

Chapter 10

All But Sleeping? Consequences of Soil Seed Banks on Neutral and Selective Diversity in Plant Species



Daniel Živković and Aurélien Tellier

Abstract An axiom of modern evolutionary theory is that intra-species genetic diversity determines the adaptive potential of any species. This diversity results from the interaction between three factors: the effective population size, natural selection and the rate of recombination. All three factors are influenced by the occurrence of long dormant stages (seed banking or seed persistence), which is an evolutionary bet hedging strategy and a key characteristic of many angiosperms, but also bacteria, fungi or invertebrates. Perhaps surprisingly, this ecological trait has so far been almost ignored in evolutionary genomics. Seed banking is expected to have a fundamental influence on neutral and selective evolutionary processes, and is therefore a key factor to comprehend angiosperm genomic evolution. Theoretical modeling aims to predict the effect of seed banking on patterns of nucleotide diversity. We first adapt for seed banks the two classical mathematical frameworks of population genetics: (1) the backward in time process of the Kingman n -coalescent, and (2) the forward in time diffusion approach. This allows us to derive population genetics quantities and statistics that can be obtained from DNA sequence data. Second, we generate new predictions on neutral diversity and past demographic inference for single and multiple populations under seed banks. Third, we compute the expected effect of seed banks on unlinked (genome wide selection) and linked (gene level selection) sites. Finally, we conclude by suggesting three hypotheses, which can be tested by contrasting polymorphism data in seed banking and non-seed banking species.

10.1 Introduction

The amount of genetic diversity and polymorphism at the molecular level (in DNA) between individuals is a cornerstone of modern evolutionary theory because it determines the phenotypic diversity on which selection acts, and it is thus key

D. Živković · A. Tellier (✉)

Section of Population Genetics, Technical University of Munich, Freising, Germany

e-mail: zivkovic@bio.lmu.de; tellier@wzw.tum.de

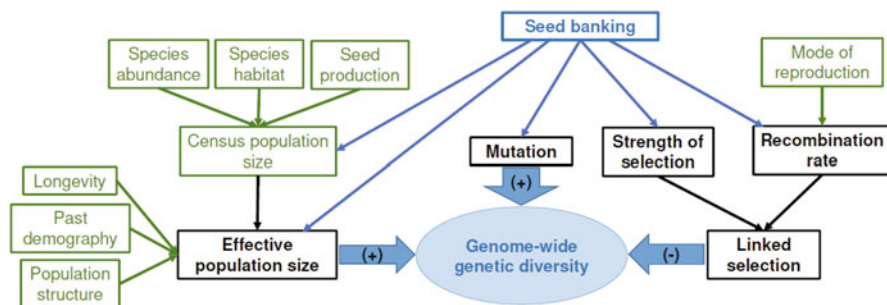


Fig. 10.1 Overview of the determinants of genetic diversity in plant species (adapted from [13]). In black the key evolutionary factors, in green the key ecological factors, and in blue the effects of seed banking. The sign on the arrows indicates the direction of correlation to genetic diversity

for predicting the adaptive potential of any species [7, 13, 40]. Our ability to predict evolutionary changes has implications for conservation biology, disease management in medicine and agriculture, plant breeding and future ecosystems management in response to global change [1, 35, 40]. Theoretical and empirical evolutionary genetics describe three main factors controlling the amount of neutral and selective genetic diversity: the species effective population size, mutation and linked selection (Fig. 10.1).

The effective population size is determined directly by species characteristics such as spatial structure, past demographic events (population increase or decrease, [35, 40]) and life span (longevity), as well as the census size, i.e. the number of observed individuals (Fig. 10.1). Several life history traits affect directly the census size, such as the size of propagules in animals [45], or in plants crucial ecological factors such as the geographic range, ecological habitat, density and abundance [47] and seed production [17].

The mutation rate is a characteristic of the species which may vary with latitude and longitude, e.g., due to exposure to UV light. It is also important to quantify if mutations occurs in the resting or dormant stages of some species: seeds for plants, eggs for insects and crustaceans (e.g. *Daphnia*), spores in bacteria and fungi [12, 26, 37, 38, 42].

Linked selection is a function of the strength of selection and the recombination rate (Fig. 10.1, [59, 60]). The strength of selection depends on the trait affected by selection and the advantage (or disadvantage) in a given environment compared to other genotypes. By linkage disequilibrium (LD, e.g. [8]) between genomic sites, natural selection affects (1) neutral genomic diversity around the targets of selection, and (2) the efficiency of selection at other neighbouring sites (or loci), the latter is termed the Hill-Robertson effect [25]. The extent of LD depends on the recombination rate with higher rates decreasing the effect of linked selection on diversity (Fig. 10.1, [11, 13]). In flowering plants recombination rate has been primarily studied as determined by the mode of reproduction (selfing, outcrossing) and the reproductive characteristics (Hermaphrodite, Dioecy or Monoecy) [9, 19, 20, 46].

Higher mutation rates and larger effective population size increase genetic diversity, while increased linked selection has the opposite effect (Fig. 10.1). The joint action of these three factors explains the occurrence of the so-called Lewontin paradox [39], stating that the census size of a species, N , defined as the number of individuals observed, does counterintuitively not correlate with genetic diversity. Census sizes can vary over several orders of magnitude between species, e.g. between elephants, humans and crustaceans, while genetic diversity varies only by two to three orders of magnitude in animals [11, 13, 36, 39, 45]. Recent studies show that such a discrepancy between expected and observed genetic diversity arises because selection counteracts the effect of a large population size on diversity: in large populations, positive and purifying selection are more efficient thus increasing the effect of linked selection and the Hill-Robertson effect [11, 13]. Resolving this paradox is instrumental to understand the adaptability of species in the current age of global climatic and anthropogenic changes [1, 36].

In plants, in particular in angiosperms (flowering plants), despite the recent progress in sequencing and the large amount of available genome data, we still do not know the relative importance of the several ecological and life history traits highlighted in Fig. 10.1 in shaping genetic diversity [11, 13, 36]. A key characteristic of many plant species is generally ignored: the ability of forming persistent seed banks with seeds remaining in the soil for many years. It is classically observed as plant adaptation to unpredictable semi-arid to desertic habitats, but is found also in many species in temperate climate in Eurasia and North-America [17, 52]. As an ecological adaptation, seed persistence has major consequences on the genomic evolution by increasing genetic diversity and also by buffering against fast ecological changes in population sizes and preventing population extinction. Seed persistence is therefore a key life-history trait linking the ecology of a species with its abundance, the production and survival of seeds over years, and consequently defining the census size above ground and the available genetic diversity (effective population size). To give a first general idea of its potential importance (see details below), let us define the germination rate, b , as the probability ($0 < b \leq 1$) that a seed will germinate on average in the next year. Theory predicts that persistent seed banks increase genetic diversity by a factor $1/b^2$ and the recombination rate by a factor $1/b$. If the germination rate is realistically small ($b < 0.5$), the effect of seed banking becomes prominent on genetic diversity and recombination rates [10, 14, 34, 62].

In this chapter we describe the effect of seed banking on shaping genetic diversity in plant species. We first build here backward and forward population genetics models which allow us to derive statistics as the expected time of coalescence, the time of fixation and the site frequency spectrum (SFS) of a sample (Sect. 10.2). We then derive a set of predictions that should be observable in polymorphism data from diverse plant species for neutral diversity (Sect. 10.3) and for natural selection (including footprints of selection in the genome, Sect. 10.4).

10.2 Model Description

Seed banking is an evolutionary strategy termed as bet hedging because it consists in reducing short-term reproductive success in favour of longer-term risk reduction to maximize individual fitness over time in variable and unpredictable environments [10, 14, 15, 54, 55]. It is also commonly found in bacteria, fungi, protozoans (including human parasites) [27, 37] and invertebrates (e.g. *Daphnia*, [42]), thus our predictions can be extended to these species. This ecological trait is classically observed as plant adaptation to unpredictable semi-arid to desertic habitats [55], but is found also in many species in temperate climate in Eurasia and North-America [2, 17]. A non-persistent seed bank is defined as seeds remaining in the soil for up to 5 years (hereafter defining non-seed banking, non-SB species), while persistent seed bank is longer (defining seed banking, SB species, [2, 17]). Note that seed bank persistence is independent of dormancy [17]. While dormancy is defined as the best timing for germination within a year, seed banking occurs over many years and seeds may alternate between periods of dormancy and non-dormant stages when in the soil.

In population genetics, it is common practice to either model an entire population forwards in time within a diffusion framework [16] or to trace a sample of a population backwards in time by applying coalescent theory [30]. In both cases a population of large and finite size N is initially assumed that evolves on a discrete generation-wise time scale and whose reproduction mechanism will basically follow a Wright–Fisher model throughout this chapter. In the backwards formulation of the usual non-SB version of this model, the entire population is replaced after reproduction and descendants pick their ancestors by random sampling from the foregoing generation, while in its SB version, ancestors can be chosen from the previous and up to m generations according to the probabilities b_1, \dots, b_m . In the dual prospective perception of time, these probabilities determine when ancestors give rise to their descendants within the subsequent m generations. We focus throughout on the average germination rate b , which is the inverse of the mean time a seed will spend in the bank, i.e. $b = 1 / \sum_{i=1}^m i b_i$. In practice, the seed bank age distribution is often assumed to follow a geometric distribution, so that seeds are more likely to germinate at earlier than later ages.

Since it is mathematically favourable to treat the ancestral (or the equivalent reproductive) process on a continuous time-scale, time is scaled in units of N generations (in a haploid model as assumed throughout) while $N \rightarrow \infty$. In the retrospective setting, for instance, the waiting times until two sampled individuals coalesce into a common ancestor can be geometrically distributed in the discrete, and exponentially distributed in the continuous model. Mutations are the main source of variation among the individuals of a population. As rare events these are assumed throughout to follow a Poisson distribution with a scaled mutation rate, $\theta = \lim_{N \rightarrow \infty} 2N\mu/b^2$, and that each of them arises on a previously unmutated (or monomorphic) site according to the infinitely-many sites model of Kimura [33]. In other words, the effective population size, N_e , consists of all plants and seeds,

while the census size is only consisting of living plants, so that we obtain the relation $N_e = N/b^2$ [28, 43]. In other words, mutations are assumed to arise in seeds and living plants [12, 58]. Alternatively, mutations may not arise in seeds, so that the scaled mutation rate reduces by a factor b [52, 62]. Except in the case of linked selection, all polymorphic sites will be assumed to be unlinked (due to recombination in between them) and therefore independent of each other.

10.2.1 The Retrospective Coalescent View for Neutrally Evolving Sites

Kaj et al. [28] introduced an urn model to describe neutral seed bank dynamics for a population of a constant size, where the population of a new generation is formed via multinomial sampling from the m previous generations in this initially discrete setting. A sample represented as balls and consisting of n balls at present is repeatedly relocated across the previous generations by sliding a window comprising the m consecutive generations as cells in a stepwise manner. Sliding the window one generation backwards, all balls from the first cell of the previous window are relocated into one of the m cells of the actual window according to the probabilities b_1, \dots, b_m . More precisely, each ball is relocated into one of the N slots of a given cell, each representing an individual of the population in the respective generation. Thereby, two types of coalescent events may occur in the sample's history: either two balls are placed into the same slot of the same cell or a ball is placed into a previously occupied slot. The probability of one coalescent event at a time is $O(1/N)$, while more than one coalescent event at a time occur with the negligible probability of $O(1/N^2)$. Therefore, coalescences occur on a 'slow' timescale of $O(N)$ steps, while the relocation process runs on a 'fast' time scale making the separation of time scales possible.

The ancestral process of the discrete seed bank model is denoted by $(A_n^N(k))_{k \geq 0}$, where $A_n^N(k)$ is the number of ancestors at step k with population size N and initial sample size n . It has been shown [28] that the continuous seed bank model is the n -coalescent [30] run on a slower time-scale by proofing that the time-rescaled ancestral process $(A_n^N([Nt]))_{t \geq 0}$ converges as $N \rightarrow \infty$ to the continuous-time Markov chain $(A_n(t))_{t \geq 0}$ with infinitesimal generator matrix $\mathbf{Q} = (q_{ij})_{i,j \in \{1, \dots, n\}}$ defined by

$$\begin{aligned} q_{ii} &= -b^2 \binom{i}{2}, & 2 \leq i \leq n, \\ q_{ii-1} &= b^2 \binom{i}{2}, & 2 \leq i \leq n, \\ q_{ij} &= 0, & \text{otherwise.} \end{aligned} \tag{10.1}$$

From Eq. (10.1) one can derive the probability that the process $A_n(t)$ is in the state of $j = n, \dots, 2$ ancestors at time t via a matrix decomposition (e.g. [61]) to obtain

$$P(A_n(t) = j) = \sum_{k=j}^n c_{nk} r_{kj} \exp\left(-b^2 \binom{k}{2} t\right), \quad (10.2)$$

where $c_{nk} = \binom{n}{k} k_{(k)} / n_{(k)}$ and $r_{kj} = (-1)^{k-j} \binom{k}{j} j_{(k-1)} / k_{(k-1)}$ are the elements of the matrices of column and row eigenvectors of \mathbf{Q} , respectively, and $a_{(b)} = a(a+1) \cdots (a+b-1)$, $a_{(0)} = 1$. The mean waiting times between coalescent events are given by

$$E(T_j) = \int_0^\infty P(A_n(t) = j) dt. \quad (10.3)$$

as the inverse of the coalescent rate.

10.2.2 The Prospective Diffusion Framework for Neutral and Selected Sites

Following allele frequencies in the limit $N \rightarrow \infty$ forward in time leads to diffusion approximations with time and allele frequencies being again measured on a continuous scale (i.e. in units of N generations). We assume two allelic types A and a with frequencies x and $1-x$. The advantage of the diffusion framework over the coalescent setting is that selection can be more straightforwardly taken into account. The effect of weak selection on the fertility of plants and on the fraction of surviving seeds can be summarized into a coefficient s , so that the scaled coefficient in the haploid model is given by $\sigma = Ns$. By making use of a perturbation approach, it has been shown [34] that in the diffusion limit, as $N \rightarrow \infty$, the probability $f(y, t) dy$ that the type- A genotype has a frequency in $(y, y+dy)$ is determined by the following forward equation (see [31] for the non-SB model):

$$\frac{\partial}{\partial t} f(y, t) = -\frac{\partial}{\partial y} \{\mu(y) f(y, t)\} + \frac{1}{2} \frac{\partial^2}{\partial y^2} \{\sigma^2(y) f(y, t)\}, \quad (10.4)$$

where the drift and the diffusion terms are, respectively, given by $\mu(y) = \sigma b y(1-y)$ and $\sigma^2(y) = b^2 y(1-y)$. For neutrality, $\mu(y) = 0$ so that the exclusive consequence of genetic drift is characterized by the diffusion term. For the derivation of the time to fixation and the SFS we require the following measures. The scale density of the diffusion process is given by

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

The speed density is obtained (up to a constant) as $\pi(y) = [\sigma^2(y)\xi(y)]^{-1}$ and the probability of absorption at $y = 0$ is given by

$$u_0(x) = \frac{\int_x^1 \xi(z) dz}{\int_0^1 \xi(z) dz}$$

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

Assuming that both $y = 0$ and $y = 1$ are absorbing states the mean time \bar{t} until one of these states is reached is given by [16]

$$\bar{t}(x) = \int_0^1 t(x, y) dy, \quad (10.5)$$

where

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

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

The time until a mutant allele is fixed conditional on fixation can be evaluated as

$$\bar{t}^*(x) = \int_0^1 t^*(x, y) dy, \quad (10.6)$$

where $t^*(x, y) = t(x, y)u_1(y)/u_1(x)$.

10.2.3 Statistical Measures for the Analysis of Genomic Data

The SFS is one of the most commonly used statistics for the analysis of genomewide distributed SNPs. It is defined as the distribution of the number of times i a mutation is observed in a population or a sample of n sequences conditional on segregation. In the coalescent setting the SFS is (either theoretically or empirically) evaluated for a sample at the present time. In the diffusion framework, the SFS of a population or a sample can often even be derived in dependence of a time variable t and an equilibrium solution can either be implied by letting $t \rightarrow \infty$ or by making use of the measures from the foregoing section. Although the forward version of the SFS has interesting applications on time-series data (i.e. DNA sequences sampled over time), we will, for convenience, focus throughout on coalescent formulations and results and only mention equilibrium solutions with reference to diffusions.

Let the site frequencies be denoted as $f_{n,i}$ ($1 \leq i \leq n-1$), where the index n shall indicate that the SFS depends on the sample size (except in the case of neutrality and a constant population size). It has been shown [22, 63] that the following equation holds for general binary coalescent trees, i.e. ancestral trees with pairwise coalescences at a time with arbitrary continuous waiting times and following the mutation model given above:

$$f_{n,i} = \frac{\theta}{2} \sum_{k=2}^{n-i+1} k \frac{\binom{n-i-1}{k-2}}{\binom{n-1}{k-1}} E(T_k). \quad (10.7)$$

Two related measures of the SFS are the expected number of segregating sites S_n equalling the total number of mutations in the infinitely-many sites model and the average number of pairwise differences Π_n . The relationships are

$$S_n = \sum_{j=1}^{n-1} f_{n,j},$$

$$\Pi_n = \frac{1}{\binom{n}{2}} \sum_{j=1}^{n-1} j(n-j) f_{n,j},$$

and the expectations of both quantities can be easily obtained via Eq. (10.7). One can also derive the normalized version of the SFS as $r_{n,i} = f_{n,i}/S_n$.

In the diffusion framework the equilibrium version of the SFS as the proportion of sites where the mutant frequency is in $(y, y+dy)$ is given by [21]

$$\hat{f}(y) = \theta \pi(y) u_0(y), \quad (10.8)$$

and the finite version can be immediately obtained via binomial sampling as

$$\hat{f}_{n,i} = \int_0^1 \binom{n}{i} y^i (1-y)^{n-1} \hat{f}(y) dy. \quad (10.9)$$

10.3 Effect of Seed Banking on Neutral Evolutionary Processes

The major effect of seed persistence is to increase genetic diversity by a storage effect of seeds in the soil. In coalescent theory terms, seed banks increase the time for two lineages to coalesce by a factor $1/b^2$ under any neutral model.

10.3.1 Constant Population Size

The expected time to coalescence is simply given by

$$E(T_j) = \left(b^2 \binom{j}{2} \right)^{-1},$$

which can, e.g., be derived by applying Eq. (10.2) to (10.3). The consideration of the average germination rate b diminishes the rate of genetic drift (Fig. 10.3, [28]) and is bounded as $1/m \leq b \leq 1$. The lower and upper bounds result from the scenarios, where all seeds, respectively, rest m and one generation in the bank. So the expected coalescent tree can be up to m^2 generations longer in the SB model compared to the usual non-SB Wright–Fisher model. Persistent seed banks, acting as a genetic storage, increase diversity by a factor $1/b^2$. For a neutral model of a constant population size, Eq. (10.7) simply yields $f_{n,i} = \theta/(b^2 i)$.

Applying diffusion theory, the population and the sample SFS can be, respectively, obtained as $\hat{f}(y) = \theta/(b^2 y)$ and $\hat{f}_{n,i} = \theta/(b^2 i)$ by applying Eqs. (10.8) and (10.9), which are equivalent to the coalescent results. The mean time to absorption and the time to fixation can be, respectively, derived via Eqs. (10.5) and (10.6) as $\bar{t}(x) = -2/b^2 (x \log(x) + (1-x) \log(1-x))$ and $\bar{t}^*(x) = -2/b^2 (1-x)/x \log(1-x)$. Therefore, the mean time to absorption/loss and the time to fixation are both slowed down in the SB model by a factor of $1/b^2$ compared to the non-SB model.

10.3.2 Estimation of Past Demographic Events

The change in coalescent rate due to SB decreases the strength of genetic drift, lengthens the time of fixation of neutral alleles and thus diminishes the genetic differentiation between populations [56]. This consequently affects the inference of past demography of a population or a species [62]. In the following, we present a simplified demographic setting that allows us to make the results of Sects. 10.2.1 and 10.2.3 applicable via a simple time rescaling argument. We assume that plants and seeds of all age classes are equivalently affected by changes in the population size such that the relative proportions of all type of seeds, and therefore the seed bank age distribution b_1, \dots, b_m and its average germination rate b remain constant over time. Furthermore, the population size changes are assumed to occur on a coalescent time scale so that the relocation process can reach an equilibrium between coalescent events as in the case of a constant population size. This particularly implies that the population size is approximately constant over any given time window, i.e. for a given m -window k_0 the relative population size function (as scaled by the population size N at present time) at the i -th cell $\rho_N(i + k_0 - 1) = N(i + k_0 - 1)/N \approx \rho_N(k_0)$. This simplification holds for a

geometrically growing population, if the growth rate and m are chosen realistically small. In the case of an instantaneous population decline, this relationship is violated, but only for $m - 1$ generations, so that for small m instantaneous changes within a window can be neglected due to the small corresponding coalescence probability for large population sizes.

In continuous time, let $\rho(t)$, which arises from $\rho_N([Nt])$ as $N \rightarrow \infty$ and time being measured in units of N generations, be piecewise continuous, bounded and follow the conventions in discrete time. The time-rescaling argument (in terms of the harmonic mean of the relative population sizes) for the coalescent approximation of the non-SB Wright–Fisher model [32], $t \rightarrow \int_0^t \rho(s)^{-1} ds$, can then be applied to the ancestral process $(A_n(t))_{t \geq 0}$ to obtain the process with time-varying population size $(A_n^\rho(t))_{t \geq 0}$. Therefore, the corresponding results to Eqs. (10.2) and (10.3) are, respectively, given by Živković and Tellier [62]

$$P(A_n^\rho(t) = j) = \sum_{k=j}^n c_{nk} r_{kj} \exp\left(-b^2 \binom{k}{2} \int_0^t \rho(s)^{-1} ds\right) \quad (10.10)$$

and

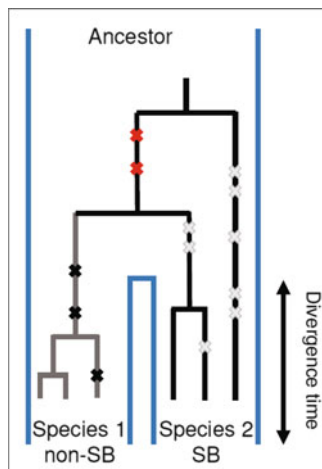
$$E(T_j^\rho) = \int_0^\infty P(A_n^\rho(t) = j) dt. \quad (10.11)$$

The SFS can be simply obtained by applying Eq. (10.11) to (10.7) and applied for the estimation of demographic parameters, which is only feasible to some extent with prior knowledge on the average germination rate b [62]. The SFS can also be obtained within the diffusion framework (e.g. via a moment based approach) to solve a time inhomogeneous-version of Eq. (10.4) with the diffusion term given by $\sigma^2(y, t) = b^2 y(1 - y)/\rho(t)$ and the drift term $\mu(y) = 0$ (see [61] for non-SB models). As an application of these results, it has been shown in a seminal study [52] that b can be estimated by applying an Approximate Bayesian Computation (ABC) method to polymorphism data with prior knowledge on metapopulation structure and N using ecological data.

10.3.3 *Seed Banks Decrease the Rate of Divergence Between Populations/Species*

As the coalescent tree is longer by a factor $1/b^2$, we also predict that the inference of recent splits between populations (or species) is affected (Fig. 10.2). Seed banking decreases the genetic differentiation, measured as F_{ST} , in spatially structured populations [49]. Note that the population migration rate should be multiplied by b if pollen only disperse, and not if pollen and seeds disperse. Ignoring seed banking thus influences the estimation of migration rates if the appropriate dispersal scaling

Fig. 10.2 In a model of speciation (without gene flow) from an ancestor into two incipient species, persistent seed banks increase the amount of incomplete lineage sorting due to the lower rate of genetic drift in species 2 (SB) compared to species 1 (non-SB). The number of shared alleles (red mutations) between species with and without seed bank is increased by seed banking compared to private alleles (black and grey crosses) (black and grey crosses)



is not taken into account. On a longer time scale, species with persistent seed banks may exhibit higher rates of shared alleles between divergent species, the so-called Incomplete Lineage Sorting (ILS), which is obscuring the phylogenetic signal of speciation (red mutations in Fig. 10.2). For example, several species in the tomato clade show evidence for seed banking [43], and high rates of ILS are found [44].

10.3.4 Seed Banks Increase the Rate of Recombination

Seed banking decreases the recombination rate per year because only seeds germinating with probability b can undergo recombination upon gamete production. However, as the overall coalescent tree is longer by a factor $1/b^2$, the net effect of longer seed persistence is to increase the amount of recombination events per locus by a factor $1/b$ (Fig. 10.3). We can thus predict that at equal census size SB species compared to non-SB species should exhibit (1) a higher genetic diversity (number of SNPs), and (2) a higher recombination rate per locus (Fig. 10.3). Using prior information on N , nucleotide diversity and recombination rate per gene, it should thus be possible to estimate b for any given species with persistent seed bank.

10.4 Effect of Seed Banking on Selective Processes

Seed banks slow down natural selection, and decrease the Hill-Robertson effect. Persistent seed banking slows down the action of positive (and purifying) selection, so that it takes longer for a selected allele to get fixed in a population [23, 34, 49]. However, we have recently demonstrated that even though selection is slower under

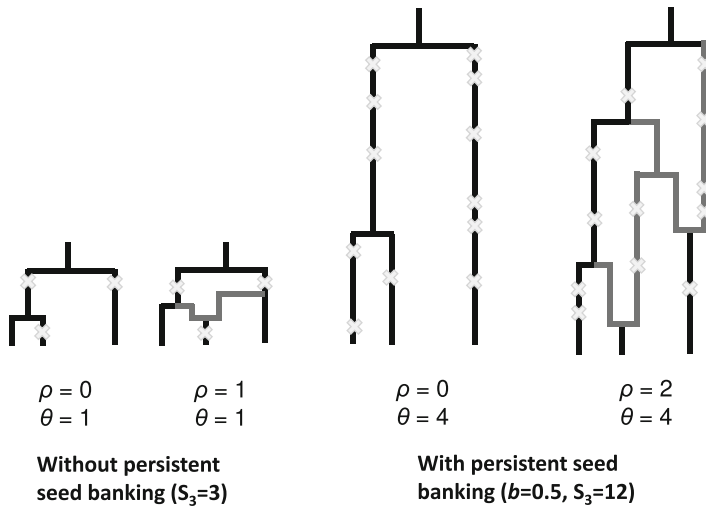


Fig. 10.3 Effect of a persistent seed bank on the length of the coalescent tree, diversity and recombination rate per gene. The comparison is made for two species of equal census size: one without ($b = 1$, left panel) and one with seed banking ($b = 0.5$, right panel). The number of segregating sites S_3 is increased by the seed bank (grey crosses). The per gene scaled recombination (ρ) and mutation (θ) rates are exemplarily given. The branches in grey are generated by recombination events

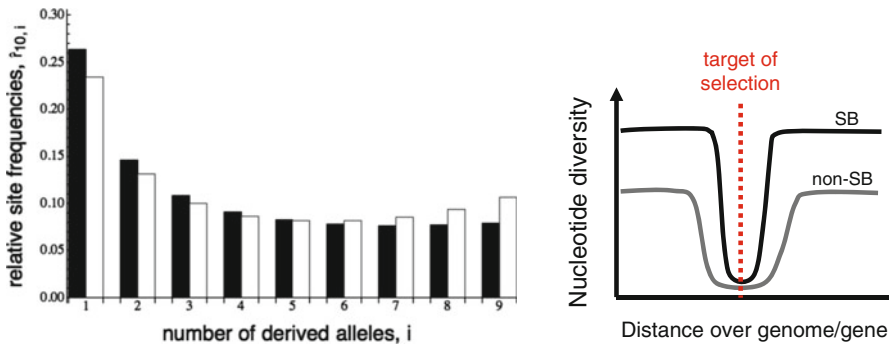


Fig. 10.4 Effect of a persistent seed bank on signatures of selection. (left panel) An increase of the strength of positive selection ($\sigma = 2$) among independent SNPs is illustrated in terms of the normalized SFS, $\hat{r}_{n,i} = f_{n,i}/S_n$; for species without (black) and with (white, $b = 0.5$) seed bank. (right panel) Persistent seed banks increase the strength of selection and recombination rates, so that narrower but deeper selective sweep signatures are expected around the target of selection in SB versus non-SB species

more persistent seed banks, the strength of selection is enhanced when observed at equilibrium (Fig. 10.4 for positive selection, [34]). This is explained by selection acting on every lineage germinating from the seed bank at rate b , whereas genetic drift acts on the time scale of b^2 .

10.4.1 Selection on Unlinked Sites

As already mentioned above, it is more straightforward to model selection within the diffusion framework than via coalescent theory. For non-SB models, a partly numerical and partly explicit solution of Eq. (10.4) can be obtained by finding a spectral representation of $f(y, t)$ for a constant population size [48] and even for piecewise population size changes [64]. Since the average germination rate b is included as a linear factor in Eq. (10.4), the underlying method can even be extended to SB models. For convenience, we will only present the equilibrium solution of the SFS for a constant population size, which anyway demonstrates the main affect of selection on seed banks.

The population and the sample SFS are, respectively, given via Eqs. (10.8) and (10.9) as

$$\hat{f}(y) = \frac{\theta}{b^2 y(1-y)} \frac{1 - \exp(-2\sigma(1-y)/b)}{1 - \exp(-2\sigma/b)}$$

and

$$\hat{f}_{n,i} = \frac{\theta n}{b^2 i(n-i)} \frac{1 - {}_1F_1(i; n; 2\sigma/b)e^{-2\sigma/b}}{1 - e^{-2\sigma/b}}.$$

For the mean time to absorption and the time to fixation one can only obtain onerous equations [34], which as well as the results for the SFS show that seed banks enhance the effect of selection (Fig. 10.4, left panel) while the considered allele takes longer to reach that equilibrium state.

10.4.2 Selection on Linked Sites

Here we propose four novel theoretical predictions which have not yet been tested with experimental data.

1. As selection is strong and recombination rates are higher in SB species, classic selective sweeps [8, 29] will exhibit a strong but reduced extent hitchhiking signature [41] around the target site (Fig. 10.4, right panel).
2. As seed banking generates higher nucleotide diversity and slows down positive selection, it is also expected that positive selection acts on standing genetic variation and incomplete sweeps (in which the selected allele has not reached fixation) should be observed (the so-called soft sweeps, [24, 57]).
3. Seed banks also favour the maintenance of polymorphism and enhance the strength of balancing selection [49, 51, 53], which should thus be more observable in SB species.

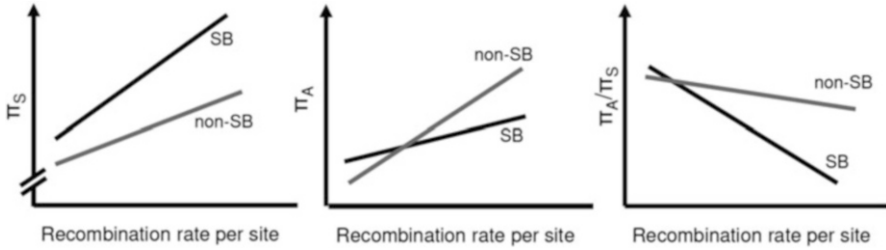


Fig. 10.5 Effect of a persistent seed bank on increasing pervasive purifying selection and decreasing the effect of background selection in the genome when comparing SB (black line) and non-SB (grey line) species. Genes with similar recombination rates are binned (see [4]). Nucleotide diversity at synonymous sites increases faster for increasing recombination rates in SB than non-SB (left panel), diversity at non-synonymous sites is driven by more efficient selection in SB versus non-SB (middle panel) predicting the ratio π_A/π_S to differ across recombination rates between SB and non-SB species (right panel)

4. The Hill-Robertson effect [25] should be reduced in outcrossing SB compared to outcrossing non-SB species. Purifying selection is also predicted to be efficient and narrow around the target sites (decreasing the extent of background selection, [5, 6]) in SB species (Fig. 10.5). This yields from the enhanced recombination rate compared to the coalescent rate and the increased strength of selection under seed banking.

Positive and purifying selection are thus predicted to be more efficient and show a reduced LD signature in SB than in non-SB species. In SB species, positive selection should be pervasive across the genome, signatures of classic selective sweeps would be narrow around the target site (Fig. 10.4), and the extent of purifying linked (background) selection should be limited (Fig. 10.5). More selective sweeps are also expected to arise from standing genetic variation and/or be incomplete in SB compared to non-SB species.

10.5 Conclusions

We have built here a model of a so-called weak seed bank for plants species, but also applicable to fungi or invertebrate species producing dormant resting stages. An assumption that we have used in both of our models is namely that the maximum time seeds can spend in the bank, m , is bounded and small compared to the time of coalescence of lineages. This condition ensures that our coalescent model converges to the n -Kingman coalescent [28], and that the separation of time scale and the Markovian property can be used to analyse the diffusion model. Seed banking is thus a unique life-history trait which links ecology and genetic diversity defining the Lewontin paradox in angiosperms. Following the hypotheses and predictions above (summarized in Table 10.1), we predict that SB species exhibit smaller census sizes,

Table 10.1 Summary of hypotheses and predictions

Hypothesis	Predictions
Seed banking increases neutral diversity and the rate of recombination	Seed banking species exhibit more diversity and higher recombination rates per locus
Seed banking enhances the efficiency of selection and decreases the Hill-Robertson effect	Signatures of selection are enhanced, selection is more pervasive, and footprints of selection narrower in genomes of SB species
Seed banking enhances the time of allele fixation and decreases genetic differentiation	Seed banking decreases the accuracy of past demographic inference and increases the amount of Incomplete Lineage Sorting and incongruence in phylogenies

i.e. species abundance, but much higher genetic diversity than non-SB species. This could be termed as an inverse Lewontin paradox for plant species when comparing SB and non-SB species, which generates the opposite pattern of diversity versus census size than found in animals. As a corollary, we predict that selfing non-SB species exhibit large census sizes but show the lowest genetic diversity compared to SB species.

An alternative model has been proposed, the so-called strong seed bank, where the time m can be infinite or at least very large compared to coalescent times [3]. These models produce very different predictions regarding the SFS and the effect of recombination, because they do not consist in a rescaled version of the Kingman n -coalescent. Strong seed banks can generate so-called multiple merger coalescents [3, 50] with signatures differing from our classic results for neutral sites. It is suggested that these models may be better adapted to bacteria species with resting stages that can survive for many years (especially for soil bacteria [37]), which is a much larger amount of time compared to their generation time [18]. As more and more full genome intra-species polymorphism data are becoming available in bacteria, fungi, invertebrates (*Daphnia*, Mosquito) and plants, it is a fascinating prospect to test the derived predictions and to assess which of the weak or strong seed bank models applies to a given species.

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

References

- Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* 11:697–709
- Baskin CC, Baskin JM (2014) *Seeds: ecology, biogeography, and evolution of dormancy and germination*, 2nd edn. Elsevier, Amsterdam
- Blath J, González-Casanova A, Kurt N, Spano D (2013) On the ancestral process of long-range seedbank models. *J Appl Probab* 50:741–759

4. Campos JL, Halligan DL, Haddrill PR, Charlesworth B (2014) The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Mol Biol Evol* 31:1010–1028
5. Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303
6. Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* 70:155–174
7. Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 10:195–205
8. Charlesworth B, Charlesworth D (2010) Elements of evolutionary genetics. Roberts & Company Publishers, Greenwood Village
9. Chen J, Glémin S, Lascoux M (2017) Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol Biol Evol* 34:1417–1428
10. Cohen D (1966) Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* 12:119–129
11. Corbett-Detig RB, Hartl DL, Sackton TB (2015) Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol* 13:e1002112
12. Dann M, Bellot S, Schepella S, Schaefer H, Tellier A (2017) Mutation rates in seeds and seed-banking influence substitution rates across the angiosperm phylogeny. *bioRxiv*. <https://doi.org/10.1101/156398>
13. Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nat Rev Genet* 17:422–433
14. Evans MEK, Dennehy JJ (2005) Germ banking: bet-hedging and variable release from egg and seed dormancy. *Q Rev Biol* 80:431–451
15. Evans MEK, Ferriere R, Kane MJ, Venable DL (2007) Bet hedging via seed banking in desert evening primroses (*Oenothera*, *Onagraceae*): demographic evidence from natural populations. *Am Nat* 169:184–194
16. Ewens WJ (2004) Mathematical population genetics: I. Theoretical introduction. Springer, Berlin
17. Fenner M, Thompson K (2004) The ecology of seeds, Cambridge University Press, Cambridge
18. González-Casanova A, Aguirre-von-Wobeser E, Espín G, Servín-González L, Kurt N, Spanò D et al Strong seedbank effects in bacterial evolution. *J Theor Biol* 356:62–70 (2014)
19. Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV et al Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol* 27:1822–1832 (2010)
20. Gossmann TI, Keightley PD, Eyre-Walker A (2012) The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol* 4:658–667
21. Griffiths RC (2003) The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theor Popul Biol* 64:241–251
22. Griffiths RC, Tavaré S (1998) The age of a mutation in a general coalescent tree. *Stoch. Models* 14:273–295
23. Hairston Jr NG, De Stasio Jr BT (1988) Rate of evolution slowed by a dormant propagule pool. *Nature* 336:239–242
24. Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352
25. Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genet Res* 8:269–294
26. Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination: Tansley review. *New Phytol* 171:501–523
27. Jones SE, Lennon JT (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci* 107:5881–5886
28. Kaj I, Krone SM, Lascoux M (2001) Coalescent theory for seed bank models. *J Appl Probab* 38:285–300

29. Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking long a recombining chromosome. *Genetics* 160:765–777
30. Kingman JFC (1982) On the genealogy of large populations. *J Appl Probab* 19A:27–43
31. Kimura M (1955) Stochastic processes and distribution of gene frequencies under natural selection. In: Cold spring harbor symposia on quantitative biology, vol 20. Cold Spring Harbor Laboratory Press, pp 33–53
32. Kimura M (1955) Random genetic drift in multi-allelic locus. *Evolution* 9:419–435
33. Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61:893–903
34. Koopmann B, Müller J, Tellier A, Živković D (2017) Fisher-Wright model with deterministic seed bank and selection. *Nat Rev Microbiol* 114:29–39
35. Lande R (1988) Genetics and demography in biological conservation. *Science* 241:1455–1460
36. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A et al (2012) Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol* 10:e1001388
37. Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 9:119
38. Levin DA (1990) The seed bank as a source of genetic novelty in plants. *Am. Nat.* 135:563–572
39. Lewontin RC The genetic basis of evolutionary change. Columbia University Press (1974)
40. Lynch M, Lande R (1998) The critical effective size for a genetically secure population. *Anim. Conserv.* 1:70–72
41. Maynard-Smith J, Haigh J (1974) Hitch-hiking effect of a favorable gene. *Genet Res* 23:23–35
42. Möst M, Oexle S, Markova S, Aidukaite D, Baumgartner L, Stich HB et al (2015) Population genetic dynamics of an invasion reconstructed from the sediment egg bank. *Mol Ecol* 24:4074–4093
43. Nunney L (2002) The effective size of annual plant populations: the interaction of a seed bank with fluctuating population size in maintaining genetic variation. *Am Nat* 160:195–204
44. Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol* 14:e1002379
45. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A et al (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263
46. Roselius K, Stephan W, Städler T (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171:753–763
47. Salguero-Gómez R (2017) Applications of the fast-slow continuum and reproductive strategy framework of plant life histories. *New Phytol* 213:1618–1624
48. Song YSS, Steinrücken M (2012) A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics* 190:1117–1129
49. Templeton AR, Levin DA (1979) Evolutionary consequences of seed pools. *Am Nat* 114:232–249
50. Tellier A, Lemaire C (2014) Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *Mol Ecol* 23:2637–2652
51. Tellier A, Brown JKM (2009) The influence of perenniality and seed banks on polymorphism in plant-parasite interactions. *Am Nat* 174:769–779
52. Tellier A, Laurent SJ, Lainer H, Pavlidis P, Stephan W (2011) Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proc Natl Acad Sci USA* 108:17052–17057
53. Turelli M, Schemske DW, Bierzychudek P (2001) Stable two-allele polymorphisms maintained by fluctuating fitnesses and seed banks: protecting the blues in *Linanthus parryae*. *Evolution* 55:1283–1298
54. Venable DL (1989) Modeling the evolutionary ecology of seed banks. In: Leck MA (ed) *Ecology of soil seed banks*. Elsevier, Amsterdam, pp 67–87
55. Venable DL, Lawlor L (1980) Delayed germination and dispersal in desert annuals: escape in space and time. *Oecologia* 46:272–282

56. Vitalis R, Glémin S, Olivieri I (2004) When genes go to sleep: the population genetic consequences of seed dormancy and monocarpic perenniality. *Am Nat* 163:295–311
57. Vy HMT, Kim Y (2015) A Composite-likelihood method for detecting incomplete selective sweep from population genomic data. *Genetics* 200:633–649
58. Waterworth WM, Footitt S, Bray CM, Finch-Savage WE, West CE (2016) DNA damage checkpoint kinase ATM regulates germination and maintains genome stability in seeds. *Proc Natl Acad Sci* 113:9647–9652
59. Wright SI, Andolfatto P (2008) The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu Rev Ecol Evol Syst* 39:193–213
60. Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol* 22:506–519
61. Živković D, Stephan W (2011) Analytical results on the neutral non-equilibrium allele frequency spectrum based on diffusion theory. *Theor Popul Biol* 79:184–191
62. Živković D, Tellier A (2012) Germ banks affect the inference of past demographic events. *Mol Ecol* 21:5434–5446
63. Živković D, Wiehe T (2008) Second-order moments of segregating sites under variable population size. *Genetics* 180:341–357
64. Živković D, Steinrücken M, Song YSS, Stephan W (2015) Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. *Genetics* 200:601–617