The Fisher-Wright model with deterministic seed bank and selection

Bendix Koopmann
Center for Mathematics, Technische Universität München, 85748 Garching, Germany

Johannes Müller
Center for Mathematics, Technische Universität München, 85748 Garching, Germany and Institute for Computational Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany

Aurélien Tellier and Daniel Živković
Section of Population Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, 85354 Freising, Germany

Seed banks are a common characteristics to many plant species, which allow storage of genetic diversity in the soil as dormant seeds for various periods of time. We investigate an above-ground population following a Fisher-Wright model with selection coupled with a deterministic seed bank assuming the length of the seed bank is kept constant and the number of seeds is large. To assess the combined impact of seed banks and selection on genetic diversity, we derive a general diffusion model. We compute the equilibrium solution of the site-frequency spectrum and derive the times to fixation of an allele with and without selection. Finally, it is demonstrated that seed banks enhance the effect of selection onto the site-frequency spectrum while slowing down the time until the mutation-selection equilibrium is reached.

INTRODUCTION

Dormancy of reproductive structures, that is seeds or eggs, is described as a bet-hedging strategy [8, 11] in plants [12, 18, 30], invertebrates, *e.g.*, Daphnia [9], and microorganisms [25] to buffer against environmental variability. Bet-hedging is widely defined as an evolutionary stable strategy in which adults release their offspring into several different environments, here specifically with dormancy at different generations in time, to maximize the chance of survival and reproductive success, thus magnifying the evolutionary effect of good years and dampening the effect of bad years [8, 11]. Dormancy and quiescence sometimes have surprising and counter-intuitive consequences, similar to diffusion in activator-inhibitor models [16]. In the following study, we focus more specifically on the evolution of dormancy in plant species [12, 18, 30], but the theoretical models also apply to microorganisms and invertebrate species [9, 25].

Seed banking is a specific life-history characteristic of most plant species, which produce seeds remaining in the soil for short to long periods of time (up to several generations), and it has large but yet underappreciated consequences [11] for the evolution and conservation of many plant species.

First, polymorphism and genetic diversity are increased in a plant population with seed banks compared to the situation without banks. This is mostly due to storage of genetic diversity in the soil [19, 26]. Seed banks also damp off the variation in population sizes over time [26]. Under unfavourable conditions at generation t, the small offspring production is compensated at the next generation $t+1$ by individuals from the bank germinating at a given rate. Under the assumption of large seed banks, the observed population sizes between consecutive generations (t and $t+1$) may then be uncoupled.

Second, seed banks may counteract habitat fragmentation by buffering against the extinction of small and isolated populations, a phenomenon known as the "temporal rescue effect" [7]. Populations which suffer dramatically from events of decrease in population size can be rescued by seeds from the bank. Improving our understanding of the evolutionary conditions for the existence of long-term dormancy and its genetic underpinnings is thus important for the conservation of endangered plant species in habitats under destruction by human activities.

Third, germ banks influence the rate of natural selection in populations. On the one hand, seed banks promote the occurrence of balancing selection for example for color morphs in *Linanthus parryae* [31] or in host-parasite co-evolution [27]. On the other hand, the storage effect is expected to decrease the efficiency of positive selection in populations, thus natural selection, positive or negative, would be slowed down by the presence of long-term seed banks. Empirical evidence for this phenomenon has been shown [17], but no quantitative model exists so far. In general terms, understanding how seed banks evolve, affect the speed of adaptive response to environmental changes, and determine the rate of population extinction in many plant species is of importance for conservation genetics under the current period of anthropologically driven climate change.

Two classes of theoretical models have been developed for studying the influence of seed banks on genetic variability. First, Kaj *et al.* [19] have proposed a backward in time coalescent seed bank model which includes the probability of a seed to germinate after a number of years

in the soil and a maximum amount of time that seeds can spend in the bank. Seed banks have the property to enhance the size of the coalescent tree of a sample of chromosomes from the above ground population by a quadratic factor of the average time that seeds spend in the bank. This leads to a rescaling of the Kingman coalescent [23] because two lineages can only coalesce in the above-ground population in a given ancestral plant. The consequence of longer seed banks with smaller values of the germination rate is thus to increase the effective size of populations and genetic diversity [19] and to reduce the differentiation among populations connected by migration [32]. This rescaling effect on the coalescence of lineages in a population has also important consequences for the statistical inference of past demographic events [35]. In practice this means that the spatial structure of populations and seed bank effects on demography and selection are difficult to disentangle [5]. Nevertheless, Tellier *et al.* [28] could use this rescaled seed bank coalescent model [19] and Approximate Bayesian Computation to infer the germination rate in two wild tomato species *Solanum chilense* and *S. peruvianum* from polymorphism data [29].

A second class of models assumes a strong seed bank effect, whereby the time seeds can spend in the bank is very long, that is longer than the population coalescent time [14], or the time for two lineages to coalesce can be unbounded. This latest model generates a seed bank coalescent [2], which may not come down from infinity and for which the expected site-frequency spectrum (SFS) may differ significantly from that of the Kingman coalescent [4]. In effect, the model of [19] represents a special case, also called a weak seed bank, where the time for lineages to coalesce is finite because the maximum time that seeds can spend in the bank is bounded.

In the following we mainly have the weak seed bank model in mind where the time in the seed bank is bounded to a small finite number assumed to be realistic for most plant species [12, 18, 29, 30]. Even if we allow for unbounded times a seed may be stored within the soil, we assume that the germination probability decreases rapidly with age such that e.g. the expected time a seed rests in the soil is finite. We develop a forward in time diffusion for seed banks following a Fisher-Wright model with random genetic drift and selection acting on one of two genotypes. The time rescaling induced by the seed bank is shown to be equivalent for the Fisher-Wright and the Moran model. We provide the first theoretical estimates of the effect of seed bank on natural selection by deriving the expected SFS of alleles observed in a sample of chromosomes and the time to fixation of an allele.

The main difficulty in the present paper is the non-Markovian character of seedbank models (with the exception of a geometric survival distribution for seeds, in which case the model can be reduced to a Markovian model, see below). The way to deal with this non-Markovian character is based on a separation of time scales. The genetic composition of the population only changes on a slow, so-called evolutionary time scale (thousands of generations), while being fairly stable on a fast, ecological time scale (tens of generations). We assume seeds to have a life span corresponding to this ecological time scale, and thus the seedbank tends to a quasi-stationary state. The non-Markovian character of the model is visible at the ecological time scale, while it vanishes on the evolutionary time-scale due to the quasi-steady-state assumption. In other words we ensure the separation of time scales by assuming that most seeds die after a few generations. We demonstrate thereafter that seed banks affect selection and genetic drift differently.

MODEL DESCRIPTION

We consider a finite plant-population of size N. The plants appear in two genotypes A and a. We assume non-overlapping generations. Let X_n denote the number of type-A plants in generation n (that is, the number of living type-a plants in this generation is $N - X_n$). Plants produce seeds. The number of seeds is assumed to be large, such that noise in the seed bank does not play a role (therefore we call the seed bank "deterministic"). The amount of seeds produced by type-A-plants in generation n is $\beta_A X_n$, that of type-a plants $\beta_a(N - X_n)$. The seeds are stored *e.g.* in the soil; some germinate in the next generation, some only in later generations, and some never.

To obtain the next generation of living plants X_n, we need to know which seeds are likely to germinate. Let $b_A(i)$ be the fraction of type-A seeds of age i able to germinate, and $b_a(i)$ that of type-a seeds. Hence, the total amount of type-A seeds that is able to germinate is given by

$$\sum_{i=1}^{\infty} b_A(i)\beta_A X_{n-i},$$

and accordingly, the total amount of all seeds that may germinate

$$\sum_{i=1}^{\infty} b_A(i)\beta_A X_{n-i} + \sum_{i=1}^{\infty} b_a(i)\beta_a(N - X_{n-i}).$$

The probability that a plant in generation n is of phenotype A is given by the fraction of type-A seeds that may germinate among all seeds that are able to germinate. The frequency process of the di-allelic Fisher-Wright model with deterministic seed bank reads

$$X_n \sim \mathrm{Bin}(N, q_n(X_\bullet)), \tag{1}$$

$$q_n(X_\bullet) = \frac{\sum_{i=1}^{\infty} b_A(i)\beta_A X_{n-i}}{\sum_{i=1}^{\infty} b_A(i)\beta_A X_{n-i} + \sum_{i=1}^{\infty} b_a(i)\beta_a(N - X_{n-i})}.$$

Next we introduce (weak) selection. The fertility of type-a plants is given by

$$\beta_a = (1 - s_1)\, \beta_A,$$

such that $s_1 = 0$ corresponds to the neutral case. Furthermore, the fraction of surviving seeds is affected. We relate $b_a(i)$ to $b_A(i)$ by

$$b_a(i) = (1 - s_2)\, b_A(i).$$

Of course, s_2 has to be small enough to ensure that $b_a(i) \in [0, 1]$. There are other ways to incorporate a fitness difference in the surviving probabilities of seeds, but we feel that this is the most simple version. If we lump s_1 and s_2 in one parameter that scales in an appropriate way for selection,

$$(1 - s_1)\,(1 - s_2) = 1 - \sigma/N,$$

(the sign is chosen in such a way that genotype A has an advantage over genotype a for $\sigma > 0$ and a disadvantage if $\sigma < 0$) then eqn. (1) for $q_n(X_\bullet)$ with selection becomes

$$\frac{\sum_{i=1}^{\infty} b_A(i) X_{n-i}}{\sum_{i=1}^{\infty} b_A(i) X_{n-i} + (1 - \sigma/N) \sum_{i=1}^{\infty} b_A(i)(N - X_{n-i})}.$$

As this ratio is homogeneous of degree zero in b_A, we assume $\sum_{i=1}^{\infty} b_A(i) = 1$. That is, $b_A(i)$ is considered a probability distribution for the survival of a (type-A) seed. We assume that the average life time of a seed is finite, $B = \sum_{i=1}^{\infty} i b_A(i) < \infty$. We will implicitly assume that $b_A(i)$ converge fast enough to zero, such that the separation of ecological and evolutionary time scale is still true. The sum $\sum_{i=1}^{\infty} b_A(i) X_{n-i}$ is a moving average. We emphasize this fact by introducing the operator

$$M_n(X_\bullet) = \sum_{i=1}^{\infty} b_A(i) X_{n-i}.$$

As a consequence, we have $M_n(N) = N$, and

$$\begin{aligned} q_n(X_\bullet) &= \frac{M_n(X_\bullet)}{M_n(X_\bullet) + (1 - \sigma/N)(N - M_n(X_\bullet))} \\ &= \frac{M_n(X_\bullet)}{N - \sigma/N\,(N - M_n(X_\bullet))}. \end{aligned} \tag{2}$$

DIFFUSION LIMIT – GEOMETRIC CASE

As indicated above, if $b_A(i)$ follow a geometric distribution, then the non-Markovian model introduced above can be reduced to a Markovian model: it is not necessary to track the age of a seed, as all seeds independent of their age have the same mortality resp. germination probability. In this case, and without selection ($\sigma = 0$), it is straight forward to obtain a diffusion limit that describes the model well on the evolutionary time scale if the population size is (finite but) large. In particular, the diffusion limit is the diffusive Moran model, where we already obtain a first indication how the scaling is affected by the seedbank. Note that the backward process has been analyzed in [3]. This neutral case with a geometric germination rate serves as a warm-up before investigating the full model.

The Fisher-Wright model without selection

We recall briefly the procedure to derive the diffusion limit for the standard Fisher-Wright model (without seed bank).

• *Model:* $X_{n+1} \sim \mathrm{Bin}(N, X_n/N)$.

• *Rescale population size:* Let $x_n = X_n/N$. Then, $X_{n+1} \sim \mathrm{Bin}(N, x_n)$. For N large, the Binomial distribution approximates a normal distribution with expectation $x_n\, N$ and variance $x_n(1 - x_n)N$. Let η_n be i.i.d. $N(0, 1)$-random variables. Then,

$$\begin{aligned} x_{n+1} = X_{n+1}/N &\approx \left(x_n\, N + (x_n(1 - x_n))^{1/2} N^{1/2} \eta_n \right) / N \\ &= x_n + N^{-1/2}\, (x_n(1 - x_n))^{1/2}\, \eta_n. \end{aligned}$$

• *Rescale time:* Now define $\Delta\tau = 1/N$, introduce the time $\tau = n\Delta\tau$, let $u_{n\Delta\tau} = x_n$, and rescale the index of the normal random variables, that is, replace η_n by $\eta_{n\Delta\tau} = \eta_\tau$. Then, $u_{\tau+\Delta\tau} - u_\tau = \Delta\tau^{1/2}\, (u_\tau(1 - u_\tau))^{1/2}\, \eta_\tau$. According to the Euler-Maruyama formula (see *e.g.* [24]), we approximate the diffusive Moran model for N large (that is, $\Delta\tau = 1/N$ small)

$$du_\tau = (u_\tau(1 - u_\tau))^{1/2}\, dW_\tau.$$

where W_t indicates the Brownian motion.

Seed bank model with a geometric germination rate and without selection

In the present section we assume that there is no selection ($\sigma = 0$), and $b(i)$ follow a geometric distribution with parameter $\mu \in (0, 1)$, $b(1) = \mu$ and $b(i) = (1 - \mu) b(i - 1)$. In this case, the delay-model is equivalent to a proper Markov chain.

• *Reformulation of the model:* Define $z_n = M_{n+1}(X_\bullet)/N = \mu \sum_{i=1}^{\infty} (1 - \mu)^{i-1} X_{n+1-i}/N$. We immediately obtain

$$\begin{aligned} z_{n+1} &= \mu \sum_{i=1}^{\infty} (1 - \mu)^{i-1} X_{n+2-i}/N \\ &= \mu\, X_{n+1}/N + \mu \sum_{i=2}^{\infty} (1 - \mu)^{i-1} X_{n+1-(i-1)}/N \\ &= \mu\, X_{n+1}/N + (1 - \mu)\, z_n. \end{aligned}$$

Next (and with the nomenclature of (2)), we have $q_{n+1}(X_\bullet) = M_{n+1}(X_\bullet/N) = z_n$. All in all, we reformulated model (1) in the present situation as

$$\begin{aligned} X_{n+1} &\sim \mathrm{Bin}(N, z_n), \\ z_{n+1} &= \mu\, X_{n+1}/N + (1-\mu)\, z_n. \end{aligned} \tag{3}$$

Note that z_n can be interpreted as the state of the seed bank (the fraction of type-A seeds that are able to germinate).

• *Rescale population size:* As this model is Markovian, it is simple to derive the diffusion limit. As usual, we start off by defining $x_n = X_n/N$, and obtain $z_n = \mu\, x_n + (1-\mu)\, z_{n-1}$, $X_{n+1} = \mathrm{Bin}(N, z_n)$. Approximating the Binomial distribution by a normal distribution for N large yields

$$x_{n+1} \approx z_n + N^{-1/2}(z_n(1-z_n))^{1/2}\eta_n,$$

where the $\eta_n \sim N(0,1)$ i.i.d.. As x_{n+1} can be expressed by z_n and z_{n+1}, the foregoing two equations give

$$\frac{z_{n+1} - (1-\mu)\, z_n}{\mu} = z_n + N^{-1/2}(z_n(1-z_n))^{1/2}\eta_n.$$

Therefore, $z_{n+1} - z_n = \mu\, N^{-1/2}\, (z_n(1-z_n))^{1/2}\, \eta_n$.

• *Rescale time:* Scaling time by N yields for $u_{n/N} = z_n$ and $\tau = n/N$

$$du_\tau = \mu\, (u_\tau\, (1-u_\tau))^{1/2} dW_\tau.$$

If we define $B = 1/\mu$ (the expected value of a geometric distribution with parameter μ), we may write this equation as

$$du_\tau = \frac{(u_\tau\, (1-u_\tau))^{1/2}}{B}\, dW_\tau. \tag{4}$$

We find a diffusive Moran model for the state of the seed bank with rescaled time scale. The factor $1/B$ has been already proposed in the paper of Kaj, Krone and Lascoux [19], who analyzed a seedbank process backward in time.

DIFFUSION LIMIT – GENERAL CASE

We expect a similar result as above to hold in the general case. A difference between the two cases is that we naturally considered the state of the seed bank before, while in the general case we will focus on the state of living plants. As discussed before, the center of the analysis below is an additional step that investigates the quasi-stationary state of the seedbank at evolutionary time scale; this additional step is necessary to deal with the non-Markovian character of our model.

Rescale population size

From (2), we immediately have

$$q_n(x_\bullet) = q_n(X_\bullet/N) = \frac{M_n(x_\bullet)}{1 - \Delta t\, \sigma(1 - M_n(x_\bullet))}.$$

Using Normal approximation of the Binomial distribution leads to

$$x_n \approx q_n(x_\bullet) + \Delta t^{1/2} \sqrt{q_n(x_\bullet)\, (1 - q_n(x_\bullet))}\, \eta_n$$

where $\eta_n \sim N(0,1)$ are i.i.d.. Taylor expansion of $q_n(x_\bullet)$ w.r.t. Δt yields in lowest order

$$\begin{aligned} &x_n - M_n(x_\bullet) - \Delta t\, \sigma\, f(M_n(x_\bullet)) \\ &= \Delta t^{1/2} f^{1/2}(M_n(x_\bullet))\, \eta_n. \end{aligned} \tag{5}$$

with $f(x) = x(1-x)$.

Perturbation approach

The leading term of eqn. (5) is $x_n - M_n(x_\bullet)$. This difference must not become too large, as all other terms in the equation are at least of order $\Delta t^{1/2}$. That is, the state x_n can only slowly drift away from $M_n(x_\bullet)$ (which represents the state of the seed bank). Hence, for a reasonable number of time steps (on the ecological time scale), $M_n(x_\bullet)$ is fairly constant. In order to disentangle the evolutionary and the ecological time scale, we introduce $\varepsilon = \Delta t^{1/2}$, expand x_n w.r.t. ε,

$$x_n = x_n^{(0)} + \varepsilon x_n^{(1)} + \varepsilon^2 x_n^{(2)} + \dots$$

and rewrite eqn. (5) as

$$x_n - M_n(x_\bullet) = \varepsilon^2\, \sigma\, f(M_n(x_\bullet)) + \varepsilon\, f^{1/2}(M_n(x_\bullet))\eta_n.$$

Taylor expansion and equating equal powers of ε yields

$$x_n^{(0)} - M_n(x_\bullet^{(0)}) = 0 \tag{6}$$

$$x_n^{(1)} - M_n(x_\bullet^{(1)}) = f^{1/2}(M_n(x_\bullet^{(0)}))\eta_n \tag{7}$$

$$\begin{aligned} x_n^{(2)} - M_n(x_\bullet^{(2)}) &= \sigma(f(M_n(x_\bullet^{(0)})) \\ &\quad + \frac{1}{2} f^{-1/2}(M_n(x_\bullet^{(0)}))\, f'(M_n(x_\bullet^{(0)}))\, M_n(x_\bullet^{(1)})\, \eta_n. \end{aligned} \tag{8}$$

Zero order: The zero order term $x_n^{(0)}$ follows a deterministic dynamics. As M_n is an averaging operator the solution becomes constant in the long run. The system saddles on the slow manifold, consisting of constant sequences. At this point it is important that $b_A(i)$ tend fast enough to zero, s.t. $x_n^{(0)}$ indeed approximates on the fast (ecological) time scale the slow manifold. We assume $x_n^{(0)} \equiv \overline{x}^0$.

First order: The recursive equation (7) is well known as an auto-regression (AR) model in the statistical modeling of time series [6]. We define $\beta = f^{1/2}(M_n(x_\bullet^{(0)})) =$

$f^{1/2}(\overline{x}^0)$ (note that β is a real number and not a random variable) and convert the AR model into a moving average equation. Thereto we introduce the back-shift operator acting on the index of a sequence, $Lz_n = z_{n-1}$, and the power series

$$\psi(x) = 1 - \sum_{i=1}^{\infty} b_A(i) x^i;$$

Eqn. (7) becomes in this notation

$$\psi(L) x_n^{(1)} = x_n^{(1)} - M_n(z_\bullet) = \beta\, \eta_n.$$

Note that $\psi(1) = 0$, which does mean that the AR model is non-stationary (this process is also called an ARIMA model for time series [6, Chapter 9]). We do not find a power series $\psi^*(x)$ well defined at $x = 1$ such that $\psi^*(x)\,\psi(x) = 1$. Therefore, we rewrite $\psi(x)$ as $\psi(x) = (1-x)\,\tilde{\psi}(x)$ (which is the defining equation of $\tilde{\psi}(x)$). As

$$\tilde{\psi}(1) = \lim_{x\to 1} \frac{\psi(x)}{(1-x)} = -\psi'(1) = \sum_{i=1}^{\infty} b_A(i)\, i = B \neq 0,$$

we do find $\psi^*(x)$ such that $\psi^*(x)\tilde{\psi}(x) = 1$, and hence $\psi^*(x)\psi(x) = 1 - x$ in a neighbourhood of $x = 1$. As an immediate consequence (used later) we have $\psi^*(1) = 1/B$. If we multiply the equation $\psi(L)x_n^{(1)} = \beta\,\eta_n$ by $\psi^*(L)$, we obtain

$$x_n^{(1)} - x_{n-1}^{(1)} = (1-L)x_n^{(1)} = \beta\psi^*(L)\eta_n$$

and

$$\begin{aligned} x_n^{(1)} &= x_{n-1}^{(1)} + \beta\,\psi^*(L)\eta_n \\ &= x_{n-2}^{(1)} + \beta\,\psi^*(L)\eta_n + \Delta t^{1/2}\,\beta\,\psi^*(L)\eta_{n-1} = \cdots \\ &\approx \beta \sum_{\ell=0}^{n} \psi^*(L)\eta_{n-\ell}. \end{aligned}$$

Let $\psi^*(z) = \sum_{i=0}^{\infty} a_i z^i$. We expand the sum above, and obtain

$$\begin{array}{rcllllll} \sum_{\ell=0}^{n} \psi^*(L)\eta_{n-\ell} & = & a_0\eta_n & +a_1\eta_{n-1} & +a_2\eta_{n-2} & +a_3\eta_{n-3} & +a_4\eta_{n-4} & +a_5\eta_{n-5} & +\cdots \\ & & & +a_0\eta_{n-1} & +a_1\eta_{n-2} & +a_2\eta_{n-3} & +a_3\eta_{n-4} & +a_4\eta_{n-5} & +\cdots \\ & & & & +a_0\eta_{n-2} & +a_1\eta_{n-3} & +a_2\eta_{n-4} & +a_3\eta_{n-5} & +\cdots \\ & & & & & +a_0\eta_{n-3} & +a_1\eta_{n-4} & +a_2\eta_{n-5} & +\cdots \\ & & & & & & +\cdots & +\cdots & +\cdots \end{array}$$

If we inspect not rows (that have $\psi^*(L)\eta_{i-\ell}$ as entries) but columns (that contain always the same random variable $\eta_{i-\ell}$), we find that the coefficient in front of one given random variable $\eta_{i-\ell}$ approximates $\psi^*(1)$ for $\ell \to \infty$.
At this point, we want to write $x_{n+1}^{(1)} \approx \beta\,\psi^*(1) \sum_{\ell=1}^{n} \eta_\ell$. This is only true, also in an approximate sense, if n is large and the state x_n does hardly change over a time scale that allows $\sum_{i=1}^{m} a_i$ to converge to $\psi^*(1) = 1/B$. If $\Delta t^{1/2}$ is small, then indeed $x_n \approx \overline{x}_0$ on the ecological time scale, as required. Hence, for Δt small we are allowed to assume

$$x_{n+1}^{(1)} \approx \beta\,\psi^*(1) \sum_{\ell=1}^{n} \eta_\ell = \frac{\beta}{B} \sum_{\ell=1}^{n} \eta_\ell.$$

Thus, $x_{n+1}^{(1)} \approx (\beta/B) \sum_{\ell=1}^{n} \eta_\ell$, and for n large

$$x_{n+1}^{(1)} - x_n^{(1)} = (\beta/B)\,\eta_n \tag{9}$$

where, as before, $\eta_n \sim N(0,1)$ i.i.d..
Second order: With $\alpha = f(\overline{x}^0)$, $\tilde{\beta} = \frac{1}{2} f^{-1/2}(\overline{x}^0)\, f'(\overline{x}^0))$, we may write

$$x_n^{(2)} + M_n(x_\bullet^{(2)}) = \alpha + \tilde{\beta}\, M_n(x_\bullet^{(1)})\eta_{n-1}. \tag{10}$$

If $\alpha \neq 0$, $x_n^{(2)}$ incorporates a deterministic trend. We first remove this trend defining $z_n = x_n^{(2)} - w_n$ with $w_n = n\,\alpha/B$. Then, $M_n(w_\bullet) = \sum_{i=1}^{\infty} b_A(i)\,(n-i)\,\alpha/B = n\,\alpha/B - \alpha$, and, with $M_n(x_\bullet^{(2)}) = M_n(z_\bullet) + M_n(w_\bullet)$,

$$\begin{aligned} & z_n + n\,\frac{\alpha}{B} - \Big(M_n(z_\bullet) + M_n(w_\bullet) \Big) - \alpha \\ & = z_n - M_n(z_\bullet) + n\,\frac{\alpha}{B} - \alpha - \Big(n\,\frac{\alpha}{B} - \alpha \Big) = \tilde{\beta}\, M_n(x_\bullet^{(1)})\eta_n. \end{aligned}$$

We obtain an AR model for z_n without trend,

$$z_n - M_n(z_\bullet) = \tilde{\beta}\, M_n(x_\bullet^{(1)})\eta_n. \tag{11}$$

It turns out, that we need not to analyze z_n in detail. It is sufficient to note that z_n is a random variable with expectation zero.

Result: All in all, we conclude

$$x_{n+1} - x_n = \varepsilon^2\,\frac{\alpha}{B} + \varepsilon\,\frac{\beta}{B}\,\eta_n + \varepsilon^2 z_n.$$

We only take into account the lowest order in the deterministic drift resp. in the random perturbations. As

$\varepsilon\,(\beta/B)\,\eta_n$ dominates $\varepsilon^2\,z_n$, we drop the latter term, replace in α, β the variable $\overline{x}^0$ by x_n, and end up with

$$x_n = x_{n-1} + \Delta t \frac{\sigma}{B} x_n(1-x_n) + \Delta t^{1/2} \frac{1}{B}\,\sqrt{x_n(1-x_n)}\,\eta_n. \quad (12)$$

Numerical simulation: We compare the result of these computations with numerical simulations. Thereto we consider the linear model

$$y_n - M_n(y_\bullet) = \Delta t\, a \,+\, \Delta t^{1/2}\, b\, \eta_n$$

with $a, b \in \mathbb{R}$. If $y_n = 0$ for $n \leq 0$, we expect that y_n (for $n \geq 1$) approximately to satisfy

$$y_n - y_{n-1} = \Delta t \frac{a}{B} \,+\, \Delta t^{1/2}\, \frac{b}{B}\, \eta_n.$$

That is, y_n is approximately normally distributed with expectation $n\,\Delta t\, a/B$, and variance $n\,\Delta t\, b^2/B^2$. For simulations, we choose $a = 1$, $\Delta t = 0.01$, $b = 2$ and $M_n(y_\bullet) = \frac{1}{m}\sum_{i=1}^m y_{n-i}$ for $m = 9$, that is, $B = 5$. The simulations show an excellent agreement with our computations (Figure 1).

Rescale time

As before, we define $u_{n\,\Delta t} = x_n$, and use the Euler-Maruyama-formula to conclude that u_t approximates for $\Delta t \to 0$ the stochastic differential equation

$$du_t = \frac{\sigma}{B}\, u_t(1-u_t)dt + \frac{1}{B}\left(u_t(1-u_t)\right)^{1/2} dW_t. \quad (13)$$

Please note that this result seems to inherit the usual stability of a diffusion limit w.r.t. the detailed model assumptions: if we start off with a Moran model instead of a Fisher-Wright model combined with a seed bank, we again obtain a diffusion limit of similar form (see Appendix).

We now change the time scale such that the variance coincides with the standard diffusive Moran model. If we define $\tau = t/B^2$, then the SDE reads

$$du_\tau = (\sigma B)\, u_\tau(1-u_\tau)d\tau + \sqrt{u_\tau(1-u_\tau)}\, dW_\tau. \quad (14)$$

Scaling of the selection parameter. We conclude, in line with previous findings (see discussion), that the appropriate scaling of time for the Fisher-Wright model with seed bank is not $1/N$ but $1/(B^2\,N)$. Moreover, the effective selection rate (w.r.t. this time) is increased by the average number of generations B the seeds sleep in the soil.

THE FORWARD DIFFUSION EQUATION FOR SEED BANK MODELS WITH SELECTION

In analogy to above, we consider a single locus and two allelic types A and a with frequencies x and $1-x$, respectively, at time zero. Time is scaled in units of $2N$ generations. In the diffusion limit, as $N \to \infty$, the probability $f(y,t)dy$ that the type-A genotype has a frequency in $(y, y+dy)$ is characterized by the following forward equation (see [20] for $B = 1$):

$$\frac{\partial}{\partial t} f(y,t) = -\frac{\partial}{\partial y}\left(a(y)\, f(y,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(y)\, f(y,t)\right),$$

where the drift and the diffusion terms are given by $a(y) = \sigma\; y(1-y)/B$ and $b(y) = y(1-y)/B^2$, respectively.

For the derivations of the frequency spectrum and the times to fixation we require the following definitions. The scale density of the diffusion process is given by

$$\xi(y) = \exp\left(-\int_0^y \frac{2a(z)}{b(z)} dz\right) = \exp\left(-2B\sigma y\right).$$

The speed density is obtained (up to a constant) as

$$\pi(y) = \left[b(y)\xi(y)\right]^{-1} = \frac{B^2 \exp\left(2B\sigma y\right)}{y(1-y)}.$$

The probability of absorption at $y = 0$ is given by

$$u_0(x) = \frac{\int_x^1 \xi(z)dz}{\int_0^1 \xi(z)dz} = \frac{\exp(2B\sigma(1-x)) - 1}{\exp(2B\sigma) - 1},$$

and $u_1(x) = 1 - u_0(x)$ gives the probability of absorption at $y = 1$.

Site-frequency spectra

The site-frequency spectrum (SFS) of a sample (*e.g.*, [10, 15, 34]) is widely used for population genetics data analysis. A sample of size k is sequenced, and for each polymorphic site the number of individuals in which the mutation appears is determined. In this way, a dataset is generated that summarizes the number of mutations $\zeta_{k,i}$ appearing in i individuals, $i = 1, \ldots, k-1$. That is, $\zeta_{k,1} = 10$ indicates that 10 mutations only appeared once, and $\zeta_{k,2} = 5$ tells us that five mutations were present in two individuals (where the pair of individuals may be different for each of the five mutations). Note that neither $\zeta_{k,0}$ nor $\zeta_{k,k}$ are sensible: a mutation that appears in none or all individuals of the sample cannot be recognized as a mutation. In practice, it is often not possible to know the ancestral state. Then the folded SFS $\eta_{k,i} = (\zeta_{k,i} + \zeta_{k,k-i})(1 + 1_{\{i=k-i\}})^{-1}$ can be used. Since both empirical observations and theoretical results for

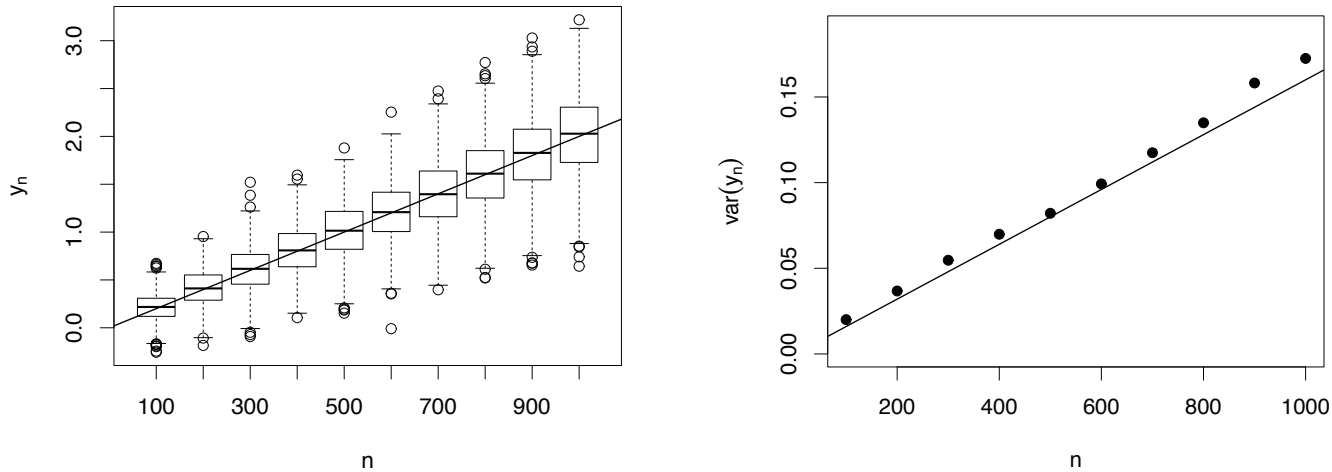

FIG. 1. Simulation of the AR model (1000 runs). Samples have been taken at time steps 100, 200,..., 1000. (left) Boxplot of the simulated time series y_n at indicated time points together with the mean according to eqn. 10 (line). (right) Variance of the simulated time series at indicated time points (dots), together with the variance according to eqn. 10 (line). For parameters used: see text.

the folded SFS follow instantaneously from the unfolded one, we only consider the unfolded version.

For the derivation of the theoretical SFS, we assume that mutations occur according to the infinitely-many sites model [21]. The scaled mutation rate is given by $\theta = 4N\,\nu$, where ν is the mutation rate per generation at independent sites. Assuming that each mutant allele marginally follows the diffusion model specified above, the proportion of sites where the mutant frequency is in $(y, y+dy)$ is given by [15]

$$\hat{f}(y) = \theta\,\pi(y)\,u_0(y) = \frac{\theta B^2}{y(1-y)}\,\frac{\exp(2B\sigma)-\exp(2B\sigma y)}{\exp(2B\sigma)-1} = \frac{\theta B^2}{y(1-y)}\,\frac{1-\exp(-2B\sigma(1-y))}{1-\exp(-2B\sigma)},$$

where $\hat{f}(y)$ denotes the equilibrium solution of the population SFS. For neutrality, we immediately obtain $\hat{f}(y) = \theta\,B^2/y$ by letting $\sigma \to 0$ in the foregoing equation.

The equilibrium solution of the SFS for a sample of size k is obtained via binomial sampling (see [33] for $B = 1$) as

$$\hat{f}_{k,i} = \binom{k}{i}\int_0^1 \hat{f}(y)y^i(1-y)^{k-i}\,dy = \theta B^2\frac{k}{i(k-i)}\,\frac{1-{}_1F_1(i;k;2B\sigma)e^{-2B\sigma}}{1-e^{-2B\sigma}},$$

where ${}_1F_1$ denotes the confluent hypergeometric function of the first kind [1]. For neutrality, we again immediately obtain $\hat{f}_{k,i} = \theta\,B^2/i$ by letting $\sigma \to 0$. For a large number of mutant sites, the relative SFS $\hat{r}_{k,i} = \hat{f}_{k,i}/\sum_{j=1}^{k-1}\hat{f}_{k,j}$ approximates the empirical distribution $\zeta_{k,i}/\sum_{j=1}^{k-1}\zeta_{k,j}$ for a constant population size. Note that the solutions for the absolute SFS assume that mutations can occur at any time. When assuming that mutations can only arise in living plants [19], θ has to be replaced by θ/B in the respective equations. Both mutation models give equivalent results for the relative SFS.

As shown in Figure 2 (left), the neutral diffusion approximation is in line with the simulation results of the original discrete model. The theoretical relative SFS for a sample of 250 individuals approximates the simulated SFS, which is obtained as an average over 10,093 repetitions. In every iteration, the sample is drawn from an initially monomorphic population of 1000 individuals after 400,000 generations (so that the population has reached an equilibrium). Figure 2 (right) illustrates the enhanced effect of selection proportional to the length of the seed bank.

Times to fixation

We assume that both $y = 0$ and $y = 1$ are absorbing states and start by considering the mean time until one of these states is reached in the diffusion process specified above. The mean absorption time $\bar{t}$ can be expressed as [13]

$$\bar{t}(x) = \int_0^1 t(x,y)dy, \tag{15}$$

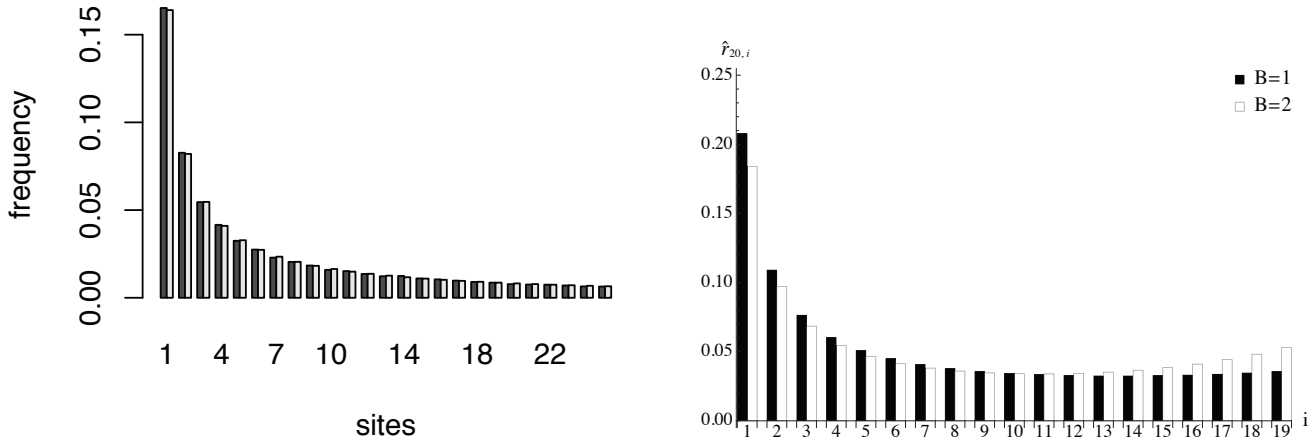

FIG. 2. (left) Simulation and theoretical prediction for the neutral relative SFS and a uniformly distributed seed bank of length $B = 10$. For the simulation of the original discrete model the population size was chosen as 1000, we started without mutations and stopped the process after 400,000 generations to calculate the SFS as an average over 10,093 repetitions. The light gray bar shows the theoretical result, the dark gray bar shows the simulation outcome. In both cases a sample of 250 individuals was drawn. (right) Theoretical results for the relative SFS of a sample of size 20 are plotted for positive selection of strength $\sigma = 2$ without ($B = 1$) and with a seed bank of length $B = 2$.

where

$$t(x,y) = 2\,u_0(x)[b(y)\xi(y)]^{-1}\int_0^y \xi(z)dz, \quad 0 \le y \le x,$$

$$t(x,y) = 2\,u_1(x)[b(y)\xi(y)]^{-1}\int_y^1 \xi(z)dz, \quad x \le y \le 1.$$

For genetic selection the integral in (15) cannot be analytically solved. For selective neutrality, we obtain $\bar{t}(x) = -2\,B^2\,(x\,\log(x) + (1-x)\,\log(1-x))$ (see *e.g.* [13] for $B = 1$) by employing the drift term, the scale density and the probabilities of absorption as specified above.
Now, we evaluate the time until a mutant allele is fixed conditional on fixation as $\bar{t^*}(x) = \int_0^1 t^*(x,y)dy$, where $t^*(x,y) = t(x,y)u_1(y)/u_1(x)$. For genic selection the mean time to fixation in dependency of x can only be derived as a very lengthy expression in terms of exponential integral functions. The neutral result is found as $\bar{t^*}(x) = -2B^2(1-x)/x\log(1-x)$ and in accordance with a classical result [22] for $B = 1$. For $x \to 0$, we obtain

$$\bar{t^*} = \frac{2\,B}{\sigma(e^{2\,B\,\sigma}-1)}\big((e^{2\,B\,\sigma}+1)\gamma - \mathrm{Ei}(2\,B\,\sigma) + \log(2\,B\,\sigma) + e^{2\,B\,\sigma}(-\mathrm{Ei}(-2\,B\,\sigma) + \log(2\,B\,\sigma))\big), \qquad \sigma > 0, \tag{16}$$

$$\bar{t^*} = 2\,B^2, \qquad \sigma = 0,$$

where γ is Euler's constant and Ei denotes the exponential integral function [1].
In Figure 3 (left), we compare the time to absorption of the original discrete seed bank model by means of simulations with the theoretical result obtained from the diffusion approximation. For b_A we use uniform distributions, where we vary the expected values between 1 and 8 corresponding to the length of the seed banks between 1 and 15. We choose an initial fraction of 0.5 for the type-A genotypes. The simulations show a good agreement between our analytical approximation and the numerical simulations. In Figure 3 (right), we show the effect of the seed bank on the times to fixation conditional on fixation of the type-A genotype for neutrality and positive selection.

DISCUSSION

Within this study, we develop a forward in time Fisher-Wright model of a deterministically large seed bank with drift occurring in the above-ground population. The time that seeds can spend in the bank is bounded and finite, as assumed to be realistic for many plant or invertebrate species. We demonstrate that scaling time in the diffusion process by a factor B^2 generates the usual Fisher-Wright time scale of genetic drift with B being defined as the average amount of time that seeds spend in the bank. The conditional time to fixation of a neutral allele is slowed down by a factor B^2 (Figure 3 (right), dotted line) compared to the absence of seed bank. These results are consistent with the backward in time coalescent model from Kaj *et al.* [19], and differs from the strong seed bank model of Blath *et al.* [2]. We evaluate the SFS based on our diffusion process and confirm agreement to the SFS obtained under discrete time Fisher-Wright simulations.

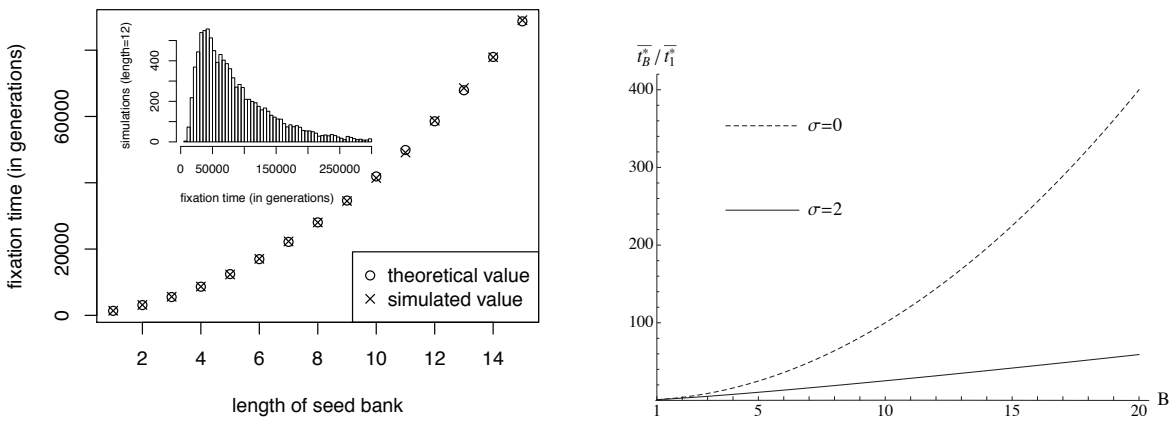

FIG. 3. (left) Simulation and theoretical prediction for the time to fixation of a seed bank model. The population size is 1000 and 50% of the individuals are initially of genotype A. We simulated 10,000 runs for each mean value. The simulated distribution of the time to fixation is shown in the histogram at the upper left corner taking the data of the simulated seed bank of length $B = 12$. (right) The ratios of the conditional fixation times with and without seedbank are plotted against the length of the seed bank B for neutrality and selection by employing (16). The additional index in the ratio is used to formally distinguish the cases with and without seed bank.

In the second part of the study, we introduce selection occurring at one of the two alleles, mimicking positive or negative selection. Two features of selection under seed banks are noticeable. First, selection is slower under longer seed banks (Figure 3 (right), solid line) confirming previous intuitive expectations [17]. Second, when computing the SFS with $B = 2$ and without seed bank ($B = 1$) under positive selection ($\sigma = 2$) we reveal a stronger signal of selection for the seed bank by means of an amplified uptick of high-frequency derived variants. This effect becomes more prominent with longer seed banks and also holds for purifying selection, under which an increase in low-frequency derived variants is induced by the seed bank. We explain this counterintuitive results as follows: longer seed banks increase, on the one hand, the selection coefficient σ generating a stronger signal at equilibrium (Figure 2 (right)), and on the other hand, the time to reach this equilibrium state (Figure 3 (right)). Our predictions are consistent with the inferred strengths of purifying selection in wild tomato species. Indeed, purifying selection at coding regions appears to be stronger in *S. peruvianum* than in its sister species *S. chilense* [28] with *S. peruvianum* exhibiting a longer seed bank [29].

This research is supported in part by Deutsche Forschungsgemeinschaft grants TE 809/1 (AT) and STE 325/14 from the Priority Program 1590 (DZ).

Appendix: Moran model with deterministic seed bank

We briefly sketch the arguments that allow to handle a Moran model with seed bank; the reasoning is completely parallel to the time-discrete case. In order to keep this appendix short, we do not take into account selection but focus on the neutral model.

Model

We start off with the individual based model. Let the population size be N, X_t the number of genotype-A-plants, δ the death rate, and $b(s)$ the distribution of the ability for a seed at age s to germinate; we require $\int_0^\infty b(s)\,ds = 1$, $B = \int_0^\infty s\,b(s)\,ds < \infty$, and $b(s)$ sufficiently smooth. Then, the rate for the transition $X_t \to X_t + 1$ is given by

$$\delta\, N\,(1 - X_t/N) \int_0^\infty b(\tau)\, X_{t-s}/N ds, \tag{17}$$

while that for a decrease of X_t by 1 reads

$$\delta\, N\,(X_t/N)\left(1 - \int_0^\infty b(\tau) X_{t-s}/N ds\right). \tag{18}$$

$$\begin{aligned} &P(X_{t+\Delta t} = X_t + 1 | X_\tau \text{ for } \tau \le t) \qquad (19)\\ &= \Delta t\, \delta\, N\,(1 - X_t/N) \int_0^\infty b(\tau)\, X_{t-s}/N ds + \mathcal{O}(\Delta t),\\ &P(X_{t+\Delta t} = X_t - 1 | X_\tau \text{ for } \tau \le t) \qquad (20)\\ &= \Delta t\, \delta\, N\,(X_t/N)\left(1 - \int_0^\infty b(\tau) X_{t-s}/N ds\right) + \mathcal{O}(\Delta t). \end{aligned}$$

Note that the delay process requires the knowledge of the complete history $\{X_s\}_{s<t}$. The usual continuous limit for

$x_t = X_t/N$ yields (with $\varepsilon = 1/N$)

$$dx_t = \delta \left(\int_0^\infty b(s)\, x_{t-s}\, ds - \; x_t \right) ds + \left\{ \varepsilon\delta \int_0^\infty b(s)(x_t + x_{t-s} - 2x_t\, x_{t-s})\, ds \right\}^{1/2} dW_t.$$

If we rescale time in the usual way, $\tau = \varepsilon t$, and define $v_\tau^\varepsilon = u_{\varepsilon\tau}^\varepsilon$, we obtain

$$dv_\tau^\varepsilon = \varepsilon^{-1}\, \delta \left(\varepsilon^{-1} \int_0^\infty b(s/\varepsilon)(v_{\tau-s}^\varepsilon - v_\tau^\varepsilon)\, ds \right) d\tau \qquad (21)$$
$$+ \left(\varepsilon^{-1}\, \delta \int_0^\infty b(s/\varepsilon)\big(v_\tau^\varepsilon + v_{\tau-s}^\varepsilon - 2\, v_\tau^\varepsilon\, v_{\tau-s}^\varepsilon\big)\, ds \right)^{1/2} dW_\tau.$$

The aim here is to find heuristic arguments indicating that v_τ^ε approximates for $\varepsilon \to 0$ the solution of a Moran diffusion process with rescaled time, paralleling equation (13).

Note that, in some sense, the terms in this time-continuous model are better to interpret than the parallel terms in the Fisher-Wright model: both terms within the brackets are moving averages, and clearly

$$\lim_{\varepsilon\to 0} \left(\varepsilon^{-1}\, \delta \int_0^\infty b(s/\varepsilon)\big(u_\tau + u_{\tau-s} - 2\, u_\tau\, u_{\tau-s}\big)\, ds \right) = \delta\, u_\tau(1 - u_\tau)$$

for a function u_τ that is reasonably smooth. For the drift term, we find similarly

$$\lim_{\varepsilon\to 0} \left(\varepsilon^{-1} \int_0^\infty b(s/\varepsilon)\, (u_{\tau-s} - u_\tau)\, ds \right) \to u_\tau - u_\tau = 0.$$

However, in eqn. (21), this bracket is divided by ε, and hence does not vanish for $\varepsilon \to 0$. If we take a closer look, we find that a deviation of x_τ from the moving average (the state of the seed bank) is punished. That is, the state of living plants can change only slower in comparison with a model without seed bank, and therefore for $\varepsilon \to 0$ we expect a diffusion model at a slower time scale.

Scaling $\varepsilon \to 0$

We drop the superscript ε in v_τ^ε, and write simply v_τ. In order to use the arguments developed above, we discetize the stochastic differential-delay equation by the Euler-Maruyama formula, and find

$$v_{\tau+\Delta\tau} = v_\tau - \varepsilon^{-1}\, \delta\, \Delta\tau \left(v_\tau - \sum_{i=1}^\infty v_{\tau-i\Delta\tau} \varphi_i^{(\Delta\tau)} \right) + \left(\delta \sum_{i=1}^\infty \varphi_i^{(\Delta\tau)} \big(v_\tau + v_{t-i\Delta\tau}^\varepsilon - 2\, v_\tau\, v_{\tau-i\Delta\tau}^\varepsilon\big) \right)^{1/2} \sqrt{\Delta\tau}\, \eta_\tau,$$

where η_τ are i.i.d. $N(0,1)$ distributed, and the weights $\varphi_i^{(\Delta t)}$ are chosen as

$$\varphi_i^{(\Delta\tau)} = b(i\, \Delta\tau/\varepsilon)(\Delta\tau/\varepsilon) + \mathcal{O}(\Delta\tau^2/\varepsilon),$$

such that $\sum_{i=1}^\infty \varphi_i^{(\Delta\tau)} = 1$. If we now define

$$\beta = \left(\delta \sum_{i=1}^\infty \varphi_i^{(\Delta\tau)} \big(v_\tau + v_{t-i\Delta\tau}^\varepsilon - 2\, v_\tau\, v_{\tau-i\Delta\tau}^\varepsilon\big) \right)^{1/2},$$

$$\psi(x) = 1 - z + \delta\Delta\tau\varepsilon^{-1} \left(z - \sum_{i=1}^\infty \varphi_i^{(\Delta t)} z^{i+1} \right),$$

we may rewrite the discretized equation for v_τ as

$$\psi(L) v_{\tau+\Delta\tau} = \beta\sqrt{\Delta\tau}\, \eta_\tau,$$

where $Lv_\tau = v_{\tau-\Delta\tau}$. We are now in the position to apply the computations about the quasi-stationary state of the seedbank (neglecting the time-dependency of β). As

$$-\psi'(1) = 1 - \delta\Delta\tau/\varepsilon + \delta \sum_{i=1}^\infty \varphi_i^{(\Delta t)} (i+1)\Delta\tau/\varepsilon$$
$$= 1 - \delta\Delta\tau/\varepsilon + \delta \sum_{i=1}^\infty b(i\, \Delta\tau/\varepsilon)(i\Delta\tau/\varepsilon)(\Delta\tau/\varepsilon)$$
$$+ \Delta\tau/\varepsilon\, \delta \sum_{i=1}^\infty \big(b(i\, \Delta\tau/\varepsilon)(\Delta\tau/\varepsilon) + \mathcal{O}(\Delta\tau^2/\varepsilon) \big),$$

we have

$$1 + \delta \int_0^\infty b(s)\, s\, ds = 1 + \delta B \quad \text{for } \Delta\tau/\varepsilon \to 0,$$

and conclude that approximately

$$v_{\tau+\Delta\tau} = v_\tau + \frac{\beta\sqrt{\Delta\tau}}{1 + \delta\, B}\, \eta_\tau.$$

Hence, for $\varepsilon \to 0$ we expect (according to these heuristic arguments) that v_τ^ε satisfies the rescale diffusion equation

$$dv_\tau = \frac{(\delta v_\tau(1 - v_\tau))^{1/2}}{1 + \delta\, B}\, dW_\tau.$$

If we define $G = 1/\delta$, the average inter-generation time of living plants, this equation becomes even closer to that derived for the Fisher-Wright case,

$$dv_\tau = \frac{(\delta v_\tau(1 - v_\tau))^{1/2}}{(1 + B/G)}\, dW_\tau \qquad (22)$$

as it becomes clear that the correction factor $1 + B/G$ measures the average time a seed rests in the soil in terms of generations.

[1] Abramowitz, M., and I. A. Stegun (1964), *Handbook of mathematical functions: with formulas, graphs, and mathematical tables* (Dover).

[2] Blath, J., B. Eldon, A. González-Casanova, N. Kurt, and M. Wilke-Berenguer (2015), Genetics **200**, 921.

[3] Blath, J., A. González Casanova, N. Kurt, and D. Spanò (2013), J. Appl. Probab. **50**, 741.

[4] Blath, J., A. González-Casanova, N. Kurt, and M. Wilke-Berenguer (2016), Ann. Appl. Prob. **26**, 857.

[5] Böndel, K. B., H. Lainer, T. Nosenko, M. Mboup, A. Tellier, and W. Stephan (2015), Mol. Biol. Evol. **32**, 2932.

[6] Brockwell, P. J., and R. A. Davis (2009), *Time Series: Theory and Methods* (Springer).

[7] Brown, J. H., and A. Kodric-Brown (1977), Ecology **58**, 445.

[8] Cohen, D. (1966), J. Theor. Biol. **12**, 119.

[9] Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester (2007), Nature **450**, 870.

[10] Etheridge, A. (2011), *Some Mathematical Models from Population Genetics*, LNM 2012 (Springer).

[11] Evans, M. E. K., and J. J. Dennehy (2005), Q. Rev. Biol **80**, 431.

[12] Evans, M. E. K., R. Ferriere, M. J. Kane, and D. L. Venable (2007), Am. Nat. **169**, 184.

[13] Ewens, W. J. (2004), *Mathematical Population Genetics: I. Theoretical Introduction* (Springer).

[14] González-Casanova, A., E. A. von Wobeser, G. Espín, L. Servín-González, N. Kurt, D. Spanò, J. Blath, and G. Soberón-Chávez (2014), J. Theor. Biol. **356**, 62.

[15] Griffiths, R. C. (2003), Theor. Popul. Biol. **64**, 241.

[16] Hadeler, K. (2013), J. Math. Biol. **66**, 649.

[17] Hairston, N. G., and B. T. Destasio (1988), Nature **336**, 239.

[18] Honnay, O., B. Bossuyt, H. Jacquemyn, A. Shimono, and K. Uchiyama (2008), Oikos **117**, 1.

[19] Kaj, I., S. M. Krone, and M. Lascoux (2001), J. Appl. Probab. **38**, 285.

[20] Kimura, M. (1955), in *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 20 (Cold Spring Harbor Laboratory Press) pp. 33–53.

[21] Kimura, M. (1969), Genetics **61**, 893.

[22] Kimura, M., and T. Ohta (1969), Genetics **61**, 763.

[23] Kingman, J. F. C. (1982), J. Appl. Probab. **19A**, 27.

[24] Kloeden, P. E., and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, Stochastic Modelling and Applied Probability, Vol. 23 (Springer).

[25] Lennon, J. T., and S. E. Jones (2011), Nat. Rev. Microb. **9**, 119.

[26] Nunney, L. (2002), Am. Nat. **160**, 195.

[27] Tellier, A., and J. K. M. Brown (2009), Am. Nat. **174**, 769.

[28] Tellier, A., I. Fischer, C. Merino, H. Xia, L. Camus-Kulandaivelu, T. Stadler, and W. Stephan (2011), Heredity **107**, 189.

[29] Tellier, A., S. J. Y. Laurent, H. Lainer, P. Pavlidis, and W. Stephan (2011), Proc. Natl. Acad. Sci. U.S.A. **108**, 17052.

[30] Tielbörger, K., M. Petruů, and C. Lampei (2012), Oikos **121**, 1860.

[31] Turelli, M., D. W. Schemske, and P. Bierzychudek (2001), Evolution **55**, 1283.

[32] Vitalis, R., S. Glemin, and I. Olivieri (2004), Am. Nat. **163**, 295.

[33] Živković, D., M. Steinrücken, Y. S. S. Song, and W. Stephan (2015), Genetics **200**, 601.

[34] Živković, D., and W. Stephan (2011), Theor. Popul. Biol. **79**, 184.

[35] Živković, D., and A. Tellier (2012), Mol. Ecol. **21**, 5434.