Review

Viability Selection at Linked Sites

Bjarki Eldon [1,†,‡] (iD)

[1] Leibniz Institute for Evolution and Biodiversity Science
Museum für Naturkunde Berlin; bjarki.eldon@mfn.berlin
† Current address: Museum für Naturkunde
Invalidenstraße 43; 10115 Berlin, Germany

Abstract: Evolutionary ecology may be described as explaining ecology through evolution and vice versa, but one may also view it as an integration of the two fields, where one takes the view that ecology and evolution are inseparable, and one can only begin to understand the biology of organisms by synthesizing the two fields. An example of such a synthesis is the biology of high fecundity and the associated concept of sweepstakes reproduction, or skewed individual recruitment success. As an illustration we consider selection at linked sites under various dominance and epistasis mechanisms in a diploid population evolving according to random sweepstakes and experiencing recurrent bottlenecks. Using simulations we give a few examples of the impact of the stated elements on selection. We show that depending on the dominance mechanisms random sweepstakes can shorten the time to fixation (conditional on fixation) of the fit type at all sites. Bottlenecks tend to increase the fixation time, with random sweepstakes counteracting the effects of bottlenecks on the fixation time. Understanding the effect of random sweepstakes, recurrent bottlenecks, dominance mechanisms and epistasis on the fate of selectively advantageous mutations may help with explaining genetic diversity in natural highly fecund populations possibly evolving under sweepstakes reproduction.

Keywords: high fecundity; random sweepstakes; natural selection; fixation; bottleneck; evolution; ecology; recruitment dynamics; epistasis

1. Introduction

Central to understanding the evolution and the ecology, and, indeed, the evolutionary ecology of natural populations is recruitment dynamics, i.e. the distribution of individual recruitment success or, in other words, the offspring number distribution. The number of offspring contributed by each individual affects population size and geographic distribution, genetic diversity, and resilience of natural populations[1]. The Wright-Fisher population model of genetic reproduction has been almost universally applied in modeling evolutionary history of natural populations. However, it may be a poor choice for highly fecund populations.

Highly fecund populations provide an important contrast to low fecundity and model organisms, have enormous reproductive potential, are diverse and found widely[1], and may share a key characteristic that is not captured by the Wright-Fisher model, namely sweepstakes reproduction[2,3]. The population model we apply allows us to define high fecundity without involving sweepstakes reproduction. There are two main forms of sweepstakes reproduction involving different mechanisms turning high fecundity into a skewed individual recruitment success. In one mechanism, referred to as random sweepstakes, Type III survivorship and chance matching of reproduction with favorable environmental conditions, for example in broadcast spawners, combine to form a jackpot of surviving offspring won by a few parents. The other form of sweepstakes reproduction has been referred to as selective sweepstakes[4]. The idea is that potential offspring are viewed as passing through selective filters on their way to maturity, and the surviving offspring have on average a different genetic constitution than the non-surviving ones; the next

time around another set of types confers advantage[5]. In our formulation a population may be highly fecund without evolving according to random or selective sweepstakes. Understanding recruitment dynamics in highly fecund populations necessarily centers on identifying sweepstakes reproduction in data, and the mechanisms of sweepstakes reproduction given evidence of it.

Random sweepstakes have been precisely formulated for both haploid and diploid populations[6–12]. As far as we know selective sweepstakes incorporating high fecundity and random sweepstakes have not been explicitly modeled. An extensive analysis of whole-genome sequences provides evidence for the evolution of the highly fecund diploid Atlantic cod being driven by positive selection[4]. A population model of recurrent sweeps of a new mutation each time, the Durrett-Schweinsberg model[13], essentially explains the Atlantic cod data, while models of random sweepstakes derived for diploid populations[6,14,15] do not, even accounting for complex demography and background selection[4]. In the Durrett-Schweinsberg model[13], the population evolves according to the Moran model of genetic reproduction, where at any given time one individual is randomly picked to produce one offspring and another individual perishes to keep the population size constant, and so by itself does not model high fecundity. To properly take high fecundity into account when modeling selective sweepstakes, where high fecundity is defined as high reproductive capacity in the absence of selection, one might want for example to model recurrent sweeps in a background of random sweepstakes. The main conclusion from the analysis of the Atlantic cod data is that strong pervasive positive selection should be considered in any further analysis of the data. Identifying sweepstakes reproduction in genomic data, and resolving the mechanisms of sweepstakes reproduction given evidence of it is one of the main open problems in biology.

To better understand if and how random and selective sweepstakes might interact to shape the evolution of natural populations it is necessary to illuminate how random sweepstakes affect the fate of advantageous mutations. Here we use simulations to consider the fate of advantageous types for a simple model of viability selection acting at linked sites in a diploid population evolving according to random sweepstakes and recurrent bottlenecks. With linked sites we can impose various forms of dominance mechanisms and epistatic interactions between the sites. Similar investigations based on simulations have focused on unlinked sites[16] and a single site[17]. The results indicate that random sweepstakes reduce both the probability of fixation of new advantageous types, and the average time to fixation conditional on fixation. We claim that the models of random sweepstakes applied in [16,17], and here as well, have better properties than the original one[9] leading to the multiple-merger Beta$(2 - \alpha, \alpha)$-coalescent. Multiple-merger coalescents [8,18–21], in which a random number of lineages merges each time, arise in different contexts, for example through recurrent strong bottlenecks[22] and recurrent selective sweeps[13], and may be essential for explaining genetic diversity in natural populations across domains of life[23].

Population models of random sweepstakes and recurrent bottlenecks are in the domain of attraction of jump diffusions and multiple-merger coalescents[22,24]. For an overview of the mathematical formulation of random sweepstakes, and the connection to jump diffusions consult e.g. [24]. We only draw attention to the SDE

$$dY_t = \sqrt{Y_{t-}(1 - Y_{t-})}dW_t + \int_{(0,t] \times (0,1] \times [0,1]} \left(\mathbb{1}_{\{u \leq Y_{t-}\}} r(1 - Y_{t-}) - \mathbb{1}_{\{u > Y_{t-}\}} rY_{t-} \right) N(dsdrdu)$$

(1)

describing the evolution of a type frequency in a haploid population evolving according to random sweepstakes [24]. In Eq (1), $\{Y_t\}$ is the solution of the SDE in Eq (1) and tracks the frequency of a given type in a haploid population evolving according to random sweepstakes, $\{W_t\}$ is a standard Brownian motion, N is a Poisson process on $[0, \infty) \times (0, 1] \times [0, 1]$ independent of $\{W_t\}$ equipped with intensity measure $dt \otimes r^{-2}F(dr) \otimes du$, and F is a finite measure without an atom at zero and corresponds to the random sweepstakes mechanism (the offspring number distribution)[24]. Comparing the probability of fixation of a neutral type evolving according to Eq (1) to one evolving according to the Wright-

Fisher diffusion (i.e. $dY_t = \sqrt{Y_{t-}(1 - Y_{t-})}dB_t$), whose fixation probability is proportional to the initial frequency of the type, would be informative about how the jump term in Eq (1) corresponding to random sweepstakes affect the evolution of the population. In [25] generalized Wright-Fisher models for haploid populations are studied, where

$$\mathbb{E}[Y_{k+1}|Y_k] = Y_k$$
$$\mathrm{Var}[Y_{k+1}|Y_k] = \frac{N}{N-1}\mathrm{Var}[Y_{k+1}|\{Y_k = 1\}]Y_k\left(1 - \frac{Y_k}{N}\right) \tag{2}$$

with $\{Y_k\} \equiv \{Y_k; k \in \mathbb{N}\}$ tracking the number of copies of a given type in a haploid population of fixed size N. The idea is to study a process that matches the Wright-Fisher model only in the first two moments of $\{Y_k\}$ and without imposing any specific constraints on the higher moments of $\{Y_k\}$[25]. Upon adding selection to the generalized Wright-Fisher model, conditions can be identified where a beneficial type fixes with probability one as $N \to \infty$[25]. Furthermore, [25] identify a continuum limit operator, where

$$Gu(x) = \gamma x(1-x)u'(x) + \sigma^2 x(1-x)\int_0^1 \frac{u(y) - u(x) - u'(x)(y - x)}{(y - x)^2}d\Omega_x(y) \tag{3}$$

(assuming absence of mutation), where Ω_x is a probability measure. Given that a generator as in Eq (3) can be identified, it can be applied, at least in theory, to obtain expressions for e.g. fixation probabilities [25,26]. In our case of a diploid population evolving under random sweepstakes, with random pairing of diploid individuals, identifying a continuum limit operator as in Eq (3) seems far from trivial. Mathematical investigations of the effect of random sweepstakes on selection are restricted to a single site in a haploid population[27–30]. Considering a model of selection in a haploid population evolving according to random sweepstakes, where the reproduction mechanism is a mixture of "small" and "large" reproduction events [29] study the SDE

$$X_t = x - \alpha \int_0^t X_s(1 - X_s)ds + \int_0^t \sqrt{F(\{0\})X_s(1 - X_s)}dW_s$$
$$+ \int_{[0,t]\times]0,1[^2} p\left(\mathbb{1}_{\{u \leq X_{s-}\}} - X_{s-}\right)\overline{M}(dsdudp) \tag{4}$$

where $\{W_t\}$ is standard Brownian, and $\overline{M}(ds, du, dp) = M(ds, du, dp) - p^{-2}dsdu\Lambda(dp)$ with M taken as a Poisson point measure on $\mathbb{R}_+ \times]0,1[\times]0,1]$ with intensity $dsdup^{-2}\Lambda(dp)$. In [29] Eq (4) is referred to as the Λ-Wright-Fisher SDE with selection. For comparison, the SDE

$$dX = \sigma X_{t-}(1 - X_{t-})dt + \sqrt{X_{t-}(1 - X_{t-})}dW_t$$

(simplified to a single population) with σ denoting the strength of selection is the starting point for investigating the time to fixation (conditional on fixation) in a structured population[31]. Equations (1), (3), and (4) have in common to include a term due to random sweepstakes, i.e. they are mixtures of a process corresponding to small families (the Wright-Fisher diffusion), and a term corresponding to large families. The model we employ in the simulations (see § 2) is also composed of such processes.

Our aim is to investigate the joint fixation of a fit type jointly at two or more linked sites in a diploid population evolving according to random sweepstakes and experiencing recurrent bottlenecks. Even though the algorithm developed for the simulations works for any number of sites, the number of possible phased L-site types increases exponentially with L, and thus there are computational restrictions (e.g. amount of computer memory) to how many sites can be included. This may not be much of a restriction, however, since adaptation may mostly be due to few new mutations each of large effect[32,33]; this has been referred to as the oligogenic model[34].

In § 2 we describe the model we apply in the simulations, in § 3 we present and discuss the results, in § 4 we briefly summarize the results and remaining open problems, in § A

we present further simulation results, and in § B we provide an overview of the simulation algorithm in the form of a pseudocode.

2. Materials and Methods

We consider a diploid population evolving in discrete (non-overlapping) generations. In any given generation the current individuals randomly form pairs, the pairs independently produce potential offspring (juveniles), and from the pool of juveniles we sample a specified number who will survive and produce a new set of reproducing individuals. Thus, we are essentially modeling a supercritical multi-type branching process, who is at times conditioned to be at a given value. Incorporating potential offspring into our model allows us to formulate the concepts of high fecundity and sweepstakes reproduction. We define high fecundity as the ability of organisms to produce numbers of juveniles at least on the order of the population size. In our formulation high fecundity is a necessary but not sufficient ingredient for sweepstakes reproduction to occur; what is missing is a mechanism turning high fecundity into sweepstakes reproduction. One such mechanism is random sweepstakes, whereby Type III survivorship and chance matching of reproduction with favorable environmental conditions, for example among marine broadcast spawners, combine to generate a jackpot so that a few individuals produce the bulk of the surviving offspring. This mechanism has previously been referred to as sweepstakes reproduction[2,3]. Natural selection is not involved in random sweepstakes, but is a key player in another conceptual mechanism turning high fecundity into sweepstakes reproduction, selective sweepstakes[4,5].

Now we formulate random sweepstakes. Our formulation is based on the haploid population model of random sweepstakes in the domain of attraction of the Beta$(2 - \alpha, \alpha)$-coalescent for $1 \leq \alpha < 2$[9]. Let $X, X_1, \ldots, X_M$ for $M \in \mathbb{N} := \{1, 2, \ldots\}$ denote i.i.d. copies of positive random variables, where $X_1, \ldots, X_M$ denote the independent and identically distributed random number of juveniles produced by M parent pairs (randomly formed) in a given generation. Note that our formulation of randomly forming pairs of parents is in contrast to some models involving "diploid" populations. Sometimes the only notion of diploidy is to assume an even number of individuals (population size $2N$), and the actual mechanism in the model is more often than not strictly haploid (e.g. [13]). Take

$$\mathbb{P}(X = k) = C\left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}\right), \quad k \in \{2, 3, \ldots, u(N)\},$$
$$\mathbb{P}(X \in \{0,1\}) = 1 - \sum_{2 \leq k \leq u(N)} \mathbb{P}(X = k), \tag{5}$$

where α and C are positive constants. For the simulations we take $\mathbb{P}(X \in \{0,1\}) = 0$ without loss of generality and choose α and C so that

$$\mathbb{P}(X = k) \geq \mathbb{P}(X = k+1), \quad k \in \{2, \ldots, u(N)\},$$
$$\mathbb{P}(X \leq u(N)) = 1,$$
$$\mathbb{P}(X_1 + \cdots + X_N \geq 2N) = 1,$$

and $u(N)$ is an upper bound on the number of juveniles. The limiting coalescent (i.e. a coalescent is a probabilistic description of the random ancestral relations of sampled gene copies from a hypothetical population evolving according to a given population model, e.g. Eq (5)) is being investigated for a planned future publication. One can argue that the assumption $u(N)/N \to K$ is biologically realistic. The juveniles represent fertilized eggs, and it is plausible, in particular for marine broadcast spawners, that the number of fertilized eggs produced even over the lifetime of a given individual is at most a fraction of the population size. The cutoff $u(N)$ also provides a means of formally defining a highly fecund population. A population evolving according to Eq (5) is taken to be highly fecund if $u(N)/N \not\to 0$ as $N \to \infty$. The model in Eq (5) is a variant of the model in [9], with $C > 0$

a normalizing constant, the law for the random number of juveniles (X) produced by a given individual behaves as

$$\lim_{x \to \infty} C x^{\alpha} \mathbb{P}(X \geq x) = 1 \qquad (6)$$

The random gene genealogy of a sample of gene copies from a haploid population evolving according to Eq (6) is described by a discrete-time Ξ-coalescent (admitting simultaneous multiple mergers of ancestral lineages) in the case $0 < \alpha < 1$; by the Beta$(2 - \alpha, \alpha)$-coalescent when $1 \leq \alpha < 2$, and if $\alpha \geq 2$ one obtains the Kingman-coalescent[9]. Thus, one may describe a population evolving according to Eq (5) as evolving according to random sweepstakes if $0 < \alpha < 2$. In addition to being a natural model of random sweepstakes, the model in Eq (5) also serves as a natural, biologically realistic, and a mathematically tractable alternative to the Wright-Fisher model. Given our formulation a highly fecund population may or may not evolve according to random sweepstakes.

For $1 < \alpha < 2$ the ancestral process describing the random ancestral relations of sampled gene copies from a haploid population evolving according to Eq (6) converges (in the sense of convergence of finite-dimensional distributions) to a multiple-merger coalescent with time measured in units of $\mathcal{O}(N^{\alpha-1})$ generations[9]. However, recovering observed number of mutations in a sample becomes problematic for estimates of α close to one[17]. One way to get around this difficulty may be to assume a simple mixture distribution on α in Eq (5). Fix $0 < \alpha_1 < 2$ and $\alpha_2 \geq 2$. Suppose with probability ε we take $\alpha = \alpha_1$, and with probability $1 - \varepsilon$ we take $\alpha = \alpha_2$. In this way we model random sweepstakes in a broadcast spawner, where most of the time (i.e. with probability $1 - \varepsilon$) individuals produce a small number of juveniles (through $\alpha = \alpha_2$), but occasionally reproduction matches favorable environmental conditions so that individuals have an increased chance of producing a large number of juveniles (through $\alpha = \alpha_1$). Thus, we can write, with $L(\alpha, u(N))$ denoting the law in Eq (5) for given α and $u(N)$,

$$X_1, \ldots, X_N \sim \begin{cases} L(\alpha_2, u(N)) & \text{with probability } 1 - \varepsilon \\ L(\alpha_1, u(N)) & \text{with probability } \varepsilon \end{cases} \qquad (7)$$

Similar mixture models of random sweepstakes as described in Eq (7) have been shown to yield non-trivial limits with less problematic scaling of time [10,35]. We claim that a scaling of ε can be obtained so that the ancestral process converges to a non-trivial limit (the details are being worked out for a planned future publication). A similar mechanism based on the Moran model, where a single parent produces a fixed number ($\lfloor \psi N \rfloor$ for some fixed $0 < \psi < 1$) of surviving offspring with probability ε, and one offspring with probability $1 - \varepsilon$, leads to a mixture of a Kingman-coalescent and a multiple-merger coalescent[10] (see also[35] for a more thorough investigation of the resulting coalescents derived from a Moran model extended in this way). Each juvenile inherits gene copies according to Mendel's laws, we exclude mutation.

If the total number of juveniles S_M produced in any given generation exceeds a fixed carrying capacity $\mathfrak{C}$ viability selection kicks in. We assume the environment can sustain at most $\mathfrak{C}$ diploid individuals. Given the phased genome of a juvenile we compute the weight as in Eq (8) for some given weight function f. For each given juvenile a random exponential is then sampled with rate the viability weight of the juvenile. The juveniles with the $\mathfrak{C}$ smallest realized exponentials (in case the total number of juveniles exceeds $\mathfrak{C}$) survive to form the next generation of reproducing individuals. This is a way of using viability selection to sample the juveniles. The probability of surviving generally increases with the number of copies of the fit type one carries. The exact configuration of the types in any given juvenile may also play a role. When $S_M \leq \mathfrak{C}$ all the juveniles survive. Viability selection only kicks in when the number of juveniles exceeds the carrying capacity. Under recurrent bottlenecks, the population may evolve neutrally for periods of time when recovering from a bottleneck.

In any given generation a bottleneck occurs with a fixed probability b. When a bottleneck occurs, a fixed number B of individuals are sampled uniformly and without

replacement to survive the bottleneck and produce juveniles. Clearly it would be more realistic to sample a random number of juveniles to survive a bottleneck; however we elect to keep b and B constant. The population size N_{t+1} at time $t+1$ is then given by

$$N_{t+1} = \min\left(\mathfrak{C}, S_{\lfloor M/2 \rfloor}\right)$$

where M is the number of individuals in generation t available for (randomly) forming $\lfloor M/2 \rfloor$ parent pairs producing juveniles,

$$M = B \mathbb{1}_{\{\text{bottleneck occurs}\}} + N_t \mathbb{1}_{\{\text{bottleneck does not occur}\}}$$

In the total absence of bottlenecks ($b = 0$) viability selection acts in every generation (recall Eq (5)). In the presence of bottlenecks ($b > 0$), the population evolves neutrally after a bottleneck occurred until the number of juveniles exceeds the carrying capacity. In § B we describe the algorithm for the simulations.

Throughout let L denote the number of linked sites. At each site there are two types, a wild type, or the type already present in the population, and a fit type arising by mutation and conferring advantage. Let $\mathcal{G}$ denote the set of all possible phased L-site types. We define a viability weight for a juvenile with genome $G \in \mathcal{G}$ as the map

$$G \mapsto 1 + sf(G) \tag{8}$$

where $s > 0$ is the strength of selection, and f is a weight function $f : \mathcal{G} \to [0, \infty)$. To give an example, write a genome G as $(g_1, \ldots, g_L) \in \{0, 1, 2\}^L$ for the genotypes at the L sites, where zero stands for homozygous for the wild type $(0/0)$, one for the heterozygous state $(0/1)$, and two for homozygous for the fit type $(1/1)$. For example, writing $\mathbb{1}_{\{A\}} := 1$ if A holds, and zero otherwise, we can take the weight function f from Eq (8) as

$$(g_1, \ldots, g_L) \mapsto \sum_{\ell=1}^{L} \mathbb{1}_{\{g_\ell = 2\}} \tag{9}$$

where we add one to the value of the weight function if the fit type is found in homozygous state at a site for each site; in other words the fit type is recessive at all the sites (see Figure 1). In Eq (9) we exclude epistasis. A juvenile with a genotype of all wild types (i.e. homozygous for the wild type at all sites) will (in general) be assigned a weight of one, thus forming a baseline fitness value.

Considering two sites, and assuming at which site a fit type occurs is irrelevant, we have the following seven configurations; with (a, b) denoting a haplotype with type a at site one and type b at site two we have the configurations $\{(0,0), (0,0)\}$, $\{(0,0), (0,1)\}$, $\{(0,1), (0,1)\}$, $\{(1,0), (0,1)\}$, $\{(0,0), (1,1)\}$, $\{(0,1), (1,1)\}$, and $\{(1,1), (1,1)\}$. In this notation configuration $\{(a,b), (c,d)\}$ means haplotypes (a,b) and (c,d) and genotypes a/c at site one and b/d at site two.

Fully taking into account the phase between two sites yields sixteen possible genomic configurations (2^{2L} if L sites), where the possible haplotypes are $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$. The possible configurations are shown in Table 1, where the entries are examples of the values of the weight function $(h_1, h_2) \mapsto [0, \infty)$, with h_1, h_2 denoting the two haplotypes.

To put our model in context, the effect of sites on an additive trait with trait value z has been formulated as

$$z = \sum_{\ell=1}^{L} \gamma_\ell \phi(g_\ell) \tag{10}$$

where g_ℓ is the genotype at site ℓ, ϕ is a map for translating the given genotype into a contribution to the trait value, and γ_ℓ is the effect of site ℓ on the trait value[36]. The

h_1/h_2	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	0	$1/2$	$1/2$	$1/2$
$(0,1)$	$1/2$	1	0	1
$(1,0)$	$1/2$	0	1	1
$(1,1)$	$1/2$	1	1	2

Table 1. The possible genomic configurations with two sites; (a,b) for $a,b \in \{0,1\}$ denotes a haplotype with type a at site one and type b at site two; the entries are example values of the weight function in Eq (8).

formulation in Eq (10) corresponds to our weight function f in Eq (8). Quite often a so-called Gaussian fitness is then assumed with the fitness computed according to

$$z \mapsto \exp\left(-\frac{1}{2}s(z - z_o)^2\right) \qquad (11)$$

where z_o denotes the optimal trait value and s the strength of selection[36]. Then the formulation in Eq (11) corresponds to our map in Eq (8).

3. Results

Let $Y_\ell(t) \in [0,1]$ denote the frequency of the fit type at site ℓ at time t, and write, for any $A \subset [0,1]$,

$$T_y(A) := \inf\{t \geq 0 : \prod_{\ell=1}^{L} Y_\ell(t) \in A, \ Y_\ell(0) = y_\ell \text{ for } 1 \leq \ell \leq L \},$$

for some given starting frequency $y \in (0,1)^L$. In the simulations we start with the fit type in one copy at each site, i.e. L individuals are heterozygous at a single site, and homozygous for the wild type at all other sites, and $y = (1/\mathfrak{C}, \ldots, 1/\mathfrak{C})$. We define 'fixation' as the event $\{Y_\ell(t) = 1 \forall \ell \in \{1, \ldots, L\}\}$. If the fit type is lost $(Y_\ell(t) = 0)$ at any site fixation by definition cannot occur. Given the definition of $T_y(A)$ we can define the probability of fixation at L sites, and the expected time to fixation conditional on fixation

$$
\begin{aligned}
p_y(1) &:= \mathbb{P}\big(T_y(\{1\}) < T_y(\{0\})\big) \\
\tau_y(1) &:= \mathbb{E}\big[T_y(\{1\})|T_y(\{1\}) < T_y(\{0\})\big]
\end{aligned}
\qquad (12)
$$

A mathematical investigation of $p_y(1)$ or $\tau_y(1)$ would necessarily have to consider the L-dimensional process $\{(Y_1(t), \ldots, Y_L(t)); t \geq 0\}$ since one would need to keep track of the number of copies of the fit type at all sites. Here, we do not aim for precise estimates of $p_y(1)$ or $\tau_y(1)$, but will only give few examples of how random sweepstakes, recurrent bottlenecks, dominance mechanisms and epistasis affect the statistics. The fixation probability carries both historical and practical weight, as it can inform about adaptation, loss of genetic variation, and the speed of emergence of resistance to antibiotics among pathogens[37]. The time to fixation (conditional on fixation) is also important[38]. A strongly advantageous type in a single population evolving according to the Moran model takes on the order of $\log(N)$ time units to fix (given it will fix) on average[13]; in a structured population the time to fixation depends on the migration rate between the subpopulations[31].

Figure 1 records examples of excursions to fixation at two linked sites with the weight function given by Eq (9). In Eq (9) the weight function increases by one if the fit type is found in homozygous state at a site; there is no epistasis. In all graphs of fixation trajectories (Fig 1, and Fig A1– A8 in § A) the trajectories in each panel were all obtained under identical conditions. We also remark that the scale of the abscissa (time axis) may differ between panels in each figure. The panels record excursions for ε (the probability of taking $\alpha = 0.75$ in Eq (5)) and b (the probability of a bottleneck in any given generation) as shown. The excursions in Figure 1 are shown as the number of diploid individuals homozygous for the

fit type at both sites relative to the population size as a function of time. The panels in the right column (b, d, f, h) record the corresponding trajectories in the left column (a, c, e, g) with the time of each trajectory normalized by the time to fixation for the trajectory so that each trajectory reaches the value one at time one. A fixation trajectory for a single mutation with selective advantage $1/N \ll s \ll 1$ has been described as consisting of "phases", with the phase in which the mutation sweeps from "low" ($k \ll N$) frequency to "high" frequency lasting on the order of $1/s$ generations[39]. The $1/s$ estimate, coupled with the approximation $\ln(2Ns)/s$ for the average time it takes a mutation to sweep in a population evolving according to the Wright-Fisher model necessarily requires $2Ns > e$. In Figure 1a we see that it takes roughly 300-400 generations for a mutation to fix in the absence of random sweepstakes, which matches reasonably well with the estimate $\mathcal{O}(\ln(2Ns)/s)$ of the time to fixation conditional on fixation for a single mutation in a diploid population evolving according to the Wright-Fisher model[40] even though in our case the trajectories represent the joint fixation at two sites. In Figure 1a we also see that the time it takes the trajectory to increase from "low" to "high" frequency is roughly 100 generations, or around tenfold the estimate $1/s$ given in [39].

The logistic differential equation

$$\frac{dp}{dt} = sp(1-p) \tag{13}$$

has been applied in describing how an advantageous type at a single site with selective advantage s increases in frequency in a population during a sweep[41]. A different approach based on random partitions governed by a stick-breaking construction has also been investigated[42]. In our case with multiple (at least two) sites, we suggest extending the random partitions approach of [42] (see also [43]) to incorporate multiple sites might be feasible. We are interested in investigating how well fixation trajectories tracing the joint fixation at two or more sites compare to the sigmoidal shape predicted by Eq (13) traditionally used to describe fixation trajectories. The shape of the trajectories in Figure 1 is clearly not captured by Eq (13). An extensive analysis of genomic data from Atlantic cod indicates that strong positive selection is pervasive, leading to convex (or U-shaped) site-frequency spectrum being consistently observed throughout the genome. Suppose that many of the mutations observed in a U-shaped site-frequency spectrum are selectively advantageous mutations traveling along a fixation trajectory. Then we would like to understand if a particular shape of a fixation trajectory predicts a U-shaped site-frequency spectrum, and another particular shape not. Knowing the potential correspondence between the shape of a fixation trajectory and the shape of site-frequency spectrum may also inform about the likely dominance mechanism of positive mutations.
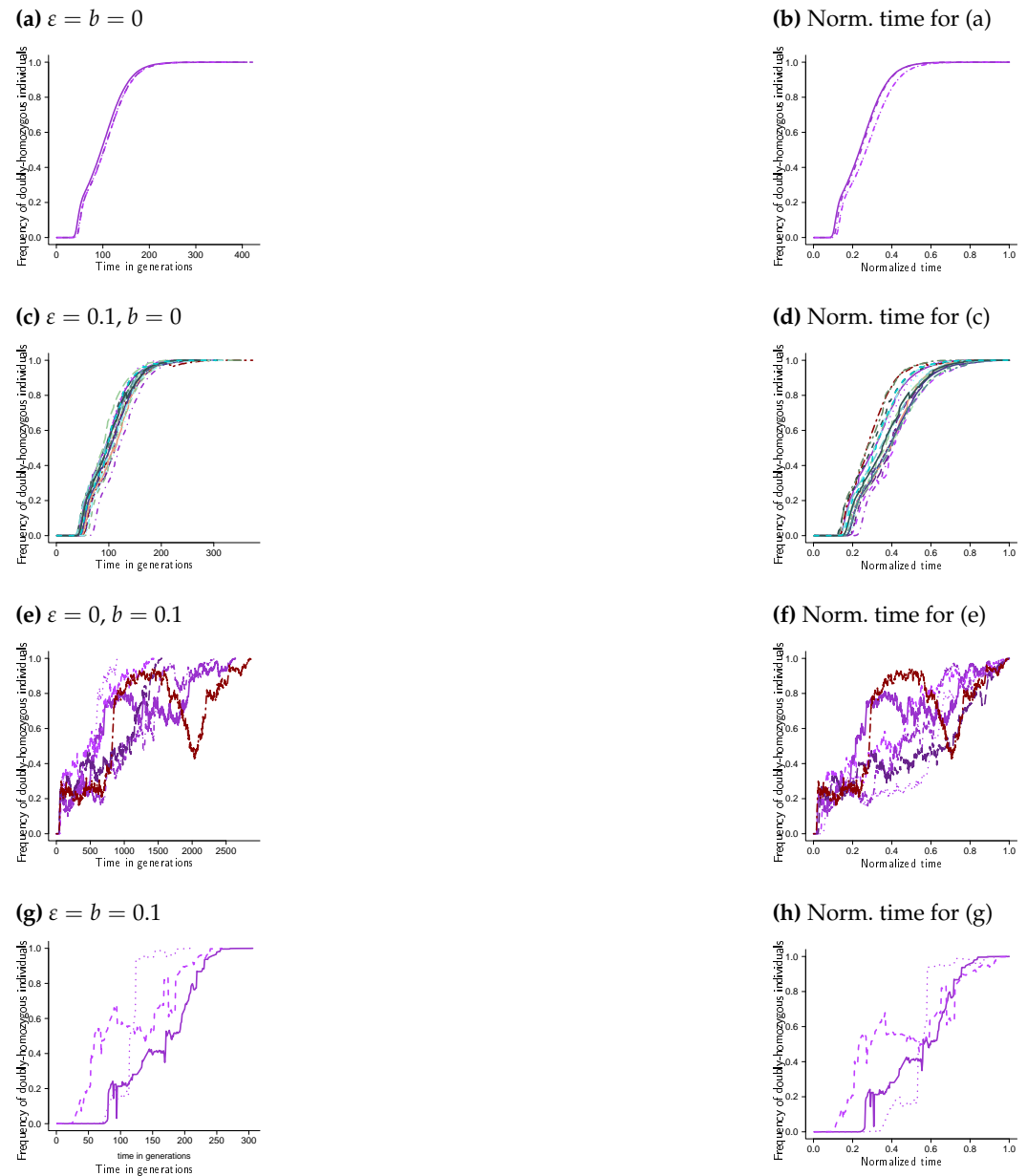
The trajectories in Figure 1a (and hence also those in Figure 1b) are from ten experiments, and in all other panels from one hundred (10^2) experiments; thus random sweepstakes and recurrent bottlenecks reduce $p_y(1)$ somewhat. We emphasize that we are not aiming for precise estimates of $p_y(1)$ or $\tau_y(1)$ defined in Eq (12). Bottlenecks in the absence of random sweepstakes clearly tend to increase the time to fixation (Figure 1e), with the effect of bottlenecks on the time to fixation negated by random sweepstakes (Figure 1g). In our framework we keep the cutoff $u(N)$ in Eq (5) fixed at the carrying capacity. When a bottleneck occurs the population size is reduced to some given value, which is less than the carrying capacity. The individuals producing juveniles right after a bottleneck do so with the cutoff higher than the current population size, resulting in individuals having a higher chance of producing a large (relative to the current population size) number of juveniles. Simulation results have shown that increasing the cutoff relative to the population size reduces the fixation time (conditional on fixation)[44]. In Figure 1c we see little effect of random sweepstakes on the time to fixation in the absence of bottlenecks; we claim this is due to taking the cutoff the same as the population size, or the carrying capacity in this case.

Normalizing the time to fixation (Figure 1b, 1d, 1f, 1h where the time to fixation for each trajectory is normalized by the time to fixation for the trajectory) informs about the time spent in each "phase" of the trajectory. For example, in Figure A2b the time to fixation is clearly dominated with the fit type in high frequency at both sites. In contrast, in Figure 1b the trajectory is increasing essentially linearly from low to high frequency and in doing so spends around 20% of the time to fixation at "intermediate" frequency. Given the estimate $1/s$ stated in [39], a type taking on average of order $\log(2Ns)/s$ generations to travel to fixation should then be spending on average about $1/\log(2Ns)$ of the time to fixation in "intermediate" frequency. For $N = 10^6$ and $s = 0.1$ this is around 8%.

Further examples of fixation trajectories are given in Figures A1– A8 in § A. Comparing the results, for the given weight functions, reveals that there are two main patterns. One pattern is the one in Figure 1, which is essentially mirrored in Figures A1, A4, A5, A7, and Figure A8. In Figure A1 we take the weight function as in Eq (A1), as in Eq (A4) for Figure A4, in Eq (A5) for Figure A5, and as in Eq (A4) for Figures A7 and A8. In Figure A7 we take $s = 0.1$, and in Figure A8 we take $s = 0.01$, the other parameter values are the same.

The other pattern is represented in Figures A2 and A3 (both for two sites), and in Figure A6 (for three sites). Taking the weight function as in Eq (A2) yields the results in Figure A2, in Eq (A3) the results in Figure A3 (two sites) and in Figure A6 (three sites). In this pattern the trajectories without random sweepstakes or bottlenecks (Figure A2a) have the shape of a completely dominant type at a single site (see e.g. [17]), although the weight function (Eq (A2) resembles one for a recessive type. In addition, random sweepstakes clearly shorten the time to fixation in the absence of bottlenecks (e.g. Figure A6c), and counteract the effect of bottleneck on the time to fixation (conditional on fixation; e.g. Figure A7g).

(a) $\varepsilon = b = 0$ **(b)** Norm. time for (a)

(c) $\varepsilon = 0.1$, $b = 0$ **(d)** Norm. time for (c)

(e) $\varepsilon = 0$, $b = 0.1$ **(f)** Norm. time for (e)

(g) $\varepsilon = b = 0.1$ **(h)** Norm. time for (g)

Figure 1. Fixation at two sites with weight function as in Eq (9). Examples of excursion to fixation at two sites ($L = 2$) with weight as given by Eq (8) with weight function as given by Eq (9) for carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$ (see Eq (5)), probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, with ε (see Eq (7)) and b (the probability of a bottleneck in a given generation) as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$ throughout, number of replicates 10^2 except ten for (a). Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

4. Discussion

We have given examples of the joint effects of random sweepstakes, recurrent bottlenecks, and dominance mechanisms on the fate of selectively advantageous mutations at linked sites. We have shown that, depending on the dominance mechanism, random

sweepstakes can shorten the time to fixation (conditional on fixation at all the sites). Bottlenecks can increase the fixation time, and random sweepstakes counteract the effect of bottlenecks on the fixation time regardless of the dominance mechanism. The shape of the fixation trajectories can be broadly divided into two categories in the absence of bottlenecks, one with a linear increase from "low" to "high" frequency, and another where the trajectories spend nearly the whole time with the fit type in high frequency at all sites. We emphasize that we do not aim for precise estimates of the probability of fixation or the time to fixation (conditional on fixation), or to investigate many possible scenarios. However, our algorithm provides a tool for further investigations of various demographic and molecular mechanisms on selection. Similar results on the interaction between bottlenecks and random sweepstakes have been obtained for selection acting at a single site[17] and unlinked sites[16].

Studied on selection acting at more than one site generally focus on how a trait value affected by the sites evolves, in particular in reference to a given optimum (e.g. [36]). The effects of the sites on the trait value may vary, and some studies have investigated the evolution of a trait given a change in the optimum when the trait value is determined by many sites (or loci) [45,46]. Our simulation algorithm can, in principle, handle any number of linked sites. However, in particular given that we need to keep track of the number of copies of the fit type at all sites, there are practical considerations (for example computer memory) that limit the number of sites that in practice can be considered. In addition, adaptation may be at least to a large degree due to a few genes of major effect[34].

Our algorithm keeps track of phased haplotypes, so we can investigate the effect of epistasis on selection. Epistasis may be a key player in adaptation[47–49]. Significant amounts of genetic variation despite pervasive strong selection may indicate that there is considerable amount of 'genetic constraints'[50]. Maybe selection acting on two or more sites with epistatic interactions between the sites can be a way of incorporating genetic constraints.

Evolution can occur "rapidly", as many examples from nature show [51–56]. Our results show that the joint fixation at two or more sites occurs on a timescale comparable to $\mathcal{O}(\ln(2Ns)/s)$, the order of time it takes a sweep of a single beneficial mutation of advantage s to complete. Understanding the time to fixation at two or more sites may be important in the battle against antibiotic and pesticide resistance[57–59].

An open problem is rigorously verifying the simulation results. Here multi-type branching process theory[60–65] may be useful. Mathematical analysis of the model of random sweepstakes given in Eq (7) and applied in the simulations is ongoing. There is also need for further mathematical theory for inferring evolutionary histories of natural populations, in particular in view of the results of [66]. There may be a high cost of selection in populations evolving according to recurrent selective sweepstakes[67]. We anticipate that models of random sweepstakes incorporating strong positive selection in the form of recurrent selective sweeps of strongly advantageous types arising by mutation will be necessary for understanding high fecundity. In particular, high fecundity and random sweepstakes might enable populations to 'pay for selection', i.e. withstand the cost of strong pervasive positive selection.

Data Availability Statement: The CWEB (C++) code written for the simulations is available at https://github.com/eldonb/fixation_many_sites

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

In this section we give further examples of fixation trajectories (Figures A1– A8) for various weight functions. Throughout the viability weight is computed as in Eq (8). In each graph a trajectory to fixation traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). Throughout we consider viability selection acting at linked sites in a diploid population evolving according to Eq (7) with $\alpha_1 = 0.75$ and $\alpha_2 = 3$ and the cutoff $u(N)$ in Eq (5) set to the carrying capacity $\mathfrak{C}$. Every time we start with the fit type in a single copy at each site for all sites. Furthermore, each diploid individual is heterozygous at most at one site to begin with, i.e. at the start each individual carries at most one copy of a fit type. The algorithm is given as a pseudocode in § B.

Let h_i for $i = 1, 2$ denote haplotype i (arbitrarily labeled) in a given juvenile. In Figure A1 we record examples of excursions to fixation at two sites with weight function (recall the definition of a viability weight in Eq (8)) as given by

$$f(G) = \mathbb{1}_{\{h_1=(1,...,1)\}} + \mathbb{1}_{\{h_2=(1,...,1)\}} \tag{A1}$$

and recall that we denote a genome by G. Compare Eq (A1) with Eq (9). In Eq (A1) the value of the weight function increases by one for each haplotype carrying the fit type at all sites.

Figure A2 records examples of excursions to fixation at two sites for the weight function

$$f(G) = \mathbb{1}_{\{\text{at least three copies of the fit type}\}} \tag{A2}$$

The idea behind the mechanism in Eq (A2) is to model essentially a recessive fit type, however the excursions in Figure A2a resemble ones for a dominant type.

In Figure A3 with weight function for two sites as given by (with h_i for $i \in \{1,2\}$ denoting haplotype i, arbitrarily labeled, in a given juvenile)

$$f(G) = \mathbb{1}_{\{\{h_1=(1,...,1)\} \cup \{h_2=(1,...,1)\}\}} \tag{A3}$$

The weight function in Eq (A3) takes the value one if and only if at least one of the two haplotypes in a given juvenile carries the fit type at all sites. In Figure A6 we record fixation trajectories for three sites with weight function given by Eq A3.

Figure A4 records trajectories for two sites when the weight function is of the form
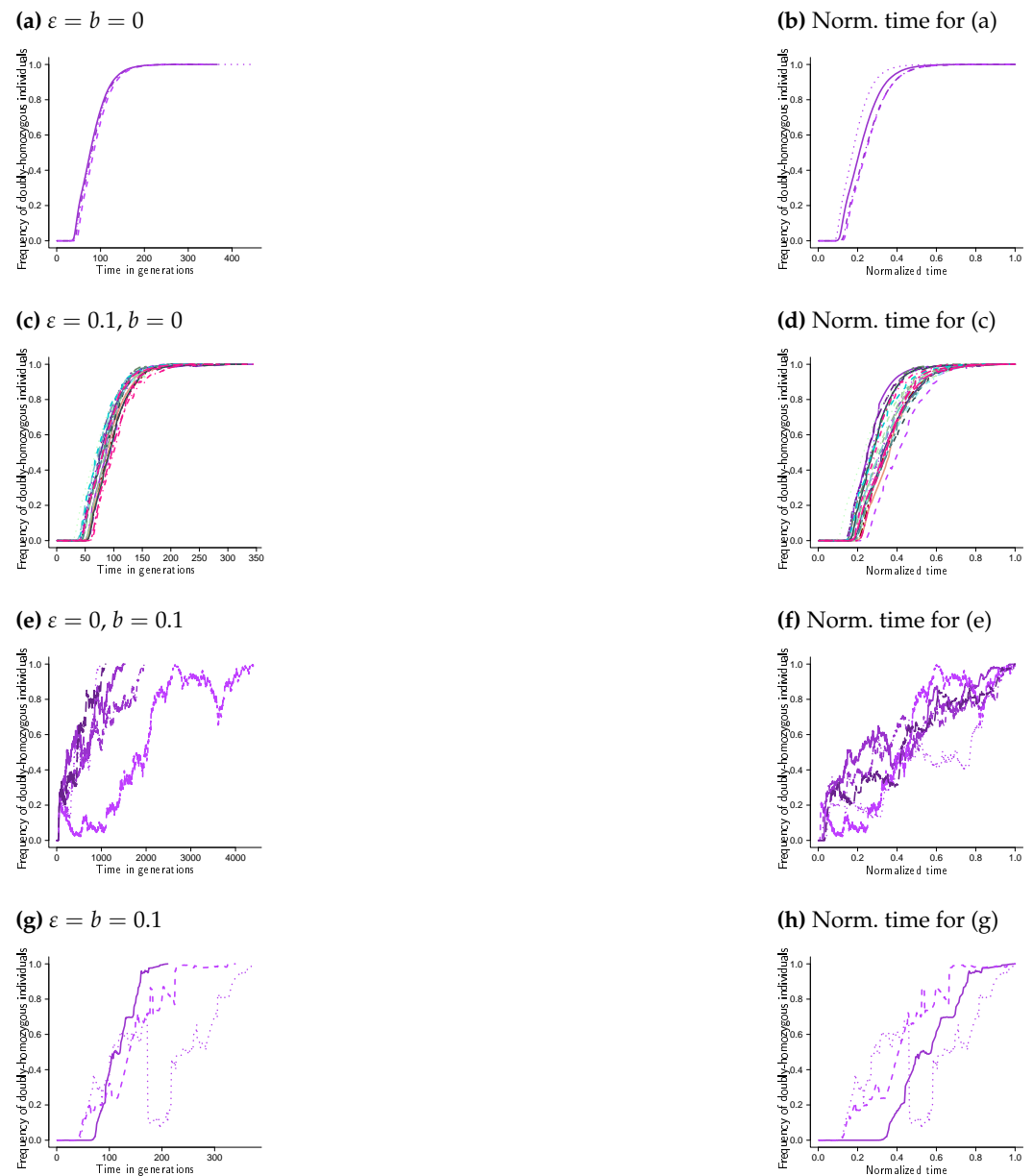
$$f(G) = \mathbb{1}_{\{\{h_1\neq(1,...,1)\} \cap \{h_2\neq(1,...,1)\}\}} \tfrac{1}{2} + \mathbb{1}_{\{\{h_1=(1,...,1)\} \cap \{h_2=(1,...,1)\}\}} \tag{A4}$$

In Eq (A4) a juvenile homozygous for the fit type at all sites has maximum fitness, but having one haplotype carrying only a fit type brings disadvantage if the other haplotype does not follow suit. In Figure A7 and Figure A8 we record fixation trajectories for three sites with weight function as in Eq A4, in Figure A8 with $s = 0.01$.

In Figure A5 we record fixation trajectories for three sites with weight function given by
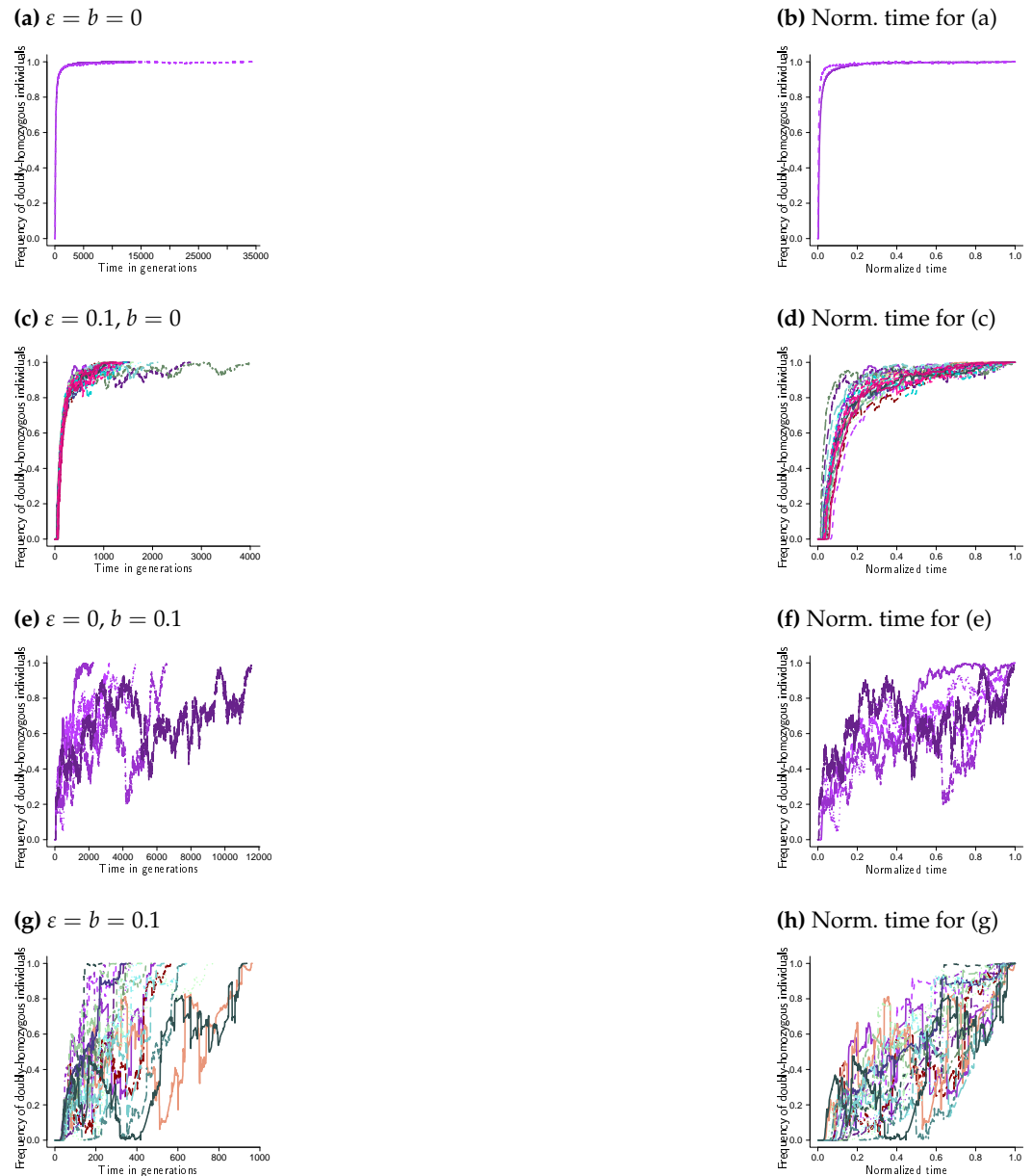
$$f(G) = \mathbb{1}_{\{\{h_1=(1,1,1)\} \cap \{h_2=(1,1,1)\}\}} \tag{A5}$$

In Eq (A5) the weight function in Eq (8) takes the value one if a juvenile is homozygous for the fit type at all three sites, otherwise the value zero.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1$, $b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0$, $b = 0.1$

(f) Norm. time for (e)

**(g)** $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A1. Two sites with weight function as in Eq (A1). Examples of trajectories to fixation at two sites with weight determined by Eq (8) with weight function as in Eq (A1) for carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$. Each panel shows the trajectories from ten experiments for (a) and (b) otherwise from 10^2 experiments. Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1$, $b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0$, $b = 0.1$

(f) Norm. time for (e)

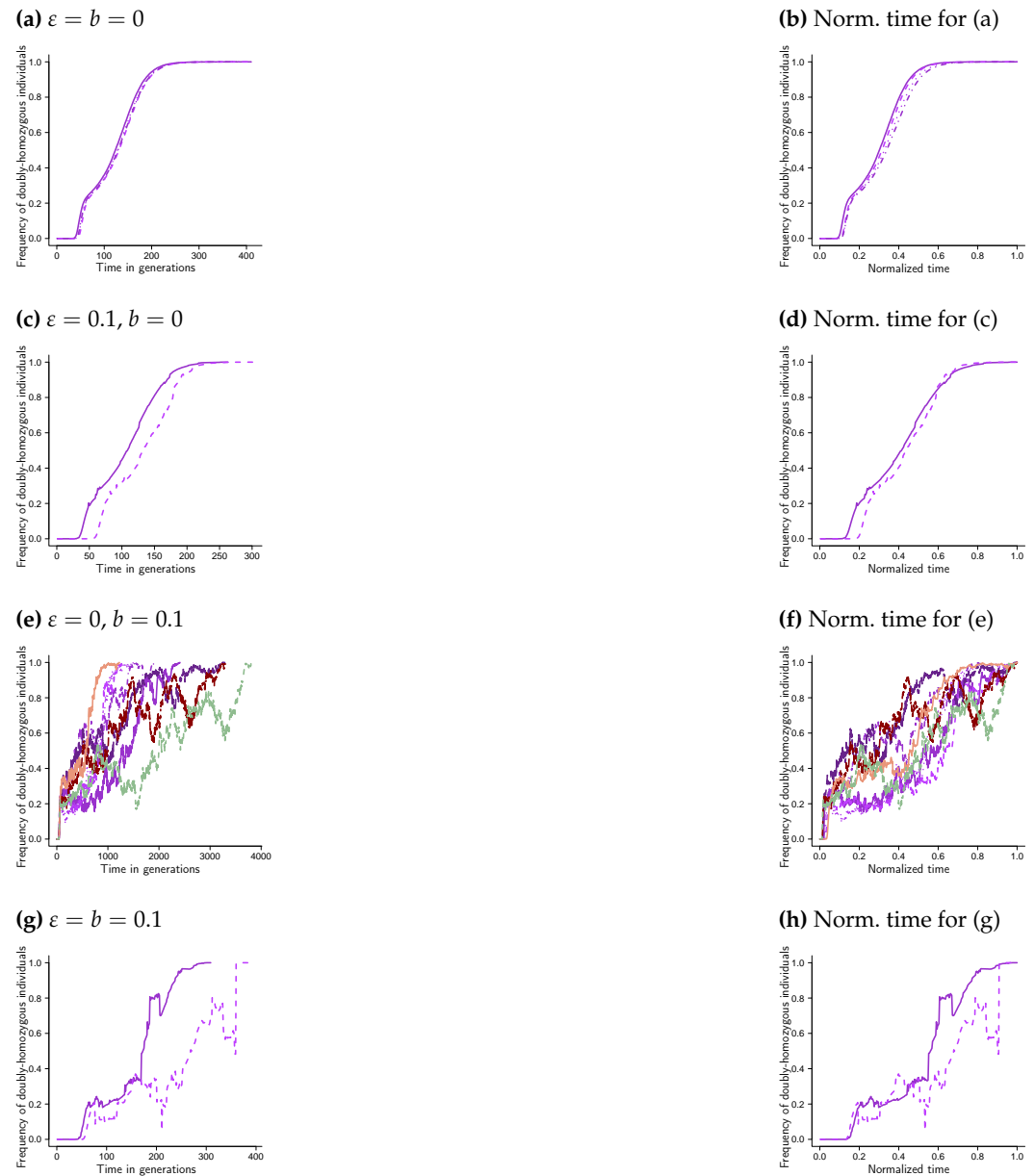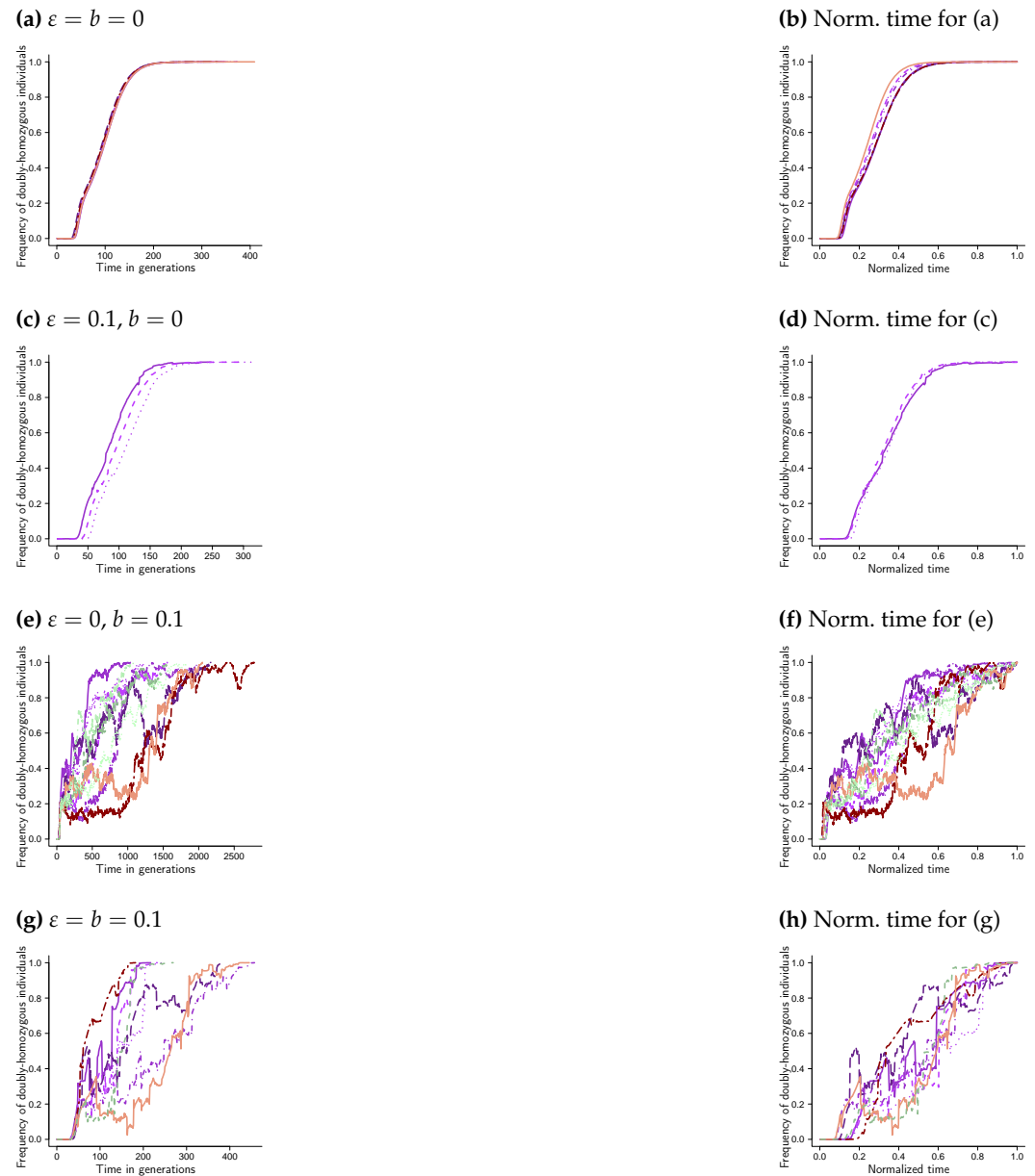**(g)** $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A2. Two sites with weight function as in Eq (A2). Examples of trajectories to fixation at two sites with weight determined by Eq (8) with weight function as in Eq (A2) for carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$, results out of ten experiments for (a); 10^2 experiments for (c); 10^3 experiments for (e) and (g). Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels. In each panel all the trajectories were obtained under the same conditions.
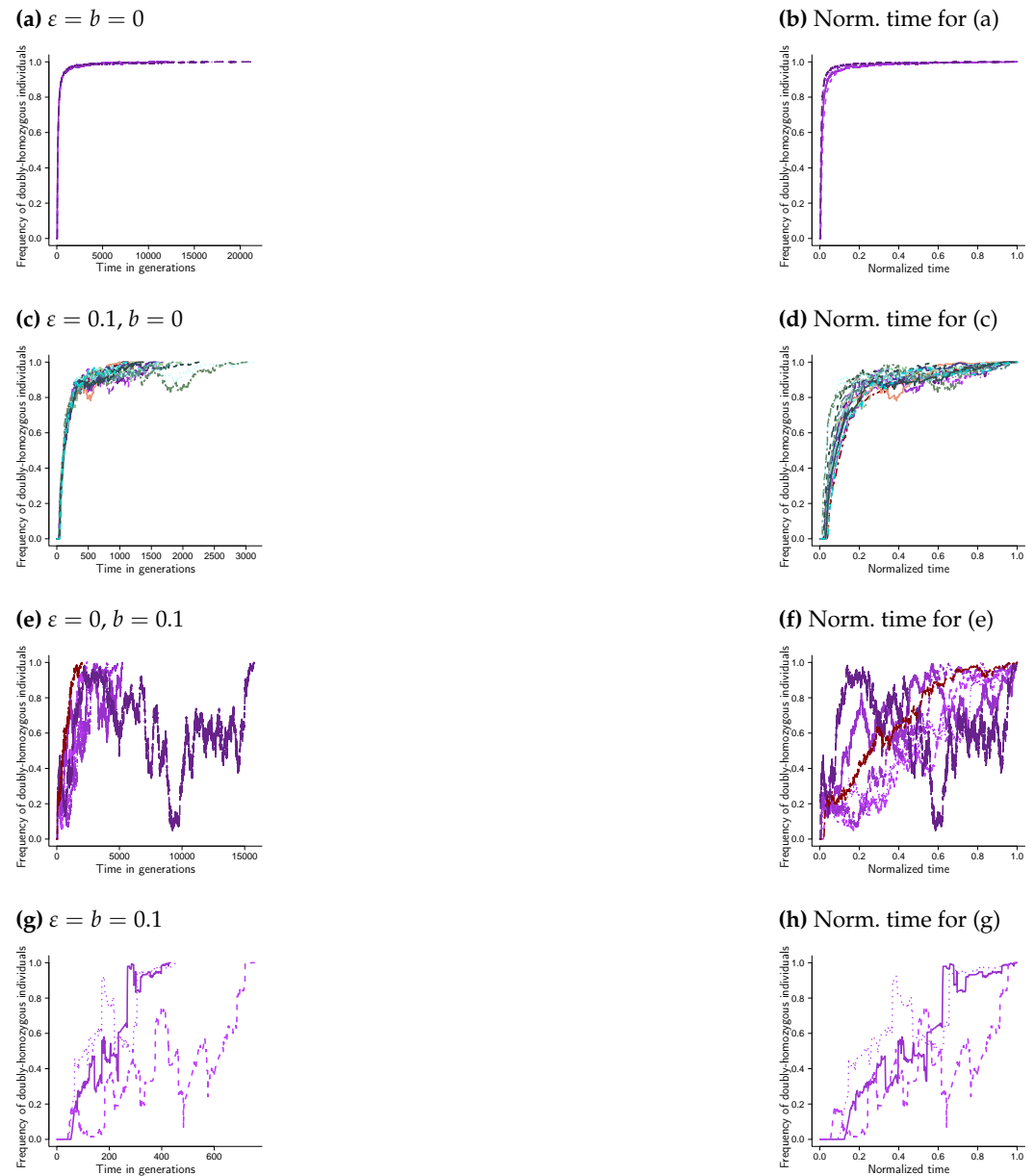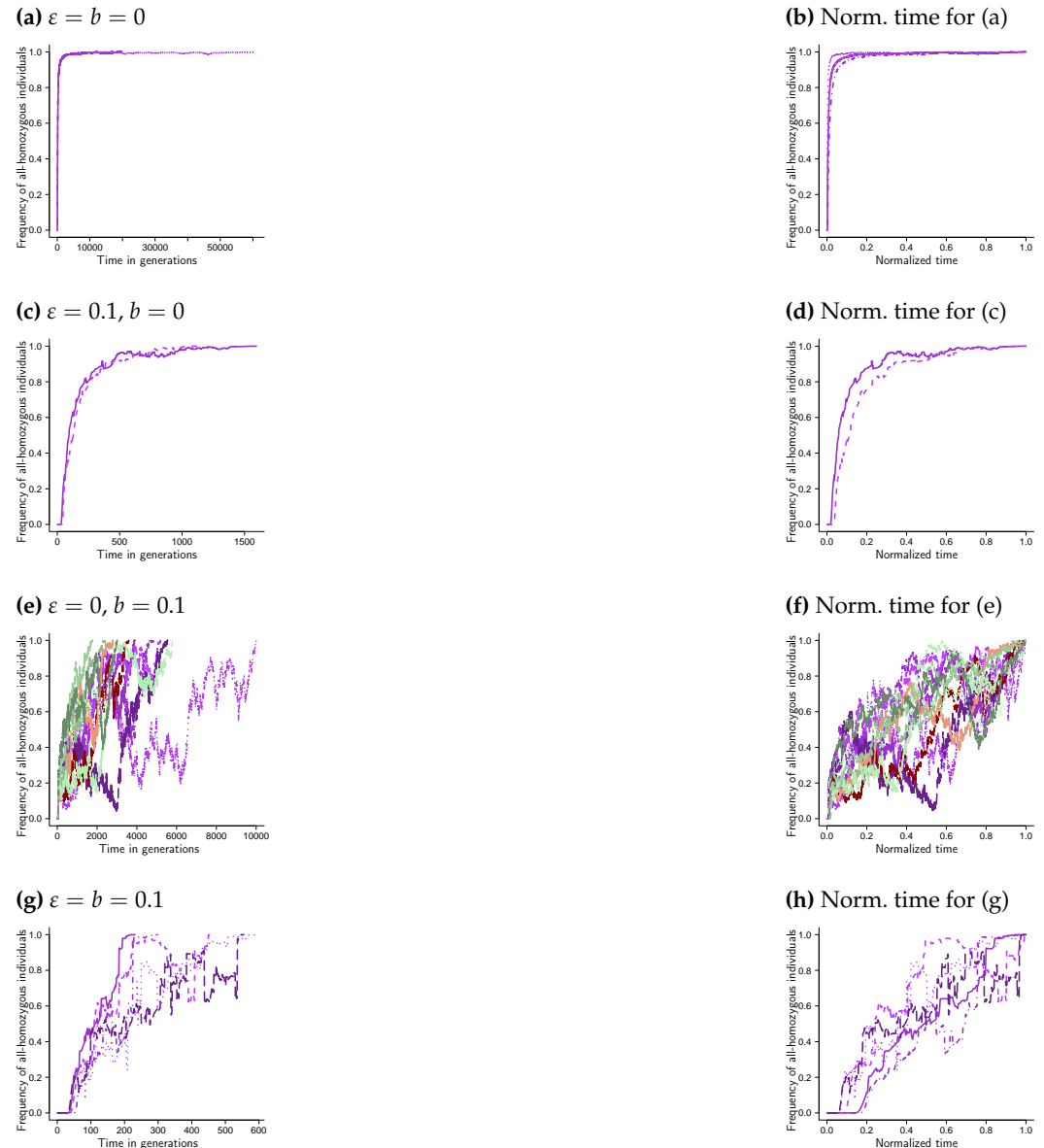
Figure A3. Two sites with weight function as in Eq (A3). Examples of excursions to fixation at two sites with weight function as given by Eq (A3) for carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$, results out of ten experiments for (a); otherwise 10^2 experiments. Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b,d,f,h) show the corresponding excursions from the left panels (a,c,e,g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1$, $b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0$, $b = 0.1$

(f) Norm. time for (e)

(g) $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A4. Two sites with weight function as in Eq (A4). Examples of excursions to fixation at two sites with weight function as given by Eq (A4) for carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$; results from ten experiments for (a) and (c); otherwise from 10^2 experiments. Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b,d,f,h) show the corresponding excursions from the left panels (a,c,e,g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1, b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0, b = 0.1$

(f) Norm. time for (e)

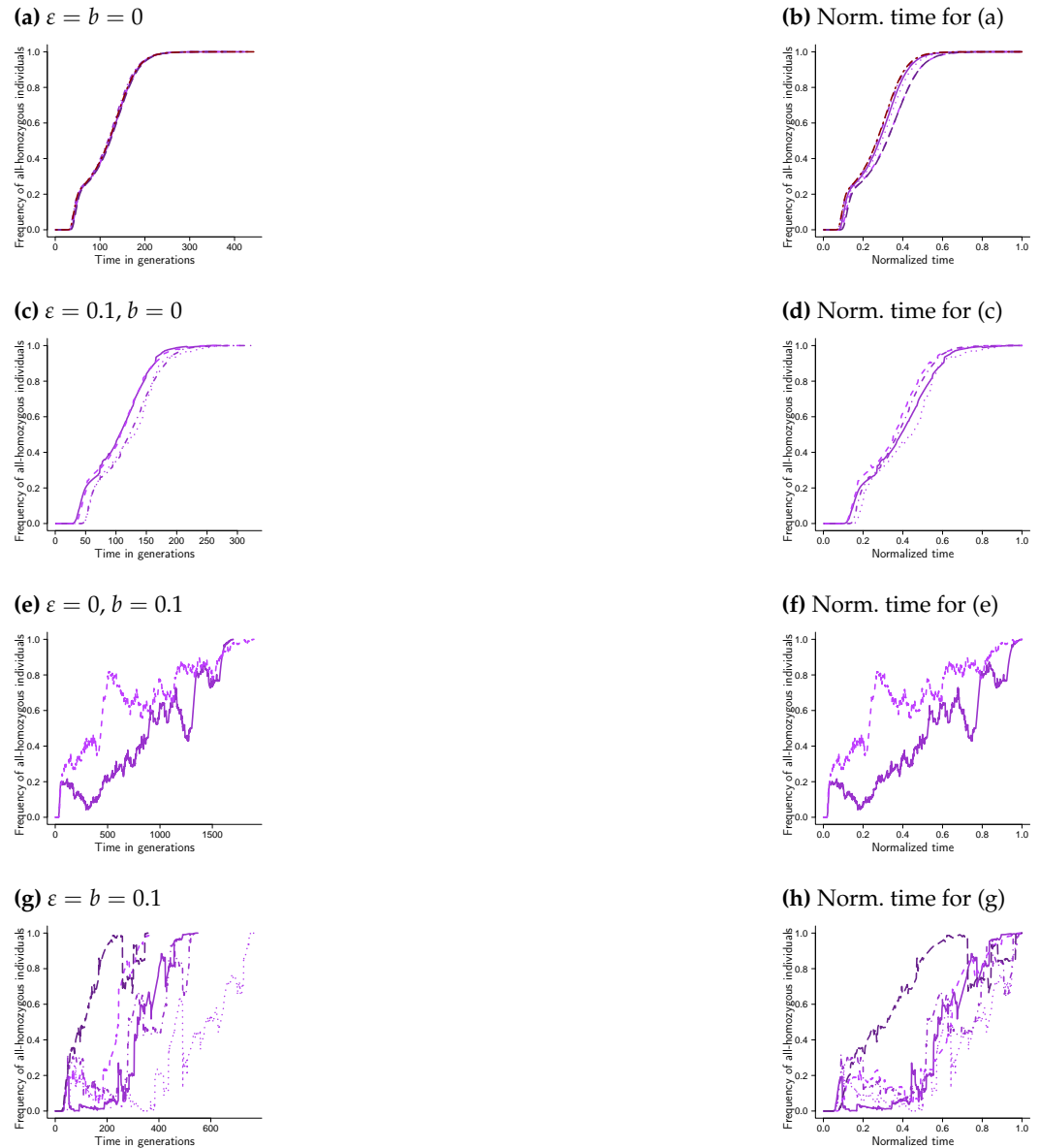**(g)** $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A5. Three sites with weight function as in Eq (A5). Examples of fixation trajectories at three sites with weight function as in Eq (A5), carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$; excursions trajectories from ten trials in (a) and (c), from 10^2 trials in (e) and (g). Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1, b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0, b = 0.1$

(f) Norm. time for (e)

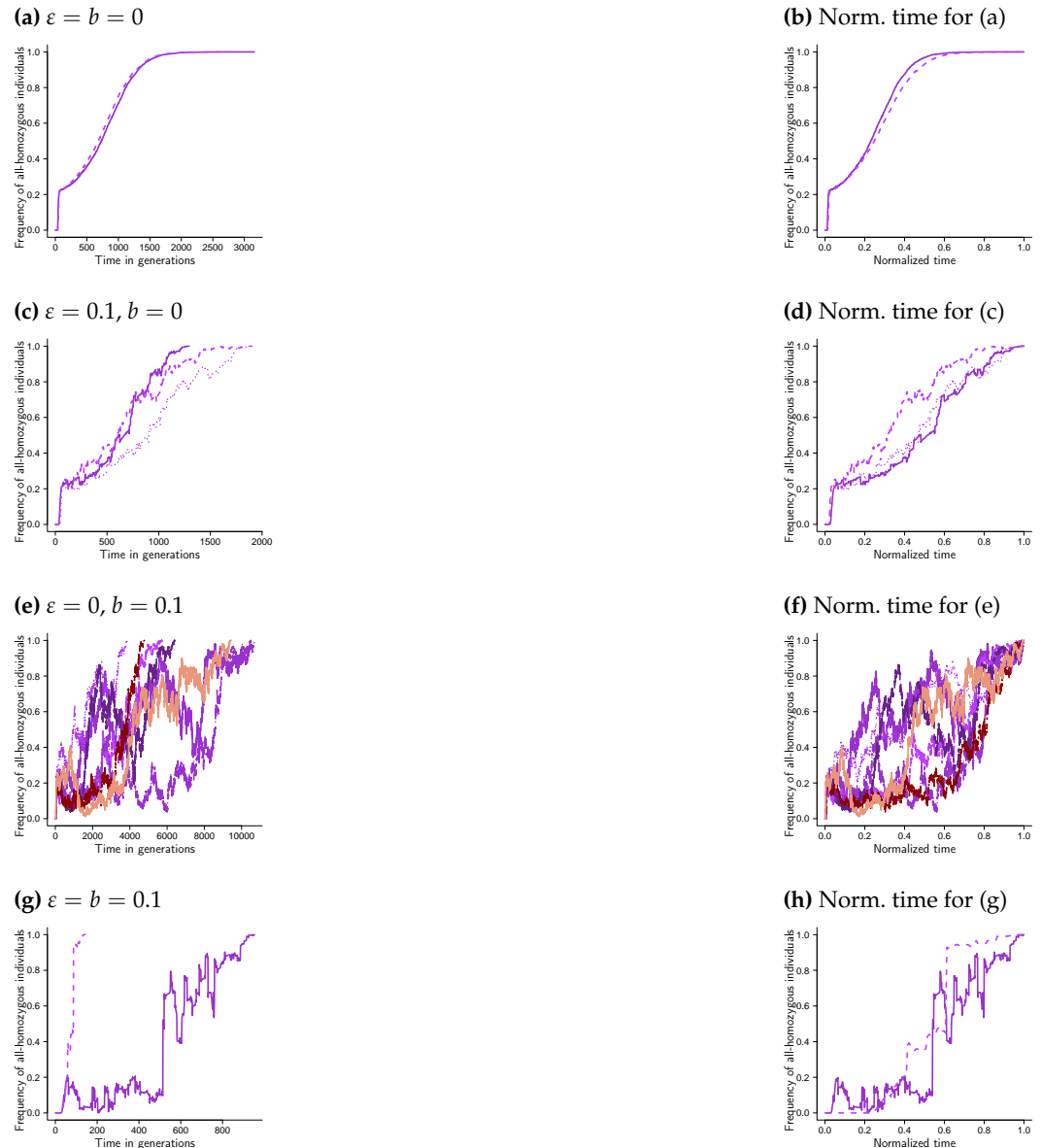**(g)** $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A6. Three sites with weight function as in Eq (A3). Examples of fixation trajectories for fixation at three sites with weight function given by Eq (A3), carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$; (a) and (c) from ten experiments, (e) and (g) from 10^2 experiments. Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1, b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0, b = 0.1$

(f) Norm. time for (e)

(g) $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A7. Three sites with weight function as in Eq (A4). Examples of fixation trajectories for fixation at three sites with weight function as in Eq (A4), carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.1$, bottleneck $B = 10^3$. In (a), (c), (e) trajectories from ten experiments, (g) from 10^2 experiments. Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

(a) $\varepsilon = b = 0$

(b) Norm. time for (a)

(c) $\varepsilon = 0.1, b = 0$

(d) Norm. time for (c)

(e) $\varepsilon = 0, b = 0.1$

(f) Norm. time for (e)

(g) $\varepsilon = b = 0.1$

(h) Norm. time for (g)

Figure A8. Three sites with weight function as in Eq (A4). Examples of fixation trajectories at three sites with weight function as given by Eq (A4) ; carrying capacity $\mathfrak{C} = 10^6$, cutoff $u(N) = \mathfrak{C}$, probability of recombination $r = 0.25$ between adjacent sites, $\alpha_1 = 0.75$, $\alpha_2 = 3$, ε (see Eq (7)) and probability of a bottleneck b as shown, strength of selection $s = 0.01$, bottleneck $B = 10^3$. In (a), (c) trajectories from ten experiments, trajectories in (e) and (g) from 10^2 experiments. Each trajectory traces, as a function of time, the number of diploid individuals homozygous for the fit type at all sites relative to the population size (the number of diploid individuals). The panels on the right (b, d, f, h) show the corresponding excursions from the left panels (a, c, e, g) where the time to fixation for each excursion is normalised by the time to fixation for the excursion. The scale of the abscissa (time axis) may vary between panels; in each panel the trajectories were all obtained under identical conditions.

Appendix B

In this section we give the algorithm for the simulations in the form of a pseudocode. Each and every time we start with the fit type in one copy at each site, with any given diploid individual initially carrying at most one copy of a fit type. Each and every time the population size is initially set to the carrying capacity $\mathfrak{C}$.

1.   Initialize the frequency of the fit type at each site, i.e.

$$(Y_1(0),\dots,Y_L(0)) \leftarrow \left(\tfrac{1}{2\mathfrak{C}},\dots,\tfrac{1}{2\mathfrak{C}}\right)$$

with any diploid individual being heterozygous at most at one site.

2.   While $\left\{0 < \prod_{\ell\in[L]} Y_\ell(t) < 1\right\}$ holds then for generation $t+1$ started with N_t individuals:

(a)   sample a random uniform $U \leftarrow$ random uniform on $(0,1)$ to check if a bottleneck occurs

(b) $M \leftarrow \mathbb{1}_{\{U\le b\}} B + \mathbb{1}_{\{U>b\}} N_t$ where b is the probability of a bottleneck and B the number of individuals surviving a bottleneck

(c)   in the case of a bottleneck the M surviving individuals are sampled uniformly at random without replacement from among the N_t individuals

(d)   the M individuals randomly form $\lfloor M/2 \rfloor$ pairs that will go on to produce juveniles

(e)   sample $S_{\lfloor M/2 \rfloor} := X_1 + \cdots + X_{\lfloor M/2 \rfloor}$ random number of juveniles according to Eq (7) produced independently by $\lfloor M/2 \rfloor$ randomly formed parent pairs

(f)   assign types to the juveniles according to Mendel's laws allowing for recombination between sites (we exclude mutation)

(g)   if $S_{\lfloor M/2 \rfloor} > \mathfrak{C}$ for each juvenile independently sample a random exponential with rate as in Eq (8), compute the $\mathfrak{C}$th largest exponential $E_{(\mathfrak{C})}$, and a juvenile with exponential $E \le E_{(\mathfrak{C})}$ survives

(h)   if $S_{\lfloor M/2 \rfloor} \le \mathfrak{C}$ all the juveniles survive so that $N_{t+1} = \min\left(S_{\lfloor M/2 \rfloor}, \mathfrak{C}\right)$ for any $t \in \mathbb{N}$

3.   Record the result of the trajectory; if the event $\left\{\prod_{\ell\in[L]} Y_\ell(t) = 0\right\}$ occurred a fit type was lost, and if $\left\{\prod_{\ell\in[L]} Y_\ell(t) = 1\right\}$ occurred the fit type reached fixation at all sites, with $[L] := \{1,2,\dots,L\}$.

The documented CWEB (C++) code is freely available at https://github.com/eldonb/fixation_many_sites.

References

1. Eldon, B. Evolutionary Genomics of High Fecundity. *Annual Review of Genetics* **2020**, *54*, 213–236. https://doi.org/10.1146/annurev-genet-021920-095932.
2. Hedgecock, D.; Pudovkin, A.I. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull Marine Science* **2011**, *87*, 971–1002.
3. Hedgecock, D. Does variance in reproductive success limit effective population sizes of marine organisms? In Proceedings of the Genetics and evolution of Aquatic Organisms; Beaumont, A., Ed.; Chapman and Hall: London, 1994; pp. 1222–1344.
4. Árnason, E.; Halldórsdóttir, K.; Koskela, J.; Eldon, B. Sweepstakes reproductive success via pervasive and recurrent selective sweeps. *bioRxiv* **2022**, *x*, y. https://doi.org/10.1101/2022.05.29.493887.
5. Williams, G. *Sex and Evolution*; Monographs in Population Biology, Princeton University Press, 1975.
6. Möhle, M.; Sagitov, S. Coalescent patterns in diploid exchangeable population models. *J Math Biol* **2003**, *47*, 337–352.
7. Sagitov, S. Convergence to the coalescent with simultaneous mergers. *J Appl Probab* **2003**, *40*, 839–854.
8. Möhle, M.; Sagitov, S. A classification of coalescent processes for haploid exchangeable population models. *Ann Probab* **2001**, *29*, 1547–1562.
9. Schweinsberg, J. Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch Proc Appl* **2003**, *106*, 107–139.
10. Eldon, B.; Wakeley, J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **2006**, *172*, 2621–2633.
11. Sargsyan, O.; Wakeley, J. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor Pop Biol* **2008**, *74*, 104–114.
12. Birkner, M.; Blath, J.; Eldon, B. An ancestral recombination graph for diploid populations with skewed offspring distribution. *Genetics* **2013**, *193*, 255–290.
13. Durrett, R.; Schweinsberg, J. A coalescent model for the effect of advantageous mutations on the genealogy of a population. *Stochastic Processes and their Applications* **2005**, *115*, 1628–1657. https://doi.org/10.1016/j.spa.2005.04.009.
14. Birkner, M.; Liu, H.; Sturm, A. Coalescent results for diploid exchangeable population models. *Electronic Journal of Probability* **2018**, *23*. https://doi.org/10.1214/18-ejp175.
15. Koskela, J.; Berenguer, M.W. Robust model selection between population growth and multiple merger coalescents. *Mathematical biosciences* **2019**, *311*, 1–12.
16. Eldon, B. Genome-wide fixation under viability selection **2022**. https://doi.org/10.1101/2022.09.19.508488.
17. Eldon, B.; Stephan, W. Sweepstakes reproduction facilitates rapid adaptation in highly fecund populations **2022**. https://doi.org/10.22541/au.165337209.92735972/v1.
18. Pitman, J. Coalescents with multiple collisions. *Ann Probab* **1999**, *27*, 1870–1902.
19. Sagitov, S. The general coalescent with asynchronous mergers of ancestral lines. *J Appl Probab* **1999**, *36*, 1116–1125.
20. Donnelly, P.; Kurtz, T.G. Particle Representations for Measure-Valued Population Models. *Ann Probab* **1999**, *27*, 166–205.
21. Schweinsberg, J. Coalescents with simultaneous multiple collisions. *Electron J Probab* **2000**, *5*, 1–50.
22. Birkner, M.; Blath, J.; Möhle, M.; Steinrücken, M.; Tams, J. A modified lookdown construction for the Xi-Fleming-Viot process with mutation and populations with recurrent bottlenecks. *ALEA Lat. Am. J. Probab. Math. Stat.* **2009**, *6*, 25–61.
23. Freund, F.; Kerdoncuff, E.; Matuszewski, S.; Lapierre, M.; Hildebrandt, M.; Jensen, J.D.; Ferretti, L.; Lambert, A.; Sackton, T.B.; Achaz, G. Interpreting the pervasive observation of U-shaped Site Frequency Spectra **2022**. https://doi.org/10.1101/2022.04.12.488084.
24. Birkner, M.; Blath, J. Measure-valued diffusions, general coalescents and population genetic inference. In *Trends in stochastic analysis*; Blath, J.; Mörters, P.; Scheutzow, M., Eds.; Cambridge University Press, 2009; pp. 329–363.
25. Der, R.; Epstein, C.L.; Plotkin, J.B. Generalized population models and the nature of genetic drift. *Theoretical Population Biology* **2011**, *80*, 80–99. https://doi.org/10.1016/j.tpb.2011.06.004.
26. Dynkin, E. *Markov Processes*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1965.
27. Etheridge, A.M.; Griffiths, R.C.; Taylor, J.E. A coalescent dual process in a Moran model with genic selection, and the Lambda coalescent limit. *Theor Popul Biol* **2010**, *78*, 77–92.
28. Foucart, C. The impact of selection in the Λ-Wright-Fisher model. *Electron. Commun. Probab* **2013**, *18*, 1–10.
29. Bah, B.; Pardoux, E. The Λ-lookdown model with selection. *Stochastic Processes and their Applications* **2015**, *125*, 1089–1126. https://doi.org/10.1016/j.spa.2014.10.014.
30. Der, R.; Plotkin, J.B. The Equilibrium Allele Frequency Distribution for a Population with Reproductive Skew. *Genetics* **2014**, *196*, 1199–1216. https://doi.org/10.1534/genetics.114.161422.
31. Greven, A.; Pfaffelhuber, P.; Pokalyuk, C.; Wakolbinger, A. The fixation time of a strongly beneficial allele in a structured population. *Electronic Journal of Probability* **2016**, *21*, 1–42. https://doi.org/10.1214/16-ejp3355.
32. Orr, H.A. The distribution of fitness effects among beneficial mutations. *Genetics* **2003**, *163*, 1519–1526.
33. Orr, H.A. The population genetics of beneficial mutations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2010**, *365*, 1195–1201. https://doi.org/10.1098/rstb.2009.0282.
34. Bell, G. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2010**, *365*, 87–97. https://doi.org/10.1098/rstb.2009.0150.

35. Huillet, T.; Möhle, M. On the extended Moran model and its relation to coalescents with multiple collisions. *Theor Popul Biol* **2013**, *87*, 5–14.

36. de Vladar, H.P.; Barton, N. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* **2014**, *197*, 749–767.

37. Patwa, Z.; Wahl, L.M. The fixation probability of beneficial mutations. *Journal of The Royal Society Interface* **2008**, *5*, 1279–1289.

38. Charlesworth, B. How Long Does It Take to Fix a Favorable Mutation, and Why Should We Care? *The American Naturalist* **2020**, *195*, 753–771. https://doi.org/10.1086/708187.

39. BARTON, N.H. The effect of hitch-hiking on neutral genealogies. *Genetical Research* **1998**, *72*, 123–133. https://doi.org/10.1017/s0016672398003462.

40. van Herwaarden, O.A.; van der Wal, N.J. Extinction Time and Age of an Allele in a Large Finite Population. *Theoretical Population Biology* **2002**, *61*, 311–318. https://doi.org/https://doi.org/10.1006/tpbi.2002.1576.

41. Kaplan, N.L.; Hudson, R.R.; Langley, C.H. The" hitchhiking effect" revisited. *Genetics* **1989**, *123*, 887–899.

42. Durrett, R.; Schweinsberg, J. Approximating selective sweeps. *Theor Popul Biol* **2004**, *66*, 129–138.

43. Kingman, J. The representation of partition structures. *J London Math Soc* **1978**, *18*, 374–380.

44. Eldon, B.; Stephan, W. Evolution of highly fecund haploid populations. *Theoretical Population Biology* **2018**, *119*, 48–56. https://doi.org/10.1016/j.tpb.2017.10.002.

45. Pritchard, J.K.; Pickrell, J.K.; Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology* **2010**, *20*, R208–R215.

46. Pritchard, J.K.; Di Rienzo, A. Adaptation–not by sweeps alone. *Nature Reviews Genetics* **2010**, *11*, 665–667.

47. Hansen, T.F. WHY EPISTASIS IS IMPORTANT FOR SELECTION AND ADAPTATION. *Evolution* **2013**, *67*, 3501–3511. https://doi.org/10.1111/evo.12214.

48. MALMBERG, R.L.; MAURICIO, R. QTL-based evidence for the role of epistasis in evolution. *Genetical Research* **2005**, *86*, 89–95. https://doi.org/10.1017/s0016672305007780.

49. Bank, C. Epistasis and Adaptation on Fitness Landscapes, 2022. https://doi.org/10.48550/ARXIV.2204.13321.

50. Walsh, B.; Blows, M.W. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annual Review of Ecology, Evolution and Systematics* **2009**, *40*, 41.

51. Reznick, D.N. *The Origin then and now*; Princeton University Press, 2011.

52. Vignieri, S.N.; Larson, J.G.; Hoekstra, H.E. The selective advantage of crypsis in mice. *Evolution: International Journal of Organic Evolution* **2010**, *64*, 2153–2158.

53. Cook, L.M.; Grant, B.S.; Saccheri, I.J.; Mallet, J. Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biology Letters* **2012**, *8*, 609–612. https://doi.org/10.1098/rsbl.2011.1136.

54. Daborn, P.; Yen, J.; Bogwitz, M.; Le Goff, G.; Feil, E.; Jeffers, S.; Tijet, N.; Perry, T.; Heckel, D.; Batterham, P.; et al. A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **2002**, *297*, 2253–2256.

55. Grant, P.R.; Grant, B.R. *How and why species multiply*; Princeton University Press, 2020.

56. Losos, J. *Lizards in an evolutionary tree*; University of California Press, 2009.

57. Roemhild, R.; Schulenburg, H. Evolutionary ecology meets the antibiotic crisis: Can we control pathogen adaptation through sequential therapy? *Evolution, Medicine, and Public Health* **2019**, *2019*, 37–45, [https://academic.oup.com/emph/article-pdf/2019/1/37/28096006/eoz008.pdf]. https://doi.org/10.1093/emph/eoz008.

58. Lehtinen, S.; Blanquart, F.; Lipsitch, M.; and, C.F. On the evolutionary ecology of multidrug resistance in bacteria. *PLOS Pathogens* **2019**, *15*, e1007763. https://doi.org/10.1371/journal.ppat.1007763.

59. Hawkins, N.J.; Bass, C.; Dixon, A.; Neve, P. The evolutionary origins of pesticide resistance. *Biological Reviews* **2018**, *94*, 135–155. https://doi.org/10.1111/brv.12440.

60. Fan, J.Y.; Hamza, K.; Jagers, P.; Klebaner, F.C. Limit theorems for multi-type general branching processes with population dependence. *Advances in Applied Probability* **2020**, *52*, 1127–1163. https://doi.org/10.1017/apr.2020.35.

61. Yakovlev, A.Y.; Yanev, N.M. Limiting distributions for multitype branching processes. *Stochastic analysis and applications* **2010**, *28*, 1040–1060.

62. Mode, C. *Multitype Branching Processes: Theory and Applications*; Modern analytic and computational methods in science and mathematics, American Elsevier Publishing Company, 1971.

63. Athreya, K.B.; Ney, P.E. *Branching Processes*; Springer Berlin Heidelberg, 1972. https://doi.org/10.1007/978-3-642-65371-1.

64. Kersting, G. A unifying approach to branching processes in a varying environment. *Journal of Applied Probability* **2020**, *57*, 196–220. https://doi.org/10.1017/jpr.2019.84.

65. Vatutin, V.A.; Dyakonova, E.E. Multitype branching processes in random environment. *Russian Mathematical Surveys* **2021**, *76*, 1019–1063. https://doi.org/10.1070/rm10012.

66. Árnason, E.; Koskela, J.; Halldórsdóttir, K.; Eldon, B. Sweepstakes reproductive success via pervasive and recurrent selective sweeps **2022**. https://doi.org/10.1101/2022.05.29.493887.

67. Haldane, J.B.S. The cost of natural selection. *Journal of Genetics* **1957**, *55*, 511–524. https://doi.org/10.1007/bf02984069.