

Noname manuscript No.
(will be inserted by the editor)

The Fisher-Wright model with deterministic seed bank and selection

Bendix Koopmann · Johannes Müller ·
Aurélien Tellier · Daniel Živković

Received: date / Accepted: date

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

12 **Keywords** diffusion · Fisher-Wright model · seed bank · selection · site-
13 frequency spectrum · times to fixation

Bendix Koopmann · Johannes Müller
Center for Mathematics, Technische Universität München, 85748 Garching, Germany
E-mail: bendix.koopmann@mytum.de · johannes.mueller@mytum.de

Aurélien Tellier
Section of Population Genetics, Center of Life and Food Sciences Weihenstephan, Technische
Universität München, 85354 Freising, Germany
E-mail: tellier@wzw.tum.de

Daniel Živković
Section of Evolutionary Biology, Department of Biology, Ludwig-Maximilians-Universität
München, 82152 Martinsried, Germany
E-mail: zivkovic@bio.lmu.de

14 1 Introduction

15 Dormancy of reproductive structures, that is seeds or eggs, is described as a
16 bet-hedging strategy (Evans and Dennehy 2005; Cohen 1966) in plants (Hon-
17 nay et al 2008; Evans et al 2007; Tielbörger et al 2012), invertebrates (*e.g.*,
18 *Daphnia*; Decaestecker et al 2007) and microorganisms (Lennon and Jones
19 2011) to buffer against environmental variability. Bet-hedging is widely de-
20 fined as an evolutionary stable strategy in which adults release their offspring
21 into several different environments, here specifically with dormancy at differ-
22 ent generations in time, to maximize the chance of survival and reproductive
23 success, thus magnifying the evolutionary effect of good years and dampening
24 the effect of bad years (Evans and Dennehy 2005; Cohen 1966). Dormancy
25 and quiescence sometimes have surprising and counterintuitive consequences,
26 similar to diffusion in activator-inhibitor models (Haderler 2013). In the follow-
27 ing study, we focus more specifically on the evolution of dormancy in plant
28 species (Honnay et al 2008; Evans et al 2007; Tielbörger et al 2012), but the
29 theoretical models also apply to microorganisms and invertebrate species (De-
30 caestecker et al 2007; Lennon and Jones 2011.)

31 Seed banking is a specific life-history characteristic of most plant species, which
32 produce seeds remaining in the soil for short to long periods of time (up to
33 several generations), and it has large but yet underappreciated consequences
34 (Evans and Dennehy 2005) for the evolution and conservation of many plant
35 species.

36 First, polymorphism and genetic diversity are increased in a plant population
37 with seed banks compared to the situation without banks. This is mostly due
38 to storage of genetic diversity in the soil (Kaj et al 2001; Nunney 2002). Seed
39 banks also damp off the variation in population sizes over time (Nunney 2002).
40 Under unfavourable conditions at generation t , the small offspring production
41 is compensated at the next generation $t + 1$ by individuals from the bank
42 germinating at a given rate. Under the assumption of large seed banks, the
43 observed population sizes between consecutive generations (t and $t + 1$) may
44 then be uncoupled.

45 Second, seed banks may counteract habitat fragmentation by buffering against
46 the extinction of small and isolated populations, a phenomenon known as the
47 “temporal rescue effect” (Brown and Kodric-Brown 1977). Populations which
48 suffer dramatically from events of decrease in population size can be rescued
49 by seeds from the bank. Improving our understanding of the evolutionary condi-
50 tions for the existence of long-term dormancy and its genetic underpinnings
51 is thus important for the conservation of endangered plant species in habitats
52 under destruction by human activities.

53 Third, germ banks influence the rate of natural selection in populations. On
54 the one hand, seed banks promote the occurrence of balancing selection for
55 example for color morphs in *Linanthus parryae* (Turelli et al 2001) or in host-
56 parasite coevolution (Tellier and Brown 2009). On the other hand, the storage
57 effect is expected to decrease the efficiency of positive selection in populations,
58 thus natural selection, positive or negative, would be slowed down by the

59 presence of long-term seed banks. Empirical evidence for this phenomenon has
60 been shown (Hairston and Destasio 1988), but no quantitative model exists so
61 far. In general terms, understanding how seed banks evolve, affect the speed
62 of adaptive response to environmental changes, and determine the rate of
63 population extinction in many plant species is of importance for conservation
64 genetics under the current period of anthropologically driven climate change.

65 Two classes of theoretical models have been developed for studying the influ-
66 ence of seed banks on genetic variability. First, Kaj et al (2001) have proposed a
67 backward in time coalescent seed bank model which includes the probability of
68 a seed to germinate after a number of years in the soil and a maximum amount
69 of time that seeds can spend in the bank. Seed banks have the property to en-
70 hance the size of the coalescent tree of a sample of chromosomes from the above
71 ground population by a quadratic factor of the average time that seeds spend
72 in the bank. This leads to a rescaling of the Kingman coalescent (Kingman
73 1982) because two lineages can only coalesce in the above-ground population in
74 a given ancestral plant. The consequence of longer seed banks with smaller val-
75 ues of the germination rate is thus to increase the effective size of populations
76 and genetic diversity (Kaj et al 2001) and to reduce the differentiation among
77 populations connected by migration (Vitalis et al 2004). This rescaling effect
78 on the coalescence of lineages in a population has also important consequences
79 for the statistical inference of past demographic events (Živković and Tellier
80 2012). In practice this means that the spatial structure of populations and seed
81 bank effects on demography and selection are difficult to disentangle (Böndel
82 et al 2015). Nevertheless, Tellier et al (2011a) could use this rescaled seed bank
83 coalescent model (Kaj et al 2001) and Approximate Bayesian Computation to
84 infer the germination rate in two wild tomato species *Solanum chilense* and
85 *S. peruvianum* from polymorphism data (Tellier et al 2011b).

86 A second class of models assumes a strong seed bank effect, whereby the time
87 seeds can spend in the bank is very long, that is longer than the population
88 coalescent time (González-Casanova et al 2014), or the time for two lineages
89 to coalesce can be unbounded. This latest model generates a seed bank coales-
90 cent (Blath et al 2015a), which may not come down from infinity and for which
91 the expected site-frequency spectrum (SFS) may differ significantly from that
92 of the Kingman coalescent (Blath et al 2015b). In effect, the model of Kaj
93 et al (2001) represents a special case, also called a weak seed bank, where the
94 time for lineages to coalesce is bounded by the maximum time that seeds can
95 spend in the bank.

96 In the following we focus on the weak seed bank model where the time in the
97 seed bank is bounded to a small finite number assumed to be realistic for most
98 plant species (Honnay et al 2008; Evans et al 2007; Tielbörger et al 2012; Tellier
99 et al 2011b). We develop a forward in time diffusion for seed banks following
100 a Fisher-Wright model with random genetic drift and selection acting on one
101 of two genotypes. The time rescaling induced by the seed bank is shown to
102 be equivalent for the Fisher-Wright and the Moran model. We provide the
103 first theoretical estimates of the effect of seed bank on natural selection by

104 deriving the expected SFS of alleles observed in a sample of chromosomes and
105 the time to fixation of an allele. Note that we do not prove every step in the
106 most rigorous sense but keep the derivations on a more intuitive level to focus
107 on the overall line of reasoning and biological implications.

108 2 Model and Diffusion Limit

109 2.1 Model description

110 We consider a finite plant-population of size N . The plants appear in two
111 genotypes A and a . We assume non-overlapping generations. Let X_n denote
112 the number of type- A plants in generation n (that is, the number of living type-
113 a plants in this generation is $N - X_n$). Plants produce seeds. The number of
114 seeds is assumed to be large, such that noise in the seed bank does not play
115 a role (therefore we call the seed bank “deterministic”). The amount of seeds
116 produced by type- A -plants in generation n is $\beta_A X_n$, that of type- a plants
117 $\beta_a(N - X_n)$. The seeds are stored *e.g.* in the soil and may germinate in the
118 next generation, but also in later generations.

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

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

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

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

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

$$X_n \sim \text{Binom}(q_n(X_{\bullet}), N), \quad \text{where} \\ q_n(X_{\bullet}) = \frac{\sum_{i=1}^{\infty} b_A(i) \beta_A X_{n-i}}{\sum_{i=1}^{\infty} b_A(i) \beta_A X_{n-i} + \sum_{i=1}^{\infty} b_a(i) \beta_a (N - X_{n-i})}. \quad (1)$$

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

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

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

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

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

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

134 (the sign is chosen in such a way that genotype A has an advantage over
 135 genotype a for $\sigma > 0$) then (1) with selection becomes

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

136 As this ratio is homogeneous of degree zero in b_A , we assume $\sum_{i=1}^{\infty} b_A(i) = 1$.
 137 That is, $b_A(i)$ is considered a probability distribution for the survival of a
 138 (type-A) seed. From now on, we will assume that the maximum and therefore
 139 also the average life time of a seed is finite, $B = \sum_{i=1}^{\infty} i b_A(i) < \infty$. The sum
 140 $\sum_{i=1}^{\infty} b_A(i) X_{n-i}$ is a moving average. We emphasize this fact by introducing
 141 the operator

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

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

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

143 2.2 Diffusion limit

144 The aim of this section is to demonstrate that under an appropriate scaling of
 145 X_n and time, the model approximates the diffusive Moran model. Before we
 146 start, we recall briefly the corresponding procedure for the standard Fisher-
 147 Wright model.

148 2.2.1 The Fisher-Wright model without selection

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

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

154 • *Rescale time:* Now define $\Delta\tau = 1/N$, introduce the time $\tau = n\Delta\tau$, let
 155 $u_{n\Delta\tau} = x_n$, and rescale the index of the normal random variables, that is,
 156 replace η_n by $\eta_{n\Delta\tau} = \eta_\tau$. Then, $u_{\tau+\Delta\tau} - u_\tau = \Delta\tau^{1/2} (u_\tau(1 - u_\tau))^{1/2} \eta_\tau$.
 157 According to the Euler-Maruyama formula (see *e.g.* Kloeden and Platen 1992),
 158 we approximate the diffusive Moran model for N large (that is, $\Delta\tau = 1/N$
 159 small)

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

160 Mostly, the approximation of the binomial distribution by a normal distribu-
 161 tion and the scaling of time is done in one step; however, as in seed bank
 162 models the different time scales are decisive, we prefer to keep these two steps
 163 separated.

164 *2.2.2 Seed bank model with a geometric germination rate and without*
 165 *selection*

166 There is one case where our model becomes particularly simple: if we have
 167 no selection, and the $b(i)$ follow a geometric distribution with parameter
 168 $\mu \in (0, 1)$. In this case, the delay-model is equivalent to a proper Markov chain.
 169 As a warm-up, we will first derive the diffusion limit for this special case.

170 **Proposition 1** *Consider the seed bank model described in section 2.1 for*
 171 *$\sigma = 0$. Define $z_n = \sum_{i=1}^{\infty} b(i)X_{n+1-i}/N$. Let $b(1) = \mu$ and $b(i) = (1 - \mu)b(i - 1)$.*
 172 *Then,*

$$z_{n+1} = \mu X_{n+1}/N + (1 - \mu) z_n, \quad \text{and} \quad X_{n+1} \sim \text{Binom}(z_n, N). \quad (3)$$

173 **Proof:** It is simple to see that $z_n = \mu \sum_{i=1}^{\infty} (1 - \mu)^{i-1} X_{n+1-i}/N$. We imme-
 174 diately obtain

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

175 Next (and with the nomenclature of (2)), we have

$$q_{n+1}(X_\bullet) = M_{n+1}(X_\bullet/N) = \sum_{i=1}^{\infty} b(i)X_{n+1-i}/N = z_n.$$

176 Hence, $X_{n+1} \sim \text{Binom}(q_{n+1}, N) = \text{Binom}(z_n, N)$. □

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

179 As this model is Markovian, it is simple to derive the diffusion limit. As usual,
 180 we start off by defining $x_n = X_n/N$, and obtain $z_n = \mu x_n + (1 - \mu) z_{n-1}$,

181 $X_{n+1} = \text{Binom}(z_n, N)$. Approximating the Binomial distribution by a normal
 182 distribution for N large yields

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

183 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
 184 foregoing two equations give

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

185 Therefore, $z_{n+1} - z_n = \mu N^{-1/2}(z_n(1 - z_n))^{1/2}\eta_n$. Scaling time by N yields
 186 for $u_{n/N} = z_n$ and $\tau = n/N$

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

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

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

189 We find a diffusive Moran model for the state of the seed bank with rescaled
 190 time scale. We expect a similar result to hold in the general case. A difference
 191 between the two cases is that we here naturally considered the state of the
 192 seed bank, while in the general case we will focus on the state of living plants.

193 2.2.3 The seed bank model with selection

194 We go through the equivalent steps for the Fisher-Wright model with deter-
 195 ministic seed bank and selection.

196 **Proposition 2** Consider the seed bank model described in section 2.1 and let
 197 $x_n = X_n/N$ and $\Delta t = 1/N$. Then, (2) becomes

$$\begin{aligned} & x_n - M_n(x_\bullet) - \Delta t \sigma M_n(x_\bullet)(1 - M_n(x_\bullet)) + \mathcal{O}(\Delta t^2) \\ &= \Delta t^{1/2} \left\{ \left(M_n(x_\bullet)(1 - M_n(x_\bullet)) \right)^{1/2} + \mathcal{O}(\Delta t) \right\} \eta_n. \end{aligned} \quad (5)$$

198 **Proof:** From (2), we immediately have

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

199 For N large, the binomial distribution can be well approximated by a normal
 200 distribution, so that

$$\begin{aligned} x_n &\approx \frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} \\ &+ \Delta t^{1/2} \left(\frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} \right)^{1/2} \left(1 - \frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} \right)^{1/2} \eta_n, \end{aligned}$$

201 where $\eta_n \sim N(0, 1)$. As the noise and the drift term scale differently, an $\Delta t^{1/2}$
 202 order approximation for this term is sufficient, and we have

$$\begin{aligned} x_n - \frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} \\ \approx \Delta t^{1/2} \left(\frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} \right)^{1/2} \left(1 - \frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} \right)^{1/2} \eta_n \\ = \Delta t^{1/2} \left\{ \left(M_n(x_\bullet)(1 - M_n(x_\bullet)) \right)^{1/2} + \mathcal{O}(\Delta t) \right\} \eta_n. \end{aligned}$$

203 Finally, we use a first-order Taylor-expansion for the drift term in Δt to obtain

$$\frac{M_n(x_\bullet)}{1 - \Delta t \sigma (1 - M_n(x_\bullet))} = M_n(x_\bullet) + \Delta t \sigma M_n(x_\bullet)(1 - M_n(x_\bullet)) + \mathcal{O}(\Delta t^2),$$

204 which yields the desired result. □

205 In the following we neglect the higher order terms. If we consider the scaling
 206 of the terms w.r.t. Δt , then the leading term is $x_n - M_n(x_\bullet)$. This difference
 207 must not become too large, as all other terms in the equation are at least of
 208 order $\Delta t^{1/2}$. That is, the state x_n can only slowly drift away from $M_n(x_\bullet)$
 209 (which represents the state of the seed bank). Hence, for a reasonable number
 210 of time steps, $M_n(x_\bullet)$ is fairly constant. In order to understand the model, we
 211 define

$$\alpha = \Delta t \sigma M_n(x_\bullet)(1 - M_n(x_\bullet)), \quad \beta = \left(M_n(x_\bullet)(1 - M_n(x_\bullet)) \right)^{1/2}.$$

212 α and β are random variables that depend on time. However, if we assume a
 213 separation of time scales, then we understand the dynamics of the model at a
 214 short time horizon by considering the surrogate model

$$y_n - M_n(y_\bullet) - \alpha = \Delta t^{1/2} \beta \eta_n, \tag{6}$$

215 according to (5), and α , β and Δt being positive, real-valued constants. This
 216 recursive equation is well known as an auto-regression (AR) model in the
 217 statistical modelling of time series. If $\alpha \neq 0$, this model incorporates a trend.
 218 We first remove this trend.

219 **Proposition 3** Assume (6) and define $z_n = y_n - w_n$ with $w_n = n \alpha / B$ and
 220 $B = \sum_{i=1}^{\infty} i b_A(i)$. Then,

$$z_n - M_n(z_\bullet) = \Delta t^{1/2} \beta \eta_n.$$

221 **Proof:** By definition of M_n , we have $M_n(w_\bullet) = \sum_{i=1}^{\infty} b_A(i) (n-i) \alpha/B =$
 222 $n \alpha/B - \alpha$. We replace y_n by $z_n + w_n$ in (6), and find with $M_n(y_\bullet) = M_n(z_\bullet) +$
 223 $M_n(w_\bullet)$,

$$\begin{aligned} z_n + n \frac{\alpha}{B} - \left(M_n(z_\bullet) + M_n(w_\bullet) \right) - \alpha \\ = z_n - M_n(z_\bullet) + n \frac{\alpha}{B} - \alpha - \left(n \frac{\alpha}{B} - \alpha \right) \\ = \Delta t^{1/2} \beta \eta_n. \end{aligned}$$

□

224 Next we convert the AR model into a moving average equation.

225 **Proposition 4** Let $z_n - M_n(z_\bullet) = \Delta t^{1/2} \beta \eta_n$, where η_n are i.i.d. $N(0, 1)$ -
 226 distributed. For $\Delta t \ll 1$, and n large, z_n satisfies approximately the recursive
 227 equation

$$z_n \approx z_{n-1} + \frac{\Delta t^{1/2} \beta}{B} \eta_n.$$

228 **Proof:** We define the back-shift operator acting on the index of a sequence,
 229 $Lz_n = z_{n-1}$, and a power series

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

230 Therewith we may write

$$\psi(L)z_n = z_n - M_n(z_\bullet) = \Delta t^{1/2} \beta \eta_n.$$

231 Note that $\psi(1) = 0$, which does mean that the AR model is non-stationary. We
 232 do not find a power series $\psi^*(x)$ well defined at $x = 1$ such that $\psi^*(x) \psi(x) = 1$.
 233 Therefore, we rewrite $\psi(x)$ as $\psi(x) = (1-x) \tilde{\psi}(x)$ (which is the defining
 234 equation of $\tilde{\psi}(x)$). As

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

235 we do find $\psi^*(x)$ such that $\psi^*(x) \tilde{\psi}(x) = 1$, and hence $\psi^*(x) \psi(x) = 1-x$ in
 236 a neighbourhood of $x = 1$. As an immediate consequence (used later) we have
 237 $\psi^*(1) = 1/B$. If we multiply the equation $\psi(L)z_n = \Delta t^{1/2} \beta \eta_n$ by $\psi^*(L)$, we
 238 obtain

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

239 and

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

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

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

241 If we inspect not rows (that have $\psi^*(L)\eta_{i-\ell}$ as entries) but columns (that
242 contain always the same random variable $\eta_{i-\ell}$), we find that the coefficient in
243 front of one given random variable $\eta_{i-\ell}$ approximates $\psi^*(1)$ for $\ell \rightarrow \infty$.

244 At this point, we want to write $z_{n+1} \approx \Delta t^{1/2} \beta \psi^*(1) \sum_{\ell=1}^n \eta_{\ell}$. This is only
245 true, also in an approximate sense, if n is large and the state z_n does hardly
246 change over a time scale that allows $\sum_{i=1}^m a_i$ to converge to $\psi^*(1) = 1/B$.
247 If $\Delta t^{1/2}$ is small, then z_n indeed changes on a time scale given by $1/\Delta t$ (for
248 our evolutionary model, we have convergence of the sum on the ecological
249 time scale, and the change of z_n on the evolutionary time scale, which are
250 completely different if the population size is large). Hence, for Δt small we are
251 allowed to assume

$$z_{n+1} \approx \Delta t^{1/2} \beta \psi^*(1) \sum_{\ell=1}^n \eta_{\ell} = \frac{\Delta t^{1/2} \beta}{B} \sum_{\ell=1}^n \eta_{\ell}.$$

252 Thus, $z_{n+1} \approx (\Delta t^{1/2} \beta/B) \sum_{\ell=1}^n \eta_{\ell}$ and $z_{n+1} - z_n \approx (\Delta t^{1/2} \beta/B) \eta_n$. □

253 We return to y_n again, and find:

254 **Corollary 1** Let $M_n(y_{\bullet}) = \sum_{i=1}^{\infty} b_A(i)y_{n-i}$, and $y_n - M_{n-q}(y_{\bullet}) + \alpha =$
255 $\Delta t^{1/2} \beta \eta_{n-q}$ for $\alpha, \Delta t, \beta \in \mathbb{R}_+$. Then, for Δt small, y_n satisfies approxi-
256 mately the recursive equation

$$y_n = y_{n-1} + \frac{\alpha}{B} + \frac{\Delta t^{1/2} \beta}{B} \eta_{n-1},$$

257 where $B = \sum_{i=1}^{\infty} i b_A(i)$.

258 [Fig. 1 about here.]

259 *Remark 1* If we start with $y_0 = 0$, we expect that y_n is (approximately) nor-
260 mally distributed with expectation $n\alpha/B$, and variance $n\Delta t\beta^2/B^2$. In order
261 to check the heuristic argumentation numerically, we took $\alpha = 0.01$, $\Delta t = 0.01$,
262 $\beta = 2$ and $M_n(x_{\bullet}) = \frac{1}{m} \sum_{i=1}^m x_{n-i}$ for $m = 9$, that is, $B = 5$. Simulations show
263 an excellent agreement with our computations (Fig. 1).

264 Now we return to the scaled Fisher-Wright model with seed bank. Though
265 $M_n(x_{\bullet})$ will change, we expect it to change on the evolutionary time scale,
266 while the generations n are still on the ecological time scale. Hence, we are
267 allowed to use corollary 1 to obtain the following result.

268 **Corollary 2** *The realizations $\{x_n\}_{n \in \mathbb{N}_0}$ of the AR model given in (5) satisfy*
 269 *for small Δt ($= 1/N$) approximately the equation*

$$x_n = x_{n-1} + \Delta t \frac{\sigma}{B} M_n(x_\bullet)(1 - M_n(x_\bullet)) + \Delta t^{1/2} \frac{1}{B} \left(M_n(x_\bullet)(1 - M_n(x_\bullet)) \right)^{1/2} \eta_n.$$

270 This formulation allows to rescale time. We work on an evolutionary time scale
 271 instead of generations. This yields an SDE.

272 **Theorem 1** *Let $u_{n\Delta t} = x_n$. If x_n only changes on the time scale given by*
 273 *$1/\Delta t$, then u_t satisfies for Δt small approximately the SDE*

$$du_t = \frac{\sigma}{B} u_t(1 - u_t)dt + \frac{1}{B} \left(u_t(1 - u_t) \right)^{1/2} dW_t. \quad (7)$$

274 **Proof:** Let $\hat{M}_{n\Delta t}(u_\bullet) = \sum_{i=0}^{\infty} b_A(i+1)u_{(n-i)\Delta t}$ (note the index shift between
 275 M_n and \hat{M}_n , which corresponds to an index shift in the next equation from
 276 x_n to $u_{t+\Delta t}$). Then,

$$u_{t+\Delta t} - u_t = \Delta t \frac{\sigma}{B} \hat{M}_{n\Delta t}(u_\bullet)(1 - \hat{M}_{n\Delta t}(u_\bullet)) \\ + \Delta t^{1/2} \frac{1}{B} \left(\hat{M}_{n\Delta t}(u_\bullet)(1 - \hat{M}_{n\Delta t}(u_\bullet)) \right)^{1/2} \eta_n.$$

277 Hence, u_t changes on the time scale determined by $1/\Delta t$, that is, slowly in
 278 comparison with n . If the $b_A(i)$ decline fast enough (resp. Δt is small enough),
 279 then x_t is fairly constant on the time scale used for the moving average, that
 280 is, $\hat{M}_{n\Delta t}(u_\bullet) \approx u_t$. □

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

285 We now change the time scale such that the variance coincides with the stan-
 286 dard diffusive Moran model.

287 **Corollary 3** *If we define $\tau = t/B^2$, then the SDE reads*

$$du_\tau = (\sigma B) u_\tau(1 - u_\tau)d\tau + \left(u_\tau(1 - u_\tau) \right)^{1/2} dW_\tau. \quad (8)$$

288 **Scaling of the selection parameter.** We conclude that the appropriate
 289 scaling of time for the Fisher-Wright model with seed bank is not $1/N$ but
 290 $1/(B^2 N)$. Moreover, the effective selection rate (w.r.t. this time) is increased
 291 by the average number of generations B the seeds sleep in the soil.

292 **3 The forward diffusion equation for seed bank models with** 293 **selection**

294 In analogy to above, we consider a single locus and two allelic types A and a
295 with frequencies x and $1 - x$, respectively, at time zero. Time is scaled in units
296 of $2N$ generations. In the diffusion limit, as $N \rightarrow \infty$, the probability $f(y, t)dy$
297 that the type- A genotype has a frequency in $(y, y + dy)$ is characterized by the
298 following forward equation (see Kimura 1955 for $B = 1$):

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

299 where the drift and the diffusion terms are given by $a(y) = \sigma y(1 - y)/B$ and
300 $b(y) = y(1 - y)/B^2$, respectively.

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

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

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

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

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

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

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

307 3.1 Site-frequency spectra

308 The site-frequency spectrum (SFS) of a sample (*e.g.*, Griffiths 2003; Živković
309 and Stephan 2011) is widely used for population genetics data analysis. A
310 sample of size k is sequenced, and for each polymorphic site the number of
311 individuals in which the mutation appears is determined. In this way, a dataset
312 is generated that summarizes the number of mutations $\zeta_{k,i}$ appearing in i
313 individuals, $i = 1, \dots, k - 1$. That is, $\zeta_{k,1} = 10$ indicates that 10 mutations
314 only appeared once, and $\zeta_{k,2} = 5$ tells us that five mutations were present
315 in two individuals (where the pair of individuals may be different for each of
316 the five mutations). Note that neither $\zeta_{k,0}$ nor $\zeta_{k,k}$ are sensible: a mutation
317 that appears in none or all individuals of the sample cannot be recognized as
318 a mutation. In practice, it is often not possible to know the ancestral state.
319 Then the folded SFS $\eta_{k,i} = (\zeta_{k,i} + \zeta_{k,k-i})(1 + 1_{\{i=k-i\}})^{-1}$ can be used. Since
320 both empirical observations and theoretical results for the folded SFS follow
321 instantaneously from the unfolded one, we only consider the unfolded version.

322 For the derivation of the theoretical SFS, we assume that mutations occur
 323 according to the infinitely-many sites model (Kimura 1969). The scaled muta-
 324 tion rate is given by $\theta = 4N\nu$, where ν is the mutation rate per generation
 325 at independent sites. Assuming that each mutant allele marginally follows the
 326 diffusion model specified above, the proportion of sites where the mutant fre-
 327 quency is in $(y, y + dy)$ is given by (Griffiths 2003)

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

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

331 The equilibrium solution of the SFS for a sample of size k is obtained via
 332 binomial sampling (see Živković et al 2015 for $B = 1$) as

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

333 where ${}_1F_1$ denotes the confluent hypergeometric function of the first kind
 334 (Abramowitz and Stegun 1964). For neutrality, we again immediately obtain
 335 $\hat{f}_{k,i} = \theta B^2/i$ by letting $\sigma \rightarrow 0$. For a large number of mutant sites, the
 336 relative SFS $\hat{r}_{k,i} = \hat{f}_{k,i} / \sum_{j=1}^{k-1} \hat{f}_{k,j}$ approximates the empirical distribution
 337 $\zeta_{k,i} / \sum_{j=1}^{k-1} \zeta_{k,j}$ for a constant population size. Note that the solutions for the
 338 absolute SFS assume that mutations can occur at any time. When assum-
 339 ing that mutations can only arise in living plants (Kaj et al 2001), θ has to
 340 be replaced by θ/B in the respective equations. Both mutation models give
 341 equivalent results for the relative SFS.

342 [Fig. 2 about here.]

343 As shown in Figure 2a, the neutral diffusion approximation is in line with
 344 the simulation results of the original discrete model. The theoretical relative
 345 SFS for a sample of 250 individuals approximates the simulated SFS, which is
 346 obtained as an average over 10,093 repetitions. In every iteration, the sample
 347 is drawn from an initially monomorphic population of 1000 individuals after
 348 400,000 generations (so that the population has reached an equilibrium). Fig-
 349 ure 2b illustrates the enhanced effect of selection proportional to the length of
 350 the seed bank.

351 3.2 Times to fixation

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

355 (Ewens 2004)

$$\bar{t}(x) = \int_0^1 t(x, y) dy, \quad (9)$$

356 where

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

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

357 For genic selection the integral in (9) cannot be analytically solved. For selec-
 358 tive neutrality, we obtain $\bar{t}(x) = -2B^2(x \log(x) + (1-x) \log(1-x))$ (see
 359 *e.g.* Ewens 2004 for $B = 1$) by employing the drift term, the scale density and
 360 the probabilities of absorption as specified above.

361 Now, we evaluate the time until a mutant allele is fixed conditional on fixation
 362 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 selec-
 363 tion the mean time to fixation in dependency of x can only be derived as a
 364 very lengthy expression in terms of exponential integral functions. The neutral
 365 result is found as $\bar{t}^*(x) = -2B^2(1-x)/x \log(1-x)$ and in accordance with a
 366 classical result (Kimura and Ohta 1969) for $B = 1$. For $x \rightarrow 0$, we obtain

$$\bar{t}^* = \frac{2B}{\sigma(e^{2B\sigma} - 1)} ((e^{2B\sigma} + 1)\gamma - \text{Ei}(2B\sigma) + \log(2B\sigma) + e^{2B\sigma}(-\text{Ei}(-2B\sigma) + \log(2B\sigma))), \quad \sigma > 0, \quad (10)$$

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

367 where γ is Euler's constant and Ei denotes the exponential integral function
 368 (Abramowitz and Stegun 1964).

369 [Fig. 3 about here.]

370 In Figure 3a, we compare the time to absorption of the original discrete seed
 371 bank model by means of simulations with the theoretical result obtained from
 372 the diffusion approximation. For b_A we use uniform distributions, where we
 373 vary the expected values between 1 and 8 corresponding to the length of the
 374 seed banks between 1 and 15. We choose an initial fraction of 0.5 for the type-
 375 A genotypes. The simulations show a good agreement between our analytical
 376 approximation and the numerical simulations. In Figure 3b, we show the effect
 377 of the seed bank on the times to fixation conditional on fixation of the type-A
 378 genotype for neutrality and positive selection.

379 4 Discussion

380 Within this study, we develop a forward in time Fisher-Wright model of a
381 deterministically large seed bank with drift occurring in the above-ground
382 population. The time that seeds can spend in the bank is bounded and finite, as
383 assumed to be realistic for many plant or invertebrate species. We demonstrate
384 that scaling time in the diffusion process by a factor B^2 generates the usual
385 Fisher-Wright time scale of genetic drift with B being defined as the average
386 amount of time that seeds spend in the bank. The conditional time to fixation
387 of a neutral allele is slowed down by a factor B^2 (Figure 3b, dotted line)
388 compared to the absence of seed bank. These results are consistent with the
389 backward in time coalescent model from Kaj et al (2001), and differs from the
390 strong seed bank model of Blath et al (2015a). We evaluate the SFS based
391 on our diffusion process and confirm agreement to the SFS obtained under
392 discrete time Fisher-Wright simulations.

393 In the second part of the study, we introduce selection occurring at one of
394 the two alleles, mimicking positive or negative selection. Two features of se-
395 lection under seed banks are noticeable. First, selection is slower under longer
396 seed banks (Figure 3b, solid line) confirming previous intuitive expectations
397 (Hairston and Destasio 1988). Second, when computing the SFS with $B = 2$
398 and without seed bank ($B = 1$) under positive selection ($\sigma = 2$) we reveal a
399 stronger signal of selection for the seed bank by means of an amplified uptick
400 of high-frequency derived variants. This effect becomes more prominent with
401 longer seed banks and also holds for purifying selection, under which an in-
402 crease in low-frequency derived variants is induced by the seed bank. We ex-
403 plain this counterintuitive results as follows: longer seed banks increase, on the
404 one hand, the selection coefficient σ generating a stronger signal at equilibrium
405 (Figure 2b), and on the other hand, the time to reach this equilibrium state
406 (Figure 3b). Our predictions are consistent with the inferred strengths of pu-
407 rifying selection in wild tomato species. Indeed, purifying selection at coding
408 regions appears to be stronger in *S. peruvianum* than in its sister species *S.*
409 *chilense* (Tellier et al 2011a) with *S. peruvianum* exhibiting a longer seed bank
410 (Tellier et al 2011b).

411 **Acknowledgements** This research is supported in part by Deutsche Forschungsgemein-
412 schaft grants TE 809/1 (AT) and STE 325/14 from the Priority Program 1590 (DZ).

413 Appendix A Moran model with deterministic seed bank

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

417 A.1 Model

418 We start off with the individual based model. Let the population size be N , X_t the number
 419 of genotype-A-plants, μ the death rate, and $b(s)$ the distribution of the ability for a seed at
 420 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
 421 smooth. Then,

$$P(X_{t+\Delta t} = X_t + 1 | X_\tau \text{ for } \tau \leq t) = \Delta t \mu N (1 - X_t/N) \int_0^\infty b(\tau) X_{t-s}/N ds + \mathcal{O}(\Delta t), \quad (11)$$

$$P(X_{t+\Delta t} = X_t - 1 | X_\tau \text{ for } \tau \leq t) = \Delta t \mu N (X_t/N) \left(1 - \int_0^\infty b(\tau) X_{t-s}/N ds \right) + \mathcal{O}(\Delta t). \quad (12)$$

422 Note that the delay process requires the knowledge of the complete history $\{X_s\}_{s < t}$. The
 423 usual continuous limit for $x_t = X_t/N$ yields (with $\varepsilon = 1/N$)

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

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

$$\begin{aligned} dv_\tau^\varepsilon &= \varepsilon^{-1} \mu \left(\varepsilon^{-1} \int_0^\infty b(s/\varepsilon) (v_{\tau-s}^\varepsilon - v_\tau^\varepsilon) ds \right) d\tau \\ &\quad + \left(\varepsilon^{-1} \mu \int_0^\infty b(s/\varepsilon) (v_\tau^\varepsilon + v_{\tau-s}^\varepsilon - 2v_\tau^\varepsilon v_{\tau-s}^\varepsilon) ds \right)^{1/2} dW_\tau. \end{aligned} \quad (13)$$

425 The aim here is to find heuristic arguments indicating that v_τ^ε approximates for $\varepsilon \rightarrow 0$ the
 426 solution of a Moran diffusion process with rescaled time, paralleling equation (7).

427 *Remark 2* In some sense, the terms in this time-continuous model are better to interpret
 428 than the parallel terms in the Fisher-Wright model: both terms within the brackets are
 429 moving averages, and clearly

$$\lim_{\varepsilon \rightarrow 0} \left(\varepsilon^{-1} \mu \int_0^\infty b(s/\varepsilon) (u_\tau + u_{\tau-s} - 2u_\tau u_{\tau-s}) ds \right) \rightarrow \mu u_\tau (1 - u_\tau)$$

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

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

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

436 A.2 Scaling $\varepsilon \rightarrow 0$

437 We drop the superscript ε in v_τ^ε , and write simply v_τ . In order to use the arguments devel-
 438 oped above, we discretize the stochastic differential-delay equation by the Euler-Maruyama
 439 formula, and find

$$\begin{aligned} v_{\tau+\Delta\tau} &= v_\tau - \varepsilon^{-1} \mu \Delta\tau \left(v_\tau - \sum_{i=1}^\infty v_{\tau-i\Delta\tau} \varphi_i^{(\Delta\tau)} \right) \\ &\quad + \left(\mu \sum_{i=1}^\infty \varphi_i^{(\Delta\tau)} (v_\tau + v_{\tau-i\Delta\tau}^\varepsilon - 2v_\tau v_{\tau-i\Delta\tau}^\varepsilon) \right)^{1/2} \sqrt{\Delta\tau} \eta_\tau, \end{aligned}$$

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

$$\varphi_i^{(\Delta\tau)} = b(i \Delta\tau/\varepsilon)(\Delta\tau/\varepsilon) + \mathcal{O}(\Delta\tau^2/\varepsilon), \quad \text{such that } \sum_{i=1}^{\infty} \varphi_i^{(\Delta\tau)} = 1.$$

441 If we now define

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

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

442 we may rewrite the discretized equation for v_τ as

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

443 where $Lv_\tau = v_{\tau-\Delta\tau}$. We are now in the position of the proof for Prop. 4 (neglecting the
444 time-dependency of β). As

$$\begin{aligned} -\psi'(1) &= 1 - \mu \Delta\tau/\varepsilon + \mu \sum_{i=1}^{\infty} \varphi_i^{(\Delta t)} (i+1) \Delta\tau/\varepsilon \\ &= 1 - \mu \Delta\tau/\varepsilon + \mu \sum_{i=1}^{\infty} b(i \Delta\tau/\varepsilon)(i \Delta\tau/\varepsilon)(\Delta\tau/\varepsilon) \\ &\quad + \Delta\tau/\varepsilon \mu \sum_{i=1}^{\infty} (b(i \Delta\tau/\varepsilon)(\Delta\tau/\varepsilon) + \mathcal{O}(\Delta\tau^2/\varepsilon)), \end{aligned}$$

445 we have

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

446 and conclude that approximately

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

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

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

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

$$dv_\tau = \frac{(v_\tau(1-v_\tau))^{1/2}}{1 + B/G} dW_\tau \quad (14)$$

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

References

- 453
454 Abramowitz M, Stegun IA (1964) Handbook of mathematical functions: with formulas,
455 graphs, and mathematical tables. Dover
456 Blath J, Eldon B, González-Casanova A, Kurt N, Wilke-Berenguer M (2015a) Genetic vari-
457 ability under the seed bank coalescent. *Genetics* 200:921–934
458 Blath J, González-Casanova A, Kurt N, Wilke-Berenguer M (2015b) A new coalescent for
459 seed-bank models. *Ann Appl Prob* (in press)
460 Böndel KB, Lainer H, Nosenko T, Mboup M, Tellier A, Stephan W (2015) North-south
461 colonization associated with local adaptation of the wild tomato species *Solanum chilense*.
462 *Mol Biol Evol* 32:2932–2943
463 Brown JH, Kodric-Brown A (1977) Turnover rates in insular biogeography: Effect of immi-
464 gration on extinction. *Ecology* 58:445–449
465 Cohen D (1966) Optimizing reproduction in a randomly varying environment. *J Theor Biol*
466 12:119–129
467 Decaestecker E, Gaba S, Raeymaekers JAM, Stoks R, Van Kerckhoven L, Ebert D,
468 De Meester L (2007) Host-parasite ‘red queen’ dynamics archived in pond sediment.
469 *Nature* 450:870–873
470 Evans MEK, Dennehy JJ (2005) Germ banking: Bet-hedging and variable release from egg
471 and seed dormancy. *Q Rev Biol* 80:431–451
472 Evans MEK, Ferriere R, Kane MJ, Venable DL (2007) Bet hedging via seed banking in
473 desert evening primroses (oenothera, onagraceae): Demographic evidence from natural
474 populations. *Am Nat* 169:184–194
475 Ewens WJ (2004) *Mathematical Population Genetics: I. Theoretical Introduction*. Springer
476 González-Casanova A, von Wobeser EA, Espín G, Servín-González L, Kurt N, Spanò D,
477 Blath J, Soberón-Chávez G (2014) Strong seed-bank effects in bacterial evolution. *J*
478 *Theor Biol* 356:62–70
479 Griffiths RC (2003) The frequency spectrum of a mutation, and its age, in a general diffusion
480 model. *Theoretical Population Biology* 64:241–251
481 Hadelér K (2013) Quiescence, excitability, and heterogeneity in ecological models. *J Math*
482 *Biol* 66:649–684
483 Hairston NG, Destasio BT (1988) Rate of evolution slowed by a dormant propagule pool.
484 *Nature* 336:239–242
485 Honnay O, Bossuyt B, Jacquemyn H, Shimono A, Uchiyama K (2008) Can a seed bank
486 maintain the genetic variation in the above ground plant population? *Oikos* 117:1–5
487 Kaj I, Krone SM, Lascoux M (2001) Coalescent theory for seed bank models. *J Appl Probab*
488 38:285–300
489 Kimura M (1955) Stochastic processes and distribution of gene frequencies under natural
490 selection. In: *Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor
491 Laboratory Press, vol 20, pp 33–53
492 Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite pop-
493 ulation due to steady flux of mutations. *Genetics* 61:893–903
494 Kimura M, Ohta T (1969) The average number of generations until fixation of a mutant
495 gene in a finite population. *Genetics* 61:763–771
496 Kingman JFC (1982) On the genealogy of large populations. *J Appl Probab* 19A:27–43
497 Kloeden PE, Platen E (1992) *Numerical Solution of Stochastic Differential Equations. Appli-*
498 *cations of Mathematics, Stochastic Modelling and Applied Probability, Vol. 23*, Springer
499 Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implica-
500 tions of dormancy. *Nat Rev Microb* 9:119–130
501 Nunney L (2002) The effective size of annual plant populations: The interaction of a seed
502 bank with fluctuating population size in maintaining genetic variation. *Am Nat* 160:195–
503 204
504 Tellier A, Brown JKM (2009) The influence of perenniality and seed banks on polymorphism
505 in plant-parasite interactions. *Am Nat* 174:769–779
506 Tellier A, Fischer I, Merino C, Xia H, Camus-Kulandaivelu L, Stadler T, Stephan W (2011a)
507 Fitness effects of derived deleterious mutations in four closely related wild tomato species
508 with spatial structure. *Heredity* 107:189–199

- 509 Tellier A, Laurent SJY, Lainer H, Pavlidis P, Stephan W (2011b) Inference of seed bank
510 parameters in two wild tomato species using ecological and genetic data. *Proc Natl Acad*
511 *Sci USA* 108:17,052–17,057
- 512 Tielbörger K, Petruň M, Lampei C (2012) Bet-hedging germination in annual plants: a
513 sound empirical test of the theoretical foundations. *Oikos* 121:1860–1868
- 514 Turelli M, Schemske DW, Bierzychudek P (2001) Stable two-allele polymorphisms main-
515 tained by fluctuating fitnesses and seed banks: protecting the blues in *Linanthus parryae*.
516 *Evolution* 55:1283–1298
- 517 Vitalis R, Glemin S, Olivieri I (2004) When genes go to sleep: The population genetic
518 consequences of seed dormancy and monocarpic perenniality. *Am Nat* 163:295–311
- 519 Živković D, Stephan W (2011) Analytical results on the neutral non-equilibrium allele fre-
520 quency spectrum based on diffusion theory. *Theor Popul Biol* 79:184–191
- 521 Živković D, Tellier A (2012) Germ banks affect the inference of past demographic events.
522 *Mol Ecol* 21:5434–5446
- 523 Živković D, Steinrücken M, Song YSS, Stephan W (2015) Transition densities and sam-
524 ple frequency spectra of diffusion processes with selection and variable population size.
525 *Genetics* 200:601–617

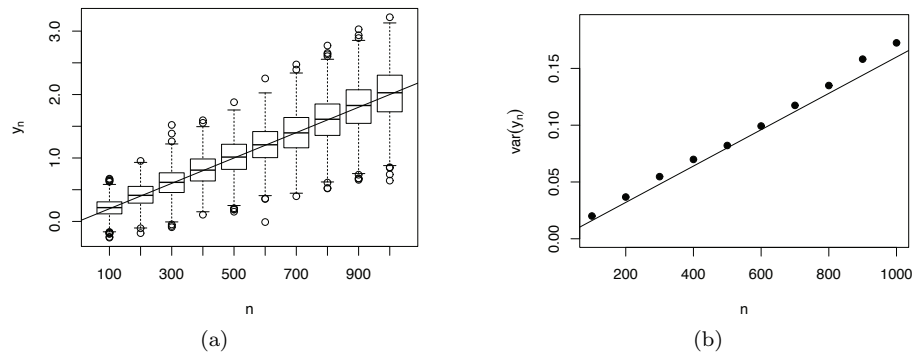


Fig. 1: Simulation of the AR model (1000 runs). Samples have been taken at time steps 100, 200, ..., 1000. (a) Boxplot of the simulated time series y_n at indicated time points together with the mean according to corollary 1 (line). (b) Variance of the simulated time series at indicated time points (dots),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

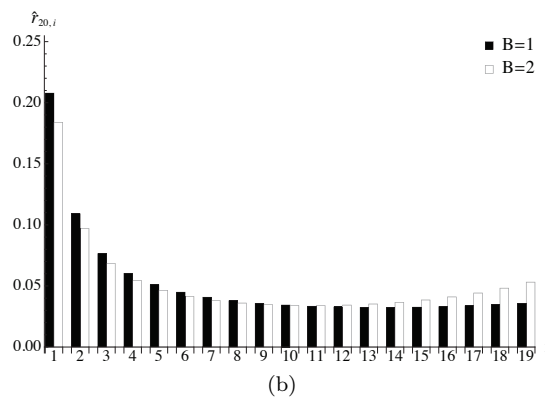
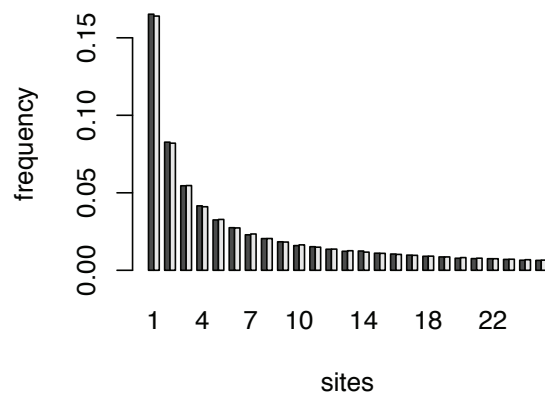


Fig. 2: (a) Simulation and theoretical prediction for the neutral relative SFS and a uniformly distributed seed bank of length $B = 10$. For the simulation of the original discrete model the population size was chosen as 1000, we started without mutations and stopped the process after 400,000 generations 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. (b) 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 $B = 2$.

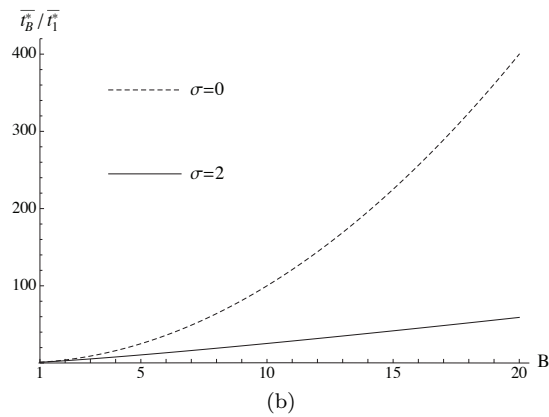
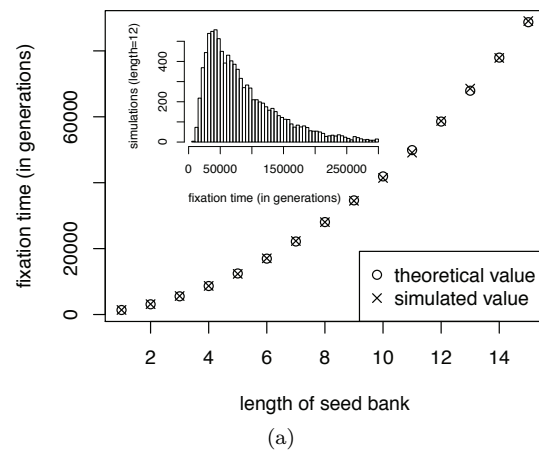


Fig. 3: (a) Simulation and theoretical prediction for the time to fixation of a seed bank model. The population size is 1000 and 50% of the individuals are initially of genotype A. We simulated 10,000 runs for each mean value. The simulated distribution of the time to fixation is shown in the histogram at the upper left corner taking the data of the simulated seed bank of length $B = 12$. (b) 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 (10). The additional index in the ratio is used to formally distinguish the cases with and without seed bank.