$

2. *For this operator we have lower and upper bounds \underline{a} and \bar{a} such that for all $\kappa \in \mathbb{R}$, $N \in \mathbb{N}$ and $x \in [0, 1]$*

$$\underline{a}(N, x, \kappa) \leq \tilde{\mathcal{A}}f_{N,\kappa}(x) \leq \bar{a}(N, x, \kappa)$$

where

$$(41) \quad \begin{aligned} \underline{a}(N, x, \kappa) = & \frac{1}{2}c\mathbb{E}\left[\frac{1}{(1-W)(1-x(1-W))}\right] - w \\ & - \alpha_{\mathfrak{s}}\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}}\right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]} \frac{1}{1+xV\mathbb{E}\left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}}\right]}\right] \\ & + \kappa w(1-x) + \kappa\alpha_{\mathfrak{s}}(1-x)\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}}\right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]}\right] \\ & - \frac{1}{2}c\mathbb{E}\left[\frac{(x(1-W)+W)^N}{W(1-W)}\right] \end{aligned}$$

$$\begin{aligned} \bar{a}(N, x, \kappa) = & \frac{1}{2}c\mathbb{E}\left[\frac{1}{W(1-W)}\right] - w(1-x^N) \\ & + w\kappa(1-x) + \alpha_{\mathfrak{s}}\kappa(1-x)\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}}\right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]}\right] \end{aligned}$$

$$(42) \quad -\alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \right. \\ \left. \times \frac{1 - \left(x - x(1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \right)^N}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right].$$

and thus

$$(43) \quad |\underline{a}(N, x, \kappa)| \leq c\beta^* + w + \alpha_s + w + |\kappa|\alpha_s + c\beta^* < \infty,$$

$$(44) \quad |\bar{a}(N, x, \kappa)| \leq c\beta^* + w + w|\kappa| + \alpha_s|\kappa| + \alpha_s\alpha^* < \infty.$$

3. In particular,

$$(45) \quad |\tilde{\mathcal{A}}f_{N,\kappa}(x)| \leq 2c\beta^* + (2 + |\kappa|)w + (1 + |\kappa| + \alpha^*)\alpha_s < \infty.$$

PROOF OF LEMMA 4.6. Proposition 5.1.3, i.e. equation (45), in particular implies that $|\tilde{\mathcal{A}}f_{N,\kappa}(x)| < \infty$ uniformly in $x \in [0, 1]$. Therefore, for any $N \in \mathbb{N}$ and $\kappa \in \mathbb{R}$

$$\left(f_{N,\kappa}(X(t)) - f_{N,\kappa}(X(0)) - \int_{[0,t]} \mathcal{A}f_{N,\kappa}(X(s)) ds \right)_{t \geq 0}$$

is a martingale and we know

$$(46) \quad \mathbb{E}_x[f_{N,\kappa}(X(t))] - f_{N,\kappa}(x) = \int_{[0,t]} \mathbb{E}_x[\mathcal{A}f_{N,\kappa}(X(s))] ds$$

holds for all $N \in \mathbb{N}$, $\kappa \in \mathbb{R}$, $x \in [0, 1]$ and $t \geq 0$.

Before we turn to the proof of the two statements of the lemma, we draw a first helpful conclusion: Under the assumption $\beta^* < \infty$, (46) implies that for any $x \in [0, 1[$ and any $t \geq 0$ the probability of the weak allele having reached fixation is zero, i.e.

$$(47) \quad \mathbb{P}_x(X(t) = 1) = 0.$$

If this were not the case there would exist $\bar{x} \in [0, 1[$ and $\bar{t} > 0$ such that $\mathbb{P}_{\bar{x}}(X(\bar{t}) = 1) > 0$. For this choice of parameters

$$\mathbb{E}_{\bar{x}}[f_{N,0}(X(\bar{t}))] - f_{N,0}(\bar{x}) \geq \mathbb{P}_{\bar{x}}(X(\bar{t}) = 1) \sum_{n=1}^N \frac{1}{n} - \sum_{n=1}^N \frac{1}{n} \bar{x}^n \xrightarrow{N \rightarrow \infty} \infty.$$

On the other hand, (46) implies

$$\mathbb{E}_{\bar{x}}[f_{N,0}(X(\bar{t}))] - f_{N,0}(\bar{x}) = \int_{[0,\bar{t}]} \mathbb{E}_{\bar{x}}[\mathcal{A}f_{N,\kappa}(X(s))] ds \\ \leq \bar{t}(2c\beta^* + (2 + |\kappa|)w + (1 + |\kappa| + \alpha^*)\alpha_s) < \infty$$

for any $N \in \mathbb{N}$, which is a contradiction and proves the claim.

Indeed, since 1 is an absorbing state the sets $\{X_t = 1\}$ are increasing in $t \geq 0$ and therefore we have actually proven that the fixation time is almost surely infinite, but we will not use this here.

We are now ready to tackle the lemma itself.

Proof of (i): Assume $\alpha_s \alpha^* + w < c\beta^*$. Equivalently, $c\beta^* - w - \alpha_s \alpha^* > 0$. For this part we will work with the lower bound \underline{a} . Note that parts of the first and third summand of \underline{a} converge (e.g. by bounded convergence) to the critical values as x tends to 1:

$$\frac{1}{2} \mathbb{E} \left[\frac{1}{(1-W)(1-x(1-W))} \right] \xrightarrow{x \rightarrow 1} \beta^*$$

and

$$\mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \frac{1}{1+xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \xrightarrow{x \rightarrow 1} \alpha^*$$

whence we may choose an $x_0 \in [0, 1[$ such that for all $x \geq x_0$

$$\begin{aligned} & c \frac{1}{2} \mathbb{E} \left[\frac{1}{(1-W)(1-x(1-W))} \right] - w \\ & - \alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \frac{1}{1+xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \\ & \geq \frac{c\beta^* - w - \alpha_s \alpha^*}{2} =: \delta > 0. \end{aligned}$$

We use κ to control the remainig $x \leq x_0$. Hence, fix $\kappa_0 \geq 0$ sufficiently large such that

$$\begin{aligned} & \kappa_0 w(1-x_0) + \kappa_0 \alpha_s(1-x_0) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x_0^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \right] \\ (48) \quad & \geq w + \alpha_s - \frac{1}{2} \mathbb{E} \left[\frac{1}{1-W} \right] + \delta. \end{aligned}$$

Note that $(1-x_0) \sum_{l=0}^{k-2} x_0^l = 1 - x_0^{k-1}$, thus the left-hand-side is decreasing in x and therefore (48) actually holds for all $x \leq x_0$.

Combined, we obtain for all but the last summand of \underline{a} : For all $x \in [0, 1]$

$$\begin{aligned} & c \frac{1}{2} \mathbb{E} \left[\frac{1}{(1-W)(1-x(1-W))} \right] - w \\ & - \alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \frac{1}{1+xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \\ & + \kappa_0 w(1-x) + \kappa_0 \alpha_s(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \right] \\ & \geq \delta \end{aligned}$$

which in turn implies

$$\underline{a}(N, x, \kappa_0) \geq \delta - c \frac{1}{2} \mathbb{E} \left[\frac{(x(1-W) + W)^N}{W(1-W)} \right]$$

for all $N \in \mathbb{N}$ and all $x \in [0, 1]$.

All these estimates serve to control the right-hand-side of (46). While we were able to do the previous estimates for deterministic x , the last term will be tackled using bounded

convergence after substituting $X(s)$ for x in the expressions above. In order to obtain a contradiction, assume

$$(49) \quad \forall x \in]0, 1[\quad \int_{[0, \infty[} \mathbb{E}_x[X(s)] ds < \infty,$$

which in particular implies that for any $x \in [0, 1[$ $\mathbb{P}_x(X_\infty = 1) = 0$ and therefore

$$(50) \quad X_{\sup} := \sup_{t \in [0, \infty[} X(t) < 1 \quad \mathbb{P}_x\text{-a.s.}$$

Without loss of generality we may assume $(X(t))_{t \geq 0}$ and W to be independent and we can estimate

$$(51) \quad \begin{aligned} \mathbb{E}_x[X(s)\underline{a}(N, X(s), \kappa_0)] &\geq \mathbb{E}_x \left[X(s) \left(\delta - \frac{c(X(s)(1-W) + W)^N}{2W(1-W)} \right) \right] \\ &\geq \mathbb{E}_x \left[X(s) \left(\delta - \frac{c(X_{\sup}(1-W) + W)^N}{2W(1-W)} \right) \right] \end{aligned}$$

for all $N \in \mathbb{N}$. Using bounded convergence and (50) we see that the right-hand-side converges

$$\mathbb{E}_x \left[X(s) \left(\delta - \frac{c(X_{\sup}(1-W) + W)^N}{2W(1-W)} \right) \right] \xrightarrow{N \rightarrow \infty} \delta \mathbb{E}_x[X(s)] > 0.$$

Assumption (49) allows us to use bounded convergence again, to estimate

$$\begin{aligned} \lim_{N \rightarrow \infty} \int_{[0, \infty[} \mathbb{E}_x[X(s)\underline{a}(N, X(s), \kappa_0)] ds \\ \geq \lim_{N \rightarrow \infty} \int_{[0, \infty[} \mathbb{E}_x \left[X(s) \left(\delta - \frac{c(X_{\sup}(1-W) + W)^N}{2W(1-W)} \right) \right] ds \\ = \delta \int_{[0, \infty[} \mathbb{E}_x[X(s)] ds := \tilde{\delta}(x) > 0 \end{aligned}$$

for any $x \in [0, 1[$. Recalling that \underline{a} was designed as a lower bound (and using (47)) this yields that for every $x \in [0, 1[$ there exists a $N_0 = N_0(x)$ such that for every $N \geq N_0$

$$(52) \quad \begin{aligned} \int_{[0, \infty[} \mathbb{E}_x[\mathcal{A}f_{N, \kappa_0}(X(s))] ds &= \int_{[0, \infty[} \mathbb{E}_x[X(s)\tilde{\mathcal{A}}f_{N, \kappa_0}(X(s))] ds \\ &\geq \int_{[0, \infty[} \mathbb{E}_x[X(s)\underline{a}(N, X(s), \kappa_0)] ds \geq \frac{\tilde{\delta}(x)}{2} > 0. \end{aligned}$$

Now choose \bar{x} sufficiently close to 1, such that

$$\log(1 - \bar{x}) + \kappa_0 \bar{x} < -1.$$

Then choose $\bar{N} \geq N_0(\bar{x})$ sufficiently large such that

$$-\sum_{n=1}^{\bar{N}} \frac{1}{n} \bar{x}^n + \kappa_0 \bar{x} < -\frac{1}{2}.$$

Recall that assumption (49) in particular implies $\mathbb{P}_{\bar{x}}(X_\infty = 0) = 1$ and therefore

$$\lim_{t \rightarrow \infty} \mathbb{E}_{\bar{x}}[f_{\bar{N}, \kappa_0}(X(t))] - f_{\bar{N}, \kappa_0}(\bar{x}) = 0 - \sum_{n=1}^{\bar{N}} \frac{1}{n} \bar{x}^n + \kappa_0 \bar{x}.$$

Together with equation (46) and (52) this implies

$$-\frac{1}{2} > \lim_{t \rightarrow \infty} \mathbb{E}_{\bar{x}}[f_{\bar{N}, \kappa_0}(X(t))] - f_{\bar{N}, \kappa_0}(\bar{x}) = \lim_{t \rightarrow \infty} \int_{[0, t]} \mathbb{E}_{\bar{x}}[\mathcal{A}f_{\bar{N}, \kappa_0}(X(s))] ds \geq \frac{\tilde{\delta}(\bar{x})}{2} > 0,$$

which clearly is a contradiction.

Hence, we have proven that assuming $\alpha_s \alpha^* + w < c\beta^*$, we know

$$\exists \bar{x} \in]0, 1[\quad \int_{[0, \infty[} \mathbb{E}_{\bar{x}}[X(s)] ds = \infty.$$

Proof of (ii): Assume $\alpha_s \alpha^* + w > c\beta^*$. Equivalently $c\beta^* - w - \alpha_s \alpha^* < 0$. For this part we will work with the upper bound \bar{a} which can be rearranged to

$$\begin{aligned} \bar{a}(N, x, \kappa) = & c\beta^* - w - \alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right. \\ & \left. \times \frac{1}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \\ & + w\kappa(1-x) + \alpha_s \kappa(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right] \\ & + wx^N + \alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right. \\ & \left. \times \frac{x^N \left(1 - (1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \right)^N}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right]. \end{aligned}$$

We proceed along the lines of the previous proof. Again, note that the third summand tends to the critical values for $x \rightarrow 1$

$$\mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \frac{1}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \xrightarrow{x \rightarrow 1} \alpha^*.$$

Therefore there exists an $x_0 \in]0, 1[$ such that for all $x \geq x_0$

$$\begin{aligned} c\beta^* - w - \alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \frac{1}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \\ \leq \frac{c\beta^* - w - \alpha_s \alpha^*}{2} =: -\delta < 0. \end{aligned}$$

Again, we control the remaining x with the help of κ . Choose $\kappa_0 \leq 0$ with $|\kappa_0|$ sufficiently large such that

$$w\kappa_0(1-x_0) + \alpha_s \kappa_0(1-x_0) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x_0^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right] \leq -c\beta^* - \delta.$$

Since the left-hand-side is increasing in x (recall that $\kappa_0 \leq 0$), the estimate actually holds for all $x \in [0, x_0[$. Thus we now know that for all $x \in [0, 1]$ we can bound the first five summands of \bar{a} by

$$\begin{aligned} c\beta^* - \alpha_s \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \frac{1}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right]} \right] \\ + \alpha_s \kappa_0 (1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right] \leq -\delta < 0. \end{aligned}$$

Note that we also find a simple bound for the last two summands and thereby obtain

$$(53) \quad \bar{a}(N, x, \kappa_0) \leq -\delta + wx^N + \alpha_s x^N$$

for all $x \in [0, 1]$ and $N \in \mathbb{N}$. We will return to this estimate below.

In order to obtain a contradiction, assume

$$(54) \quad \exists x \in]0, 1[\quad \int_{[0, \infty[} \mathbb{E}_x[X(s)] ds = \infty.$$

Fix one such x for the remainder of the proof. By monotone (or bounded) convergence,

$$\begin{aligned} \mathbb{E}_x[X(s)(-\delta + (w + \alpha_s)X(s)^N)] \\ \xrightarrow{N \rightarrow \infty} (-\delta + w + \alpha_s) \underbrace{\mathbb{P}_x(X(s) = 1)}_{\stackrel{(47)}{=} 0} - \delta \underbrace{\mathbb{E}_x[X(s) \mathbb{1}_{\{X(s) < 1\}}]}_{\stackrel{(54)}{> 0}} \\ (55) \quad = -\delta \mathbb{E}_x[X(s)] < 0. \end{aligned}$$

We can now estimate the right-hand-side of (46), where we again use monotone (or bounded) convergence in the last step: For every $t > 0$

$$\begin{aligned} \int_{[0, t]} \mathbb{E}_x[\mathcal{A}f_{N, \kappa_0}(X(s))] ds &\leq \int_{[0, t]} \mathbb{E}_x[X(s) \bar{a}(N, X(s), \kappa_0)] ds \\ &\stackrel{(53)}{\leq} \int_{[0, t]} \mathbb{E}_x[X(s)(-\delta + (w + \alpha_s)X(s)^N)] ds \\ (56) \quad &\xrightarrow{N \rightarrow \infty} -\delta \int_{[0, t]} \mathbb{E}_x[X(s)] ds < 0. \end{aligned}$$

By assumption (54) we may choose $\bar{t} > 0$ such that

$$\int_{[0, \bar{t}]} \mathbb{E}_x[X(s)] ds \geq -4 \frac{\log(1-x) - \kappa_0 x}{\delta}$$

and by (56) an $\bar{N} = \bar{N}(\bar{t})$ such that

$$(57) \quad \int_{[0, \bar{t}]} \mathbb{E}_x[\mathcal{A}f_{\bar{N}, \kappa_0}(X(s))] ds \leq 2(\log(1-x) + \kappa_0 x).$$

Now that we have the right-hand-side under control, we turn to the left-hand-side of (46). Remembering that $\kappa_0 \leq 0$, we crudely estimate

$$\mathbb{E}_x[f_{\bar{N}, \kappa_0}(X(\bar{t}))] - f_{\bar{N}, \kappa_0}(x) = \sum_{n=1}^{\bar{N}} \frac{1}{n} \mathbb{E}_x[X(t)^n] - \kappa_0 \mathbb{E}_x[X(t)] - \sum_{n=1}^{\bar{N}} \frac{1}{x^n} + \kappa_0 x$$

$$\geq -\sum_{n=1}^{\bar{N}} \frac{1}{n} x^n + \kappa_0 x \geq \log(1-x) + \kappa_0 x$$

which is a contradiction to (46) given (57). \square

5.3.3. Proof of Proposition 5.1: Bounds on the generator. It will be convenient to consider the summands of the generator individually. Therefore, we denote by \mathcal{A}_s the part of the generator describing the rare selection mechanism, i.e.

$$(58) \quad \mathcal{A}_s f(x) := \int_{[0,1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s(dy)$$

and the part describing the coordinated random genetic drift as

$$(59) \quad \mathcal{A}_c f(x) := \mathcal{A}f(x) - \sigma \frac{1}{2} x(1-x) f''(x) + wx(1-x) f'(x) - \mathcal{A}_s f(x).$$

Recall that we abbreviate $\alpha_s := \Lambda_s([0, 1])$.

PROOF OF PROPOSITION 5.1. We begin with a few observations on the functions (40) and the generators applied to them. Their first derivative is

$$f'_{N,\kappa}(x) = \sum_{n=1}^N x^{n-1} - \kappa = \begin{cases} \frac{1-x^N}{1-x} - \kappa, & \text{if } x \neq 1, \\ N - \kappa, & \text{if } x = 1. \end{cases}$$

Since the calculations are simple, but tedious, we consider the parts of the generator separately for clarity, using their representations given in Lemma 4.7. Recall that we assumed $\sigma = 0$.

Taking the expectation with respect to V , then in the next step plugging in the definition of the derivative we obtain the following expression:

$$\begin{aligned} \mathcal{A}_c f_{N,\kappa}(x) &= \frac{1}{2} cx(1-x) \mathbb{E} [f''_{N,\kappa}(x(1-W) + VW)] \\ &= \frac{1}{2} cx(1-x) \mathbb{E} \left[\frac{1}{W} (f'_{N,\kappa}(x(1-W) + W) - f'_{N,\kappa}(x(1-W))) \right] \\ &= x \frac{1}{2} c(1-x) \underbrace{\mathbb{E} \left[\frac{1}{W} \left(\frac{1 - (x(1-W) + W)^N}{1 - (x(1-W) + W)} - \frac{1 - (x(1-W))^N}{1 - x(1-W)} \right) \right]}_{=: \tilde{\mathcal{A}}_c f_{N,\kappa}(x)}. \end{aligned}$$

Note that the expression $f'_{N,\kappa}(1)$ can only occur for $x = 1$ and since in this case the right-hand-side is 0 if we assume something in the spirit of $0 \cdot \infty = 0$, we take the liberty to write this expression for all $x \in [0, 1]$.

Following the same reasoning we obtain a similar expression for the coordinated selection term for any $x \in [0, 1]$:

$$\begin{aligned} \mathcal{A}_s f_{N,\kappa}(x) &= -\alpha_s x(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right. \\ &\quad \left. \times f'_{N,\kappa}(x + V \mathbb{E}[x^{K_{Y^s}} - x \mid Y^s]) \right] \end{aligned}$$

$$\begin{aligned}
&= -x\alpha_s(1-x)\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]}\right. \\
&\quad \left.\times \left(\frac{1 - \left(x - x(1-x)V\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right]\right)^N}{1 - \left(x - x(1-x)V\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right]\right)} - \kappa\right)\right].
\end{aligned}$$

Let $\tilde{\mathcal{A}}_s f_{N,\kappa}(x)$ be such that $\mathcal{A}_s f_{N,\kappa}(x) = x\tilde{\mathcal{A}}_s f_{N,\kappa}(x)$. Define

$$\tilde{\mathcal{A}} f_{N,\kappa}(x) := \tilde{\mathcal{A}}_c f_{N,\kappa}(x) - w(1-x)f'_{N,\kappa}(x) + \tilde{\mathcal{A}}_s f_{N,\kappa}(x)$$

which is the claim of 1. in this proposition.

From these expressions, we can deduce bounds for $\tilde{\mathcal{A}} f_{N,\kappa}(x)$ that are monotone in N .

We obtain a lower bound, by omitting certain positive terms in $\tilde{\mathcal{A}} f_{N,\kappa}$ as follows: For $x < 1$, define

$$\begin{aligned}
\underline{a}(N, x, \kappa) &:= \frac{1}{2}c(1-x)\mathbb{E}\left[\frac{1}{W}\left(\frac{1 - (x(1-W) + W)^N}{1 - (x(1-W) + W)} - \frac{1}{1 - x(1-W)}\right)\right] \\
&\quad - w(1-x)\left(\frac{1}{1-x} - \kappa\right) \\
&\quad - \alpha_s(1-x)\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]}\right. \\
&\quad \left.\times \left(\frac{1}{1 - \left(x - x(1-x)V\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right]\right)} - \kappa\right)\right].
\end{aligned}$$

Then, trivially $\tilde{\mathcal{A}} f_{N,\kappa}(x) \geq \underline{a}(N, x, \kappa)$ for any $x \in [0, 1[$ (and $N \in \mathbb{N}$, $\kappa \in \mathbb{R}$).

We rearrange the terms of \underline{a} to obtain a more convenient form. Separating the terms that do/do not depend on N and κ , then canceling several $(1-x)$ -terms in a second step

$$\begin{aligned}
\underline{a}(N, x, \kappa) &= \frac{1}{2}c(1-x)\mathbb{E}\left[\frac{1}{(1-W)(1-x)(1-x(1-W))}\right] \\
&\quad - \frac{1}{2}c(1-x)\mathbb{E}\left[\frac{(x(1-W) + W)^N}{W(1-W)(1-x)}\right] - w \\
&\quad - \alpha_s(1-x)\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]}\right. \\
&\quad \left.\times \frac{1}{(1-x)\left(1 + xV\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right]\right)}\right] \\
&\quad + \kappa w(1-x) + \kappa\alpha_s(1-x)\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]}\right] \\
&= \frac{1}{2}c\mathbb{E}\left[\frac{1}{(1-W)(1-x(1-W))}\right] - w \\
&\quad - \alpha_s\mathbb{E}\left[\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right] \frac{1}{\mathbb{E}[K_{Y^s}-1 \mid Y^s]} \frac{1}{1 + xV\mathbb{E}\left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s\right]}\right]
\end{aligned}$$

$$\begin{aligned}
& + \kappa w(1-x) + \kappa \alpha_{\mathfrak{s}}(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]} \right] \\
& - \frac{1}{2} c \mathbb{E} \left[\frac{(x(1-W) + W)^N}{W(1-W)} \right].
\end{aligned}$$

By extending $\underline{a}(N, 1, \kappa) := -w - \alpha_{\mathfrak{s}} \alpha^*$, we do indeed obtain (41) and

$$(60) \quad \tilde{\mathcal{A}}f_{N,\kappa}(x) \geq \underline{a}(N, x, \kappa)$$

for all $N \in \mathbb{N}$, $\kappa \in \mathbb{R}$ and $x \in [0, 1]$.

Note that we can estimate

$$|\underline{a}(N, x, \kappa)| \leq c\beta^* + w + \alpha_{\mathfrak{s}} + w + |\kappa|\alpha_{\mathfrak{s}} + c\beta^* < \infty$$

and the right-hand-side does not depend on x nor N , which is (43).

In a similar fashion we define an upper bound by omitting certain negative terms in $\tilde{\mathcal{A}}f_{N,\kappa}$: For $x < 1$, define

$$\begin{aligned}
\bar{a}(N, x, \kappa) &:= \frac{1}{2} c(1-x) \mathbb{E} \left[\frac{1}{W} \left(\frac{1}{1 - (x(1-W) + W)} \right) \right] - w(1-x) \left(\frac{1-x^N}{1-x} - \kappa \right) \\
& - \alpha_{\mathfrak{s}}(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]} \right. \\
& \quad \left. \times \left(\frac{1 - \left(x - x(1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \right)^N}{1 - \left(x - x(1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \right)} - \kappa \right) \right] \\
&= \frac{1}{2} c \mathbb{E} \left[\frac{1}{W(1-W)} \right] - w(1-x^N) \\
& + w\kappa(1-x) + \alpha_{\mathfrak{s}}\kappa(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]} \right] \\
& - \alpha_{\mathfrak{s}} \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \frac{1}{\mathbb{E}[K_{Y^{\mathfrak{s}}}-1 \mid Y^{\mathfrak{s}}]} \right. \\
& \quad \left. \times \frac{1 - \left(x - x(1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right] \right)^N}{1 + xV \mathbb{E} \left[\sum_{l=0}^{K_{Y^{\mathfrak{s}}}-2} x^l \mid Y^{\mathfrak{s}} \right]} \right]
\end{aligned}$$

and extend it by $\bar{a}(N, 1, \kappa) := c\beta^*$ to obtain (42). As before, by definition,

$$(61) \quad \tilde{\mathcal{A}}f_{N,\kappa}(x) \leq \bar{a}(N, x, \kappa)$$

for all $N \in \mathbb{N}$, $\kappa \in \mathbb{R}$ and $x \in [0, 1]$. In addition, if we recall that $-x(1-x) \sum_{l=0}^k x^l = x^k - x$, we can simply estimate

$$|\bar{a}(N, x, \kappa)| \leq c\beta^* + w + w|\kappa| + \alpha_{\mathfrak{s}}|\kappa| + \alpha_{\mathfrak{s}}\alpha^* < \infty$$

and the right-hand-side does not depend on x nor N , as claimed in (44).

In particular, combining the above observations, we now know that

$$|\tilde{\mathcal{A}}f_{N,\kappa}(x)| \leq 2c\beta^* + (2 + |\kappa|)w + (1 + |\kappa| + \alpha^*)\alpha_s < \infty$$

uniformly in $N \in \mathbb{N}$ and $x \in [0, 1]$ and thus also (45), i.e. 3. in this proposition holds and the proof is completed. \square

5.3.4. Proof of Lemma 4.7: the Griffiths representation. Recall that we abbreviated the parts of the generator \mathcal{A} corresponding to rare selection and coordinated genetic drift as \mathcal{A}_s and \mathcal{A}_c in (58) and (59) respectively.

PROOF. It is sufficient to find the correct representations for \mathcal{A}_c and \mathcal{A}_s . Theorem 1 in [20] already states

$$\mathcal{A}_c f(x) = \frac{1}{2}cx(1-x)\mathbb{E}[f''(x(1-W) + VW)].$$

On the other hand, in the spirit of the proof of said theorem, we can calculate

$$\begin{aligned} \mathcal{A}_s f(x) &= \int_{[0,1]} (f(\mathbb{E}[x^{K_y}]) - f(x)) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s(dy) \\ &= \int_{[0,1]} (f(x + \mathbb{E}[x^{K_y} - x]) - f(x)) \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s(dy) \\ &= \int_{[0,1]} \int_{[0,1]} \mathbb{E}[x^{K_y} - x] f'(x + u\mathbb{E}[x^{K_y} - x]) du \frac{1}{\mathbb{E}[K_y - 1]} \Lambda_s(dy) \\ &= -\alpha_s \int_{[0,1]} \int_{[0,1]} x(1-x) \mathbb{E} \left[\sum_{l=0}^{K_y-2} x^l \right] \frac{1}{\mathbb{E}[K_y - 1]} \\ &\quad \times f'(x + u\mathbb{E}[x^{K_y} - x]) du \frac{\Lambda_s(dy)}{\Lambda_s([0,1])} \\ &= -\alpha_s x(1-x) \mathbb{E} \left[\mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \frac{1}{\mathbb{E}[K_{Y^s} - 1 \mid Y^s]} \right. \\ &\quad \left. \times f' \left(x - x(1-x)V \mathbb{E} \left[\sum_{l=0}^{K_{Y^s}-2} x^l \mid Y^s \right] \right) \right]. \end{aligned}$$

We used once again the simple observation that $x^k - x = -x(1-x) \sum_{l=0}^{k-2} x^l$ for any $k \in \mathbb{N}$ and $x \in [0, 1]$ if we interpret the empty sum as zero. \square

Acknowledgements. We thank Martin Slowik for pointing out a mistake in an earlier version of this manuscript, which we corrected with Lemma 4.5. We are also grateful to an anonymous referee for suggesting to comment on hypergeometric duality function, as an alternative to our sampling duality. ACG was supported by grants UNAM PAPIIT IA100419 and CONACYT CIENCIA BASICA A1-S-14615. MWB was supported by the SPP 1590, projects BL 1105/5-1 2 and KU 2886/1-1, and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689) and acknowledges support from DFG CRC/TRR 388 “Rough Analysis, Stochastic Dynamics and Related Fields”, Project A09.

REFERENCES

- [1] Afanasyev, V. I., Geiger, J., Kersting, G., & Vatutin, V. A. (2005). Criticality for branching processes in random environment. *Ann. Probab.*, 33(2), 645–673.
- [2] Alkemper, R. & Hutzenthaler, M., (2007). Graphical representation of some duality relations in stochastic population models. *Electron. Commun. Probab.*, Volume 12, paper no. 21, 206–220.
- [3] Athreya, K. B., & Karlin, S. (1971). On Branching Processes with Random Environments: I: Extinction Probabilities. *Ann. Math. Statist.*, 42(5), 1499–1520.
- [4] Bansaye, V., Caballero, M. E., & Méléard, S. (2019). Scaling limits of population and evolution processes in random environment. *Electron. J. Probab.*, 24, paper no. 19, 1–38.
- [5] Bansaye, V., Pardo, J. C., & Smadi, C. (2013). On the extinction of Continuous State Branching Processes with catastrophes. *Electron. J. Probab.*, 18, paper no. 106, 1–31.
- [6] Barton, N., Etheridge, A., & Veber, A. (2010). A New Model for Evolution in a Spatial Continuum. *Electron. J. Probab.*, 15, paper no. 7, 162–216.
- [7] Bertoin, J., & Le Gall, J. F. (2003). Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields*, 126, 261–288.
- [8] Biswas, N., Etheridge, A., & Klimek, A. (2021). The spatial Lambda-Fleming-Viot process with fluctuating selection. *Electron. J. Probab.* 26:1–51.
- [9] Blath, J., González Casanova, A., Kurt, N., & Wilke-Berenguer, M. (2016). A new coalescent for seed-bank models. *Ann. Appl. Probab.*, 26(2), 857–891.
- [10] Boinghoff, C., & Hutzenthaler, M. (2012). Branching diffusions in random environment. *Markov Proc. Rel. Fields*, 18(2), 269–310.
- [11] Donihue, C. M., Herrel, A., Fabre, A. C., Kamath, A., Geneva, A., Schoener, T., Kolbe, J., & Losos, J. (2018). Hurricane-induced selection on the morphology of an island lizard. *Nature*, 560(7716), 88–91.
- [12] Etheridge, A., & Griffiths, R. (2019). A coalescent dual process in a Moran model with genic selection. *Theor. Popul. Biol.*, 675(4), 320–330.
- [13] Foucart, C. (2013). The impact of selection in the Λ -Wright-Fisher model. *Electron. Commun. Probab.*, 18, paper no. 72, 1–10.
- [14] Geiger, J., & Kersting, G. (2000). The survival probability of a critical branching process in random environment. *Teor. Veroyatnost. i Primenen.*, 45(3), 607–615.
- [15] Geiger, J., Kersting, G., & Vatutin, V. A. (2003). Limit theorems for subcritical branching processes in random environment. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(4), 593–620.
- [16] Giardinà, C., Kurchan, J., Redig, F., & Vafayi, K. (2009). Duality and Hidden Symmetries in Interacting Particle Systems. *J. Stat. Phys.*, 135, 25–55.
- [17] Gillespie, J.H. (1973). Natural selection with varying selection coefficients - a haploid model. *Genet. Res., Camb.*, 21, 115–120.
- [18] González Casanova, A., Pardo, J. P., & Perez, J. L. (2021). Branching processes with interactions: the subcritical cooperative regime. *Adv. in Appl. Probab.*, 53(1):251–278.
- [19] González Casanova, A., & Spanò, D. (2018). Duality and fixation in Ξ -Wright-Fisher processes with frequency-dependent selection. *Ann. Appl. Probab.*, 28(1), 250–284.
- [20] Griffiths, R. C. (2014). The Λ -Fleming-Viot process and a connection with Wright-Fisher diffusion. *Adv. in Appl. Probab.*, 46(4), 1009–1035.
- [21] He, H., Li, Z., & Xu, W. (2018). Continuous-state branching processes in Lévy random environments. *J. Theor. Probab.*, 31(4), 1952–1974.
- [22] Hutzenthaler, M., Pfaffelhuber, P., & Printz, C. (2015). Stochastic averaging for multiscale Markov processes with an application to a Wright-Fisher model with fluctuating selection. *arXiv:1504.01508v2*.
- [23] Jansen, S., & Kurt, N. (2014). On the notion(s) of duality for Markov processes. *Probab. Surveys*, 11, 59–120.
- [24] Kallenberg, O. (2002). *Foundations of Modern Probability* (2nd ed.). Springer-Verlag.
- [25] Karlin, S., & Levikson, B. (1974). Temporal fluctuations in selection intensities case of small population size. *Theor. Pop. Biol.*, 6(3), 383–412.
- [26] Karlin, S., & Lieberman, U. (1974). Random temporal variation in selection intensities: the case of large population size. *Theor. Popul. Biol.*, 6(3), 355–382.
- [27] Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- [28] Kingman, J.F.C. (1982). The coalescent. *Stochastic Process. Appl.*, 13(3), 235–248.
- [29] Krone, S. M., & Neuhauser, C. (1997). Ancestral processes with selection. *Theor. Pop. Biol.*, 51(3), 210–237.
- [30] Li, Z., & Pu, F. (2012). Strong solutions of jump-type stochastic equations. *Electron. Commun. Probab.*, 17, paper no. 33, 1–13.
- [31] Möhle, M. (1999). The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli*, 5(5), 761–777.

- [32] Möhle, M., & Sagitov, S. (2001). A Classification of Coalescent Processes for Haploid Exchangeable Population Models. *Ann. Probab.*, 29(4), 1547-1562.
- [33] Papaspiliopoulos, O., Ruggiero, M., & Spanò, D. (2016). Conjugacy properties of time-evolving Dirichlet and gamma random measures. *Electron. J. Statist.*, 10, paper no. 2, 3452-3489.
- [34] Peña-Miller, R., Fuentes-Hernandez, A., Reding, C., Gudelj, I., & Beardmore, R. (2014). Testing the optimality properties of a dual antibiotic treatment in a two-locus, two-allele model. *J. R. Soc. Interface*, 11(96), 20131035.
- [35] Pitman, J. (1999). Coalescents With Multiple Collisions. *Ann. Probab.*, 27(4), 1870-1902.
- [36] Schweinsberg, J. (2000). Coalescents with Simultaneous Multiple Collisions. *Electron. J. Probab.*, 5, paper no. 12, 50 pp.
- [37] Smith, W. L., & Wilkinson, W. E. (1969). On Branching Processes in Random Environments. *Ann. Math. Statist.*, 40(3), 814-827.