Seed Banks Alter the Rate and Direction of Molecular Evolution in *Bacillus subtilis*

William R. Shoemaker[1,2*], Evgeniya Polezhaeva[1], Kenzie B. Givens[1,3], Jay T. Lennon[1*]

1 Department of Biology, Indiana University, Bloomington, IN, 47405, USA
2 Current affiliation: Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA, 90095, USA
3 Current affiliation: Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington IN, 47408, USA

***Contact: williamrshoemaker@gmail.com, lennonj@indiana.edu**

Abstract

Fluctuations in the availability of resources constrains the growth and reproduction of individuals, which in turn effects the evolution of their respective populations. Many organisms are able to respond to fluctuations by entering a reversible state of reduced metabolic activity, a phenomenon known as dormancy. This pool of dormant individuals (i.e., a seed bank) does not reproduce and is expected to act as an evolutionary buffer, though it is difficult to observe this effect directly over an extended evolutionary timescale. Through genetic manipulation, we analyze the molecular evolutionary dynamics of *Bacillus subtilis* populations in the presence and absence of a seed bank over 700 days. We find that the ability to enter a dormant state increases the accumulation of genetic diversity over time and alters the trajectory of mutations, findings that are recapitulated using simulations based on a simple mathematical model. While the ability to form a seed bank does not alter the degree of negative selection, we find that it consistently alters the direction of molecular evolution across genes. Together, these results show that the ability to form a seed bank affects the direction and rate of molecular evolution over an extended evolutionary timescale.

Introduction

Nature is rarely static. Temporal fluctuations in abiotic and biotic environmental factors often reduce the rate that an organism can grow and reproduce. To contend with such fluctuations, many species enter a reversible state of reduced metabolic activity, an adaptation known as dormancy [1]. In this state, individuals can endure environmental stressors until they subside, a temporary cessation of short-term reproductive efforts in favor of long-term reproductive gains. This evolutionary trade-off, and the life-history strategies through which it is implemented, have received substantial attention by means of theoretical [2–6] and empirical investigations [7–9], spurred in-part by the observation that dormancy has independently evolved multiple times across the tree of life [10–12].

While the trade-off aspect of dormancy has been of considerable historical interest, life-history traits do not operate in a population genetic vacuum. The fitness benefit of a life-history trait is often a consequence of its effect on birth-death processes [13], population dynamics that sequentially alter the dynamics and fates of genetic variation. The ability to enter a dormant state is no exception. The accumulation of dormant

individuals within a system can lead to the formation of seed banks [10], demographic structures that can reshape the molecular evolutionary dynamics of a population.

The presence of a seed bank primarily alters the molecular evolutionary dynamics of a population through two means. First, the ability to enter a dormant state dampens the accumulation of *de novo* genetic diversity and its subsequent fluctuations, as dormant individuals do not reproduce and the vast majority of mutations are typically acquired during the process of genome replication [1, 14]. Second, seed banks can act as reservoirs of genetic and phenotypic diversity. These reservoirs reduce the efficiency of natural selection [15–17], dampen the loss of genetic diversity due to random genetic drift [17–20], and allow for temporary deleterious variants to be retained until environmental changes occur [21, 22]. Alternatively stated, the presence of a seed bank reduces the rate of molecular evolution while increasing the maximum amount of genetic diversity that can be retained. Finally, because the ability to form a seed bank is the result of a life-history strategy maintained by natural selection, it is possible that the formation of seed banks restricts the targets of molecular evolution, constraining the direction of evolution as well as its rate. While substantial progress has been made towards developing mathematical models that describe these patterns within the discipline of theoretical population genetics [16, 17, 17–20], there remains a comparative lack of empirical data necessary to evaluate and test their central predictions.

The role of seed banks as an evolutionary buffer means that time is an essential factor when considering an empirical system. Namely, if the per-generation rate of change of genetic diversity is reduced by a given amount it is necessary to observe an additional proportionate number of generations. This constraint makes it challenging to directly observe seed bank dynamics over extended evolutionary timescales. Given their short generation times, large population sizes, propensity for rapid adaptation, and the prevalence of dormancy among their lineages [23], microorganisms are the ideal group of organisms to characterize the extent that dormancy alters molecular evolutionary dynamics. In addition, certain lineages of microorganisms have evolved the ability to form complex protective structures (i.e., endospores) that allow them to survive and form long-lasting seed banks [24, 25]. While these structures are not the only means through which microorganisms can enter a dormant state [26], their existence provides a means through which the formation of seed banks can be genetically manipulated. Furthermore, while evolution experiments have been performed using dormancy-capable microorganisms [27], questions pertaining to dormancy have primarily been restricted to examining the phenotypic decay of endospore formation via the acquisition of *de novo* mutations under relaxed selection [28–30], whereas the effect of endospore formation on the molecular evolutionary dynamics of microorganisms remain unexplored.

In this study, we examine the molecular evolutionary dynamics of *Bacillus subtilis* populations that differ in their ability to form protective endospores, a non-reproductive structure that is the primary mechanism through which this species enters a dormant state to form a seed bank. Replicate populations were maintained for over 700 days, generating a molecular fossil record which was reconstructed to determine how the presence of a seed bank altered the trajectories of *de novo* mutations. We then recapitulated the dynamics we observed using simulations based on a stochastic model of molecular evolution in dormancy-capable populations. Finally, we identified the sets of genes that were enriched for mutations within each transfer-regime for dormancy-capable and incapable populations, allowing us to quantify parallel evolution among replicate populations as well as the degree of divergent evolution between dormancy-capable and incapable populations.

Materials and Methods

B. subtilis Δspo0A mutant construction

To manipulate endospore formation we deleted *spo0A*, the master regulatory gene for sporulation pathways in *B. subtilis*. Gene deletion was performed using Gibson assembly and PCR amplified dsDNA fragments upstream and downstream of *spo0A*. Purified ligated plasmid was transformed into *E. coli* DH5α and plasmid DNA was purified from cultured positive transformants. Purified plasmid product was injected into *E. Coli* TG1, positive transformants were selected, and plasmid DNA was purified before a single *B. subtilis* NCIB 3610 colony was grown in inoculate containing a purified plasmid aliquot and identified using antibiotic plating. Transformation was confirmed via PCR and loss of antibiotic resistance was confirmed via antibiotic plating (see Supplemental Methods for additional detail).

B. subtilis Δspo0A evolution experiment

Fitness assay

We performed fitness assays to determine the degree that the ability to form a protective endospore provided a fitness advantage across energy-limited environments. At late-exponential phase, aliquots of *B. subtilis* WT and *B. subtilis* Δ*spo0A* cultures were transferred to three replicate flasks with new media at an equal ratio. Flasks were periodically plated over 100 days in duplicate for a range of serial dilutions. Colonies were distinguishable by their morphology, providing estimates of WT (N_{WT}) and Δ*spo0A* ($N_{\Delta spo0A}$) population sizes. The relative log fitness after t days was defined as

$$X(t) \equiv \log\left[\frac{N_{\Delta spo0A}(t)}{N_{\mathrm{WT}}(t)} \cdot \frac{N_{\mathrm{WT}}(0)}{N_{\Delta spo0A}(0)}\right] \tag{1}$$

Because we t taking measurements from a population that will inevitably go extinct in an increasingly energy-limited environment and we do not know the number of generations that occur after exponential growth, we chose to examine X at each given time point rather than use the typical fitness per-unit time estimate $\Delta X = \frac{1}{\Delta t} \cdot X(t)$.

Transfer protocol, sequencing, and variant calling

Energy-limitation was manipulated by extending the time between transfers for microbial populations. We performed our energy-limited evolution experiments using *Bacillus subtilis* NCIB 3610 Δ*spo0A* (ASM205596v1) using a previously described methodology [25]. To briefly summarize, a single colony was isolated from each strain and grown overnight in 10 mL of PYE media with 0.2% glucose and 0.1% casamino acids in a 50 mL Erlenmeyer flask and split into replicate populations. Five replicates were transferred as 1 mL aliquots into 9 mL of medium every 1, 10, or 100 days for 700 days at 25°C and 250 RPM. All replicate populations were cryopreserved in 20% glycerol at -80°C every 100 days. Populations were regularly plated to test for contamination. An identical experiment was concurrently run with the WT strain, which was previously described [31].

Under our chosen transfer regime with 1:10 dilutions and assuming that the decline in population size is negligible, replicate populations evolved for $\log_2(N_f/N_i) \approx 3.3$ generations per-transfer, with a cumulative ∼3,300, 330, and 30 generations for 1, 10, and 100-day transfer regimes, respectively. However, that is not necessarily the case as population size can decline by several orders of magnitude depending on the transfer regime, meaning that the number of generations per-transfer can be substantially higher than ≈ 3.3. In addition, CFU counts suggest that Δ*spo0A* has a different

time-dependent net rate of growth than the wild type, meaning that 10 and 100-day transfer regimes of those two strains undergo a slightly different number of generations each transfer. While we cannot account for any generations that occur within a transfer, we can account for the change in population size after resource replenishment for each transfer regime and strain to get a more accurate estimate of the timescale of the experiment. However, an unknown number of generations likely occurred while populations remained in stationary phase due to cells using dead cells as a nutrient source. This "cryptic growth" [25, 31–33] suggests that the estimated number of generations likely represents the true number of generations for 1-day transfers, while it is closer to an estimate of the minimum number of generations for 10 and 100-day transfers.

DNA extraction, library preparation, and pooled population sequencing was performed on all Δ*spo0a* timepoints for all replicate populations as previously described [31]. The first 20 bp of all reads were trimmed and all read pairs where at least one pair had a mean Phred quality less than 20 were removed `cutadept v1.9.1` [34]. Candidate variants were identified using a previously published approach [35] that relied on alignments generated from `breseq v0.32.0` [36], which was modified as previously described [31].

Mutation trajectory analyses

We estimated the frequency of the mth mutation candidate in the pth population at the tth timepoint using the naive estimator $\hat{f}_{pmt} \equiv A_{pmt}/D_{pmt}$, where A_{pmt} and D_{pmt} are the total number of reads containing the alternate allele and the total depth of coverage, respectively. We examined the accumulation of mutations by time t as the sum of derived allele frequencies

$$M(t) \equiv \sum_m \hat{f}_{pmt} \tag{2}$$

Given that the log of $M(t)$ over time often appeared to saturate in this study as well as in previous studies [31, 35], we modelled the relationship between $M(t)$ and t using the following equation.

$$\log_{10} M(t^*) = \log_{10} M(t^* = 0) + [\log_{10} M]_{\max} \frac{t^*}{t^*_{1/2} + t^*} \tag{3}$$

Where $[\log_{10} M]_{\max}$ is the maximum value of $\log_{10} M(t^*)$ and $t^*_{1/2}$ is the value of t^* where $\log_{10} M(t^*)$ is half of $[\log_{10} M]_{\max}$. The variable t^* represents the shift in time so that Eq.3 reduces to the intercept parameter ($\log_{10} M(0)$) at the first temporal sample, in this case, $t^* = t - 100$ days. We then multiplied $t_{1/2}$ by the estimated minimum number of per-day generations, the product of which we define as $\tau_{1/2}$. While this model is phenomenological in that we do not posit a microscopic mechanism, much like the Michaelis–Menten kinetic model from which it is derived [37], it effectively captures the hyperbolic pattern. Numerical optimization was performed over 54 initial conditions using the BFGS algorithm in Python using `statsmodels` [38].

We examined three different measures to determine how the ability to enter a dormant state affected molecular evolutionary dynamics, defined as

$$f_{\max} \equiv \max\left(\{f(t) : t = 1, ..., T\}\right) \tag{4a}$$

$$\frac{|\Delta f|}{\Delta\tau} \equiv \frac{|f(\tau + \Delta\tau) - f(\tau)|}{\Delta\tau} \tag{4b}$$

$$Qf(\tau) \equiv \frac{f(\tau + \Delta\tau)}{f(\tau)} \tag{4c}$$

First, $f_{\max}$ is the maximum estimated frequency of a given mutation over T observations. Second, $|\Delta f|/\Delta\tau$ is the magnitude of change in f between two observations. Finally, $Qf(\tau)$ is the direction of change for f between two observations, where Q corresponds to the quotient. We compared the empirical cumulative distribution functions of WT and Δ*spo0A* for all three measures in Eq.4 for each transfer time using the Kolmogorov–Smirnov test. To identify fixed mutations, we using a previously published hidden Markov model [35] to infer whether a given mutation eventually became fixed within a replicate population over the course of the experiment. The ratio of nonsynonymous to synonymous polymorphic mutations pN/pS was calculated in each population, where the total number of observed mutations of each class was weighted by the relative frequency of nonsynonymous and synonymous sites in all genes in the genome.

Parallelism at the gene level

We identified potential targets of selection by examining the distribution of nonsynonymous mutations across genes using a previously published approach [35]. To briefly summarize, gene-level parallelism was assessed by calculating the *multiplicity* of each gene as

$$m_i = n_i \cdot \frac{\overline{L}}{L_i} \tag{5}$$

where n_i and L_i is the number of mutations observed and the length of the ith gene and $\overline{L}$ is the mean length of all genes. Under this definition, the null hypothesis is that all genes have the same multiplicity $\overline{m} = n_{tot}/N_{genes}$. Using the observed and expected values, we can quantify the net increase of the log-likelihood of the alternative hypothesis relative to the null

$$\Delta\ell = \sum_i n_i \log\left(\frac{m_i}{\overline{m}}\right) \tag{6}$$

where significance was assessed using permutation tests. To identify specific genes that are enriched for mutations, we calculated the P-value of each gene as

$$P_i = \sum_{n \geq n_i} \frac{\left(\frac{n_{tot} L_i}{\overline{L} N_{genes}}\right)^n}{n!} e^{-\frac{n_{tot} L_i}{\overline{L} N_{genes}}} \tag{7}$$

where FDR correction was performed by defining a critical P-value (P^*) based on the survival curve of a null Poisson distribution. We then defined the set of significant genes for each strain-transfer combination for $\alpha = 0.05$ as:

$$I = \{i : P_i \leq P^*(\alpha)\} \tag{8}$$

(Con/di)vergence at the gene level

We tested for divergent/convergent evolution by examining the overlap of I between WT and Δ*spo0A* populations and populations from all three transfer regimes and comparing them to a null hypergeometric distribution as previously described [31]. To examine convergent/divergent evolution among enriched genes, we calculated the vector of relative multiplicities ($\mathcal{M}_i = m_i / \sum m_i$) and compared the mean absolute difference between I genes for a given pair of transfer regimes or genetic backgrounds as

$$\langle \Delta \mathcal{M} \rangle = \frac{1}{I} \sum_{i=1}^{I} \left| \mathcal{M}_i^1 - \mathcal{M}_i^2 \right| \tag{9}$$

Null distributions of $\langle \Delta \mathcal{M} \rangle$ were generated by constructing a gene-by-strain mutation count matrix for each transfer regime and randomizing combinations of mutation counts constrained on the total number of mutations acquired within each gene across strains and the number of mutations acquired within each strain. Randomization was performed 10,000 using a Python implementation of the ASA159 algorithm and observed values of $\langle \Delta \mathcal{M} \rangle$ were standardized relative to the null distribution ($Z_{\langle \mathcal{M} \rangle}$) [39, 40]. Finally, we identified genes that were preferentially enriched in one background or transfer regime over the other using the Skellam distribution as previously described [41].

Simulating evolution with a seed bank

We performed simulations to determine whether the empirical patterns of genetic diversity we observed were consistent with outcomes expected from seed bank effects. Given that a population consists of N active and M dormant individuals, we can model dormancy through a per-capita probability of entering the seed bank of c and a per-capita probability of exiting the seed bank of $c \cdot K$, where $K = N/M$ so that that the size of the active and dormant portions of the population remain constant through time. Active individuals are then subject to resuscitation, dormancy, mutation, selection, and drift while seed banks are only subject to resuscitation and dormancy.

Given that the rapid evolution we observed was likely driven by selection, we focused on a distribution of fitness effects with beneficial mutations $\rho(s)$ that followed an exponential distribution with mutations occurring at a per-individual rate U_b. Though, ultimately, the rate that *de novo* mutations accumulate during adaptation is mediated by the distribution of fitness effects [42], we note that there is no *a priori* reason to suspect that the removal of endospore formation would alter the distribution of fitness effects of *B. subtilis*. The following forward-in-time master equation was simulated

$$
\begin{aligned}
p_N(n, X, t+dt) = & \underbrace{(1-c)(1+X-\overline{X}(t))n \cdot dt\, p_N(n-1, X, t)}_{\text{Pr [birth]}} \\
& + \underbrace{(1-c)n \cdot dt\, p_N(n+1, X, t)}_{\text{Pr [death]}} \\
& + (1-c)nU_b dt \\
& \times \underbrace{\int_0^\infty ds \rho(s) p_N(1, X, t) p_N(n-1, X-s, t)}_{\text{Pr [mutation]}} \\
& - \underbrace{n \cdot c \cdot dt\, p_N(n+1, X, t)}_{\text{Pr [dormancy]}} \\
& + \underbrace{m \cdot c \cdot K \cdot dt\, p_M(m, X, t)}_{\text{Pr[resuscitation]}} \\
& + \underbrace{[1 - ndt(3 - c + X - \overline{X}(t) + U_b) p_N(n, X, t)]}_{\text{Pr [nothing]}} \\
p_M(m, X, t+dt) = & \underbrace{n \cdot c \cdot dt\, p_N(n+1, X, t)}_{\text{Pr [dormancy]}} \\
& - \underbrace{m \cdot c \cdot K \cdot dt\, p_M(m+1, X, t)}_{\text{Pr[resuscitation]}} \\
& + \underbrace{[1 - m \cdot c \cdot K \cdot dt\, p_M(m+1, X, t)]}_{\text{Pr [nothing]}}
\end{aligned}
\tag{10a}
$$

These coupled master equations are extensions of previous theoretical efforts to characterize adaptation dynamics in asexual microbial populations [43], where we have incorporated the seed bank and resuscitation and dormancy dynamics. Simulations were performed with $U_b = 10^{-4}$/indiv.. While the true value of U_b is unknown, we elected for a value that was on the order of magnitude of 10% of the total mutation rate (all non-lethal mutations) obtained from a previously published mutation accumulation experiment [44]. The scale parameter of $\rho(s)$ was set to 10^{-2}. Simulations were run for 3,300 generations with $N = 10^6$ for $c = 10^{-5}$ and values of M ranging from $10^1 - 10^6$. Only values of M were manipulated as the same transition rates can be obtained by manipulating c or K. Ten replicate simulations were performed for each value of M. All simulations were performed using custom Python scripts.

Results and Discussion

By reconstructing the molecular fossil record of our experiment, we were able to examine the trajectories of *de novo* mutations for all replicate populations. The molecular evolutionary dynamics of *B. subtilis* largely followed our predictions, as the presence of a seed bank increased the maximum amount of genetic diversity retained by a population and decreased its rate of accumulation. Measures that capture distinct features of mutation trajectories were also largely consistent with our predictions, results that were validated with forward-time population genetic simulations. By comparing gene mutation counts between transfer-strain combinations, we determined that the ability to enter a dormant state radically altered the targets of molecular evolution, though the strength and direction of the effect was environment-dependent.

The results of this long-term experiment test, confirm, and challenge long-standing hypotheses regarding the effect of seed banks on the dynamics of molecular evolution.

The effect of seed banks on the accumulation and fate of *de novo* mutations

Despite the inability to form protective structures, all replicate populations of Δ*spo0A* from all transfer regimes survived the experiment. This pattern of consistent survival is a marked difference from a previous experiment that used an identical energy-limitation design, where bacterial taxa from diverse phyla (Bacteroidetes, Proteobacteria, and Deinococcus-Thermus) frequently went extinct in 10 and 100-day transfer regimes [31]. This comparison suggests that *Bacillus* is exceptionally capable of surviving and evolving in harsh environments, even without access to what is generally considered its primary survival strategy.

Through pooled population sequencing, we reconstructed the trajectories of *de novo* mutations for all replicate populations from all transfer regimes (Fig. S1-2). By examining the sum of derived allele frequencies at a point in time ($M(t)$) we were able to examine how *de novo* mutations accumulated over the course of the experiment for all replicate population (Fig. 1a-c, S3). As expected, populations capable of forming a seed bank tended to have higher value of $M(t)$ in the 1 and 100-day transfer regimes, though this difference was not as clear for the 10-day regime. Seed banks can effect these diversity accumulation curves through two ways: 1) increasing the amount of diversity that can be retained and 2) reducing the rate that diversity accumulates.

The observation that values of $M(t)$ tended to asymptote for Δ*spo0A* and WT replicate populations spurred the development of a phenomenological model with two intuitive parameters (Eq. 3) that can be used to test our predictions, the maximum value of the logarithm of $M(t)$ ($[\log_{10} M]_{\max}$) and the number of generations required for $\log_{10} M$ to reach half of its maximum value ($\tau_{1/2}$). After obtaining estimates of these parameters via numerical optimization, we found that $[\log_{10} M]_{\max}$ remained fairly constant as the time between transfers increased for the WT, but sharply decreased for Δ*spo0A* (Fig. 1d). In contrast, $\tau_{1/2}$ remained constant for Δ*spo0A*, but decreased as transfer time increased (Fig. 1e). This pattern confirms our prediction that the rate of accumulation of genetic diversity will be higher in populations that cannot form a seed bank. While the trends we observed were generally robust, the error in our estimates for the 100-day transfer regime was considerable (Table S2). This error was likely a result of the small number of mutations acquired among populations in the 100-day treatment as well as their small population sizes, increasing the variance in our estimates of $M(t)$. Regardless, parameter estimates consistently change in directions predicted by the seed bank effect.

Given seed banks altered the accumulation of genetic diversity, we examined how the presence of a seed bank altered the fate of a given mutation. The presence of seed banks reduce the efficiency of selection and retain genetic diversity [15], suggesting that a given mutation would have a lower probability of extinction as well as a lower substitution rate [16, 17]. Our estimates of the probability of extinction for each replicate population provide little evidence to support this claim, as there is substantial variation across replicate populations for a given strain-transfer combination (Fig. 1f). While the rate of fixation was similar for WT and Δ*spo0A* in the 1-day regime, as expected given that the ability to form a seed bank does not contribute to survival on that time scale, few fixation events occurred within replicate populations in the 10-day and 100-day transfer regimes (Fig. 1g). This paucity of fixations meant that the quantity could not be estimated, and comparisons could not be performed. While the direction of selection could be evaluated by examining the proportion of nonsynonymous

to synonymous mutations (pN/pS), estimates of pN/pS were consistently less than one across transfer regimes and there was no difference in pN/pS values between WT and Δ*spo0A* populations within a given transfer regime (Fig. S4). This lack of difference suggests that purifying selection was predominant regardless of whether a population was capable of entering the seed bank or in environments where the ability to enter a dormant state would be favorable.

Seed banks alter the dynamics of segregating mutations

While seed banks had a clear effect on the accumulation of *de novo* genetic diversity, though its effect on the fates of mutations was not as straightforward. Measures such as the substitution rate and the probability of extinction are advantageous in that their value in the presence of a seed bank can be compared to predictions from existing mathematical models [16, 17, 19]. However, their emphasis on the final state of a given mutation (i.e., fixation or extinction) means that such measures do not necessarily capture the dynamics mutations exhibit over their sojourn times. To circumvent this limitation, we leveraged the temporal structure of the data by examining three measures of each mutation trajectory: 1) the maximum frequency that a mutation reached (f_{max}), 2) the set of per-generation magnitudes of frequency changes ($\frac{|\Delta f|}{\Delta \tau}$) over time, and 3) the set of changes in the direction of allele frequency changes between observations over time (Qf). By calculating survival distributions for each empirical measure for each strain-treatment combination, we can determine the degree that seed banks altered the molecular evolutionary dynamics of *B. subtilis*.

Across transfer regimes, Δ*spo0A* populations had higher values of f_{max} than their corresponding WT transfer regime and the significance of the distance between the distributions was tested and confirmed via Kolmogorov–Smirnov tests (Fig. 2a). Interestingly, WT and Δ*spo0A* populations had the smallest distance in the 100-day transfer regimes, which is likely due to the comparatively brief evolutionary timescale ($\sim$30 gens.) reducing the maximum attainable frequency of a *de novo* mutation. While the condition that a mutation reached a sufficiently high frequency guarantees fixation under a single-locus model of evolution [45], this is not necessarily the case when multiple beneficial mutations simultaneously segregate, and recombination is not sufficiently high so that the population is in quasi-linkage equilibrium. Nevertheless, given the absence of fixation events in many replicate populations we argue that the distribution of f_{max} values is a sufficient proxy for the fixation rate. Overall, the values of f_{max} exhibited in the presences and absence of a seed bank are consistent with the prediction that the presence of a seed bank will reduce the rate of molecular evolution.

The presence of a seed bank generated a similar effect on the magnitude of allele frequency changes between observations ($\frac{|\Delta f|}{\Delta \tau}$) and the direction of frequency changes (Qf; Fig. 2b,c). There was a small, but significant difference where Δ*spo0A* had a higher magnitude of change. This distance increased considerably for the 10-day transfer regime, consistent with the prediction that the presence of a seed bank would buffer temporal changes in allele frequencies. Though, again, much of this distance disappeared for the 100-day transfer regime, a likely result of said treatment's comparatively brief evolutionary timescale.

While the distributions of these three intuitive measures matches our predictions regarding the effect of seed banks on molecular evolutionary dynamics, they are population genetic quantities that are not typically examined. This oversight is arguably due to the historic difficulty of obtaining temporally resolved frequency trajectories for a large number of mutations, rather than the measures themselves being uninformative. To evaluate how these measures typically behave in the presence of a seed bank, we simulated a master equation describing an adapting population that

incorporates seed bank dynamics using previously proposed mathematical formulations [17, 19] (Eq. 10). Briefly, in a population with N active individuals $c \cdot N$ individuals enter a seed bank of size M, while a corresponding number of individuals are resuscitated from their dormant state. The effect of this dynamic can be captured by a single parameter, the average number of generations that an individual spends in a dormant state ($\langle T_d \rangle = M/c \cdot M$). By performing simulations that manipulate the quantity $\langle T_d \rangle$ and calculating the measures in Eq. 4, we were able to recapitulate our empirical survival distributions (Fig. 3). These principled simulations combined with empirical observations confirm long-standing untested hypotheses regarding the effect of seed banks dynamics of molecular evolution.

Parallelism and (con/di)vergence at the gene level

To determine whether endospore formation as a life-history strategy affects the targets of molecular evolution in addition to its direction, it is first necessary to identify the potential contributors towards adaptation among replicate populations within a given experimental condition. To identify potential contributors towards adaptation, we examined the distribution of nonsynonymous mutation counts across genes within each strain-transfer combination. The log-likelihood that some number of genes were enriched for nonsynonymous mutations ($\Delta\ell$) was significant across 1, 10, and 100-day transfer regimes for both strains (Table S3). However, values of $\Delta\ell$ tended to be slightly higher for the WT strain, suggesting that a higher level of parallel evolution occurred when the life-history strategy was present. To determine whether this difference between strains was real or an artifact of the higher number of mutations acquired by the WT strain across transfer regimes, we randomly sub-sampled mutations to obtain a distribution of $\Delta\ell$ values for each strain-transfer combination. While the difference in $\Delta\ell$ values between strains was greatly reduced, the WT strain consistently had a slightly higher level of genome-wide parallelism across transfer regimes (Fig. S5). This increased level of parallelism suggests that the presence of a seed bank can make evolution more predictable. However, it is worth considering that this result may be an aftereffect of this particular experiment, given that *spo0A* is a master regulatory gene that controls cellular processes in addition to endospore formation [46, 47], meaning that the decrease in parallelism in Δ*spo0A* could be a consequence of pleiotropy.

To deconstruct this genome-wide pattern of parallelism, we examined the mutation counts at each gene corrected for gene size (i.e., their multiplicity [35]). As expected, based on the likelihood tests, the multiplicity of the WT was consistently higher than Δ*spo0A* across transfer regimes (Fig. S8a-c). Identifying the set of significantly enriched genes revealed that genes enriched within the WT for a given transfer regime tended to also be enriched within Δ*spo0A* (Fig. S8d-f). This pattern of consistent enrichment occurred across transfer regimes as well as among transfer regime comparisons within a given strain (Fig. S5,6), suggesting that, generally, the direction of evolution at the gene-level tended towards convergence rather than divergence. We found that this is the case, as the degree of overlap in enriched genes relative to a null distribution [31] suggests convergent evolution (Fig. S9). While certain transfer regime and strain comparisons had stronger signals of convergence than others, overall convergent evolution overwhelmingly occurred.

Similar to a previous analysis [31], it is likely that gene identity was again too coarse a measure to determine whether convergent or divergent evolution occurred. While Δ*spo0A* is a master regulatory gene, its removal may only have slightly perturbed the rates of evolution for a large number of genes in a given environment. If true, then it is arguably more appropriate to examine the difference in mutation counts among enriched genes in order to assess whether convergent or divergent evolution occurred. By examining the mean absolute difference in mutation counts across enriched genes

between two transfer regimes ($\langle \Delta\mathcal{M} \rangle$) and standardizing the observed value with respect to an appropriate null distribution ($Z_{\langle \Delta\mathcal{M} \rangle}$) we can establish whether convergent or divergent evolution occurred. The WT strain exhibited significant divergent evolution for the 1-day vs. 10-day and 1-day vs. 100-day comparisons, a result that is consistent with the WT surviving resource-poor environments by forming endospores as a life-history strategy (Fig. 5a). This conclusion is strengthened by the evidence of convergent evolution for the 10-day vs. 100-day comparison, though it was ultimately not significant. For Δ*spo0A* the pattern is inverted, as there is a significant signal of convergent evolution for the 1-day vs. 10-day comparison. For the 1-day vs. 100-day and 10-day vs. 100-day comparisons we find a strong signal of divergent evolution, suggesting that the direction of molecular evolution shifts between 10 and 100 days for Δ*spo0A* with no such shift occurring between 1 and 10 days.

To examine how seed bank formation affected the direction of molecular evolution within a given transfer regime, we repeated the convergent/divergent analysis between the WT and Δ*spo0A* populations, analyses that can be bolstered through comparisons to fitness estimates (Fig. 5b). Divergent evolution overwhelmingly occurred for all three comparisons, though it was at its highest for the 10-day transfer regime (Fig. 5c). This increased divergence for the 10-day transfer regime, along with evidence of convergent evolution for the 1-day vs. 10-day comparison, may reflect adaptation of Δ*spo0A* to this specific regime. Our estimates of $\langle \Delta\mathcal{M} \rangle$ can be compared to fitness estimates, allowing us to examine how the sign and magnitude of fitness changes with the direction of molecular evolution. Starting from the 1-day transfer regime, we found that the fitness of Δ*spo0A* was effectively zero, a result that is consistent with endospore formation having a negligible fitness effect in environments where it would not be advantageous. The increase in fitness for the 10-day transfer regime corresponded with an increase in the degree of divergent evolution in the 10-day transfer regime (Fig. 5d). Though by day 100 the fitness benefit held by Δ*spo0A* had dissipated and the magnitude of divergent evolution had diminished. This result suggests that the ability to form endospores actually conferred a fitness disadvantage at the 10-day mark. However, the fitness benefit was temporary and the decrease in the degree of divergent evolution suggests that it is unlikely that Δ*spo0A* was able to adapt to the harsh 100-day environment. An analysis of the set of genes that were enriched for mutations within a specific strain-transfer combination supports this conclusion [41]. The 100-day Δ*spo0A* was the only strain-transfer combination with no unique enriched genes (Table ; File S2), suggesting that Δ*spo0A* may have been unable to adapt to this extremely resource poor environment.

While it is unlikely that Δ*spo0A* was able to adapt to 100 days of energy limitation, conversely the 10-day Δ*spo0A* harbored the highest number of unique enriched genes for all three Δ*spo0A* transfer regimes, suggesting that adaptation may have occurred even in the absence of endospore formation as a life-history strategy. The mechanism responsible for the temporary gain in fitness of Δ*spo0A* is unknown, though it is likely partially due to the recycling of dead cells, a phenotype that allows individuals to exploit an untapped resource [25, 48, 49]. Naturally, dormant cells cannot use this resource as their metabolism is effectively nonexistent and, in the case of endospores, completely inert, leaving Δ*spo0A* with unrestricted access. Regardless, for the purposes of this study, the removal of endospore formation as a life-history strategy provides a clear temporary fitness benefit in environments where resources become increasingly limited.

Finally, given that endospore formation did not occur in the 1-day transfer regime, it is possible that the pathways encoding said life-history trait were susceptible to decay due to relaxed selective pressure. By calculating the fraction of nonsynonymous mutations in genes that encode for endospore formation and calculating the difference between Δ*spo0A* and the WT, we found that endospore associated genes were slightly

enriched in the WT for all transfer regimes, a difference that was significant using null distributions simulated using binomial sampling (Table S4). Operating under the premise that the majority of endospore-forming genes are nonfunctional in Δ*spo0A* populations, for WT populations in the daily transfer regime this result can be viewed as the outcome of positive selection for the removal of endospore formation as an energetically costly trait in resource-replete environments [50, 51]. Prior studies [28–30] as well as spore accumulation assays support this conclusion, as by day 500 the ability to form endospores was rapidly reduced for populations in the 1-day transfer regime, to the point that it took 10 days for 10% of the population to form endospores (Fig. S10). However, the question remains as to why 10 and 100-day WT populations acquired a greater fraction of nonsynonymous mutations. Endospore formation undergoes no noticeable decline in these transfer regimes (Fig. S10), suggesting that these mutations had a negligible or even positive effect.

This study tested lost-standing predictions and generated new questions for future research. Given that *Bacillus* incurs an environment-dependent fitness effect when endospore formation is removed, it is worth investigating the quantitative effects of mutation on endospore formation and, ultimately their effect on fitness. Given that limited evolutionary timescales can be obtained in experiments that mimic the environments where the ability to enter a dormant state has the greatest fitness benefit, alternative approaches may be necessary to probe the effect of mutation on endospore formation. One promising option is mass barcoding (e.g., RB-TnSeq), sidestepping the slow input of mutations via slow generation times by generating large mutant libraries, the frequencies of which can then be tracked [52]. Such approaches have already been leveraged to identify environment-dependent fitness effects as well as the mode in which traits mediate fitness effects [53]. Given that endospore formation is a complex trait with many loci that are likely interacting, it may be a suitable candidate to apply recently developed models that predict the form of the distribution of fitness effects when epistatic interactions are prevalent [54].

Conclusion

We demonstrated that the ability to form seed banks altered the molecular evolutionary dynamics of microbial populations. Populations capable of forming seed banks consistently accumulated higher levels of genetic diversity and had a reduced rate of molecular evolution. Through forward-time simulations, we were able to recapitulate empirical observations on the effect of seed banks on the rate and direction of allele frequency changes as well as the maximum attainable frequency. In addition to testing previously proposed predictions on the effect of seed banks on genetic diversity, new patterns were found. Specifically, we determined that endospore formation has the capacity to alter the direction of molecular evolution within a population. Stated inversely, the absence of endospore formation contributed to a substantial signal of divergent evolution for populations that were intermediately transferred. This signal of divergence, alongside the observation that the absence of endospore formation provided a substantial fitness benefit, suggests that adaptation to energy-limited environments may be possible in the absence of a highly conserved life-history strategy. Though any such adaptation would likely be transitory, as the absence of endospore formation resulted in an increasingly strong fitness disadvantage as the degree of energy-limitation increased.

Data and code Availability

Raw sequence data of Δ*spo0A* lines are available on the NCBI Sequencing Read Archive under BioProject ID PRJNA639642. Raw sequence data for WT lines was previously published [31] and is available under BioProject ID PRJNA639414. Reproducible code to perform all simulations and analyses is available on GitHub under the repository: Bacillus_Evol_Timeseries. Processed data and annotations are available on Zenodo under the DOI: 10.5281/zenodo.5549311.

Acknowledgements

We thank M. Behringer, T. Doak, D. A. Drummond, P. Foster, M. Lynch, J. McKinlay, and members of the Lennon lab for their helpful feedback at various stages of the project. We thank A. G. Casanova, D. A. Schwartz, and P. G. Wall for feedback on initial drafts. We thank H. Long for his help in setting up library construction, K. Miller for assisting with sample collection, and members of the Kearns lab for sharing their plasmid protocols and assisting with troubleshooting. This work was supported by US Army Research Office Grant W911NF-14-1-0411 and the Society for the Study of Evolution Graduate Research Excellent Grant (GREG) Rosemary Grant Advanced Award. Computing resources for simulations was supported by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, the National Science Foundation under Grant No. CNS-0521433, and Shared University Research grants from IBM, Inc., to Indiana University.

Author contributions

W.R.S., E.P., and J.T.L. designed the project; W.R.S., E.P., and K.B.G. conducted the experiments and generated the data; W.R.S. built and implemented the model; W.R.S. performed all statistical analyses; W.R.S., E.P., K.B.G., and J.T.L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Lennon, J. T., den Hollander, F., Wilke-Berenguer, M. & Blath, J. Principles of seed banks and the emergence of complexity from dormancy. *Nature Communications* **12**, 4807 (2021). URL https://www.nature.com/articles/s41467-021-24733-1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Applied mathematics;Biodiversity;Evolutionary theory;Theoretical ecology Subject_term_id: applied-mathematics;biodiversity;evolutionary-theory;theoretical-ecology.

2. Buoro, M. & Carlson, S. M. Life-history syndromes: Integrating dispersal through space and time. *Ecology Letters* **17**, 756–767 (2014). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12275.

3. Rubio de Casas, R., Donohue, K., Venable, D. L. & Cheptou, P.-O. Gene-flow through space and time: dispersal, dormancy and adaptation to changing environments. *Evolutionary Ecology* **29**, 813–831 (2015). URL https://doi.org/10.1007/s10682-015-9791-6.

4. Vitalis, R., Rousset, F., Kobayashi, Y., Olivieri, I. & Gandon, S. The joint evolution of dispersal and dormancy in a metapopulation with local extinctions and kin competition. *Evolution* **67**, 1676–1691 (2013). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.12069.

5. Venable, D. L. & Brown, J. S. The selective interactions of dispersal, dormancy, and seed size as adaptations for reducing risk in variable environments. *The American Naturalist* **131**, 360–384 (1988). URL https://www.journals.uchicago.edu/doi/abs/10.1086/284795.

6. Venable, D. L. & Lawlor, L. Delayed germination and dispersal in desert annuals: Escape in space and time. *Oecologia* **46**, 272–282 (1980). URL https://doi.org/10.1007/BF00540137.

7. Tellier, A., Laurent, S. J. Y., Lainer, H., Pavlidis, P. & Stephan, W. Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proceedings of the National Academy of Sciences* **108**, 17052–17057 (2011). URL https://www.pnas.org/content/108/41/17052. Publisher: National Academy of Sciences Section: Biological Sciences.

8. Hairston, N. G. *et al.* Rapid evolution revealed by dormant eggs. *Nature* **401**, 446–446 (1999). URL https://www.nature.com/articles/46731. Number: 6752 Publisher: Nature Publishing Group.

9. Huang, X. *et al.* The earliest stages of adaptation in an experimental plant population: strong selection on QTLS for seed dormancy. *Molecular Ecology* **19**, 1335–1351 (2010). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2010.04557.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-294X.2010.04557.x.

10. Shoemaker, W. R. & Lennon, J. T. Evolution with a seed bank: The population genetic consequences of microbial dormancy. *Evolutionary Applications* **11**, 60–75 (2018). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5748526/.

11. Guppy, M. & Withers, P. Metabolic depression in animals: physiological perspectives and biochemical generalizations. *Biological Reviews of the Cambridge Philosophical Society* **74**, 1–40 (1999).

12. Willis, C. G. *et al.* The evolution of seed dormancy: environmental cues, evolutionary hubs, and diversification of the seed plants. *New Phytologist* **203**, 300–309 (2014). URL `https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.12782`. _eprint: https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.12782.

13. Doebeli, M., Ispolatov, Y. & Simon, B. Towards a mechanistic foundation of evolutionary theory. *eLife* **6**, e23804 (2017). URL `https://doi.org/10.7554/eLife.23804`.

14. Kunkel, T. A. DNA replication fidelity. *The Journal of Biological Chemistry* **279**, 16895–16898 (2004).

15. Schoen, D. J., David, J. L. & Bataillon, T. M. Deleterious mutation accumulation and the regeneration of genetic resources. *Proceedings of the National Academy of Sciences* **95**, 394–399 (1998). URL `https://www.pnas.org/content/95/1/394`. Publisher: National Academy of Sciences Section: Biological Sciences.

16. Koopmann, B., Müller, J., Tellier, A. & Živković, D. Fisher-Wright model with deterministic seed bank and selection. *Theoretical Population Biology* **114**, 29–39 (2017).

17. Blath, J., Casanova, A. G., Eldon, B., Kurt, N. & Wilke-Berenguer, M. Genetic variability under the seedbank coalescent. *Genetics* **200**, 921–934 (2015). URL `https://www.genetics.org/content/200/3/921`.

18. Kaj, I., Krone, S. M. & Lascoux, M. Coalescent theory for seed bank models. *Journal of Applied Probability* **38**, 285–300 (2001). URL `https://www.cambridge.org/core/journals/journal-of-applied-probability/article/coalescent-theory-for-seed-bank-models/3137A15F299546B0AA701FA974C19E46`.

19. Blath, J., Casanova, A. G., Kurt, N. & Wilke-Berenguer, M. A new coalescent for seed-bank models. *Annals of Applied Probability* **26**, 857–891 (2016). URL `https://projecteuclid.org/euclid.aoap/1458651822`.

20. van Gestel, J., Ackermann, M. & Wagner, A. Microbial life cycles link global modularity in regulation to mosaic evolution. *Nature Ecology & Evolution* **3**, 1184–1196 (2019). URL `http://www.nature.com/articles/s41559-019-0939-6`.

21. Nunney, L. The effective size of annual plant populations: The interaction of a seed bank with fluctuating population size in maintaining genetic variation. *The American Naturalist* **160**, 195–204 (2002). URL `https://www.journals.uchicago.edu/doi/full/10.1086/341017`. Publisher: The University of Chicago Press.

22. Ellner, S. & Hairston, N. G. Role of overlapping generations in maintaining genetic variation in a fluctuating environment. *The American Naturalist* **143**, 403–417 (1994). URL `https://www.journals.uchicago.edu/doi/abs/10.1086/285610`. Publisher: The University of Chicago Press.

23. Lennon, J. T. & Jones, S. E. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews. Microbiology* **9**, 119–130 (2011).

24. Weller, C. & Wu, M. A generation-time effect on the rate of molecular evolution in bacteria. *Evolution* **69**, 643–652 (2015).

25. Shoemaker, W. R. *et al.* Microbial population dynamics and evolutionary outcomes under extreme energy limitation. *Proceedings of the National Academy of Sciences* **118** (2021). URL `https://www.pnas.org/content/118/33/e2101691118`. Publisher: National Academy of Sciences Section: Biological Sciences.

26. Sachidanandham, R. & Yew-Hoong Gin, K. A dormancy state in nonspore-forming bacteria. *Applied Microbiology and Biotechnology* **81**, 927–941 (2009).

27. Zeigler, D. R. & Nicholson, W. L. Experimental evolution of Bacillus subtilis. *Environmental Microbiology* **19**, 3415–3422 (2017).

28. Maughan, H., Masel, J., Birky, C. W. & Nicholson, W. L. The Roles of Mutation Accumulation and Selection in Loss of Sporulation in Experimental Populations of Bacillus subtilis. *Genetics* **177**, 937–948 (2007). URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2034656/`.

29. Maughan, H. *et al.* The Population Genetics of Phenotypic Deterioration in Experimental Populations of Bacillus Subtilis. *Evolution* **60**, 686–695 (2006). URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2006.tb01148.x`.

30. Maughan, H., Birky, C. W. & Nicholson, W. L. Transcriptome Divergence and the Loss of Plasticity in Bacillus subtilis after 6,000 Generations of Evolution under Relaxed Selection for Sporulation. *Journal of Bacteriology* **191**, 428–433 (2009). URL `https://jb.asm.org/content/191/1/428`.

31. Shoemaker, W. R., Polezhaeva, E., Givens, K. B. & Lennon, J. T. Molecular Evolutionary Dynamics of Energy Limited Microorganisms. *Molecular Biology and Evolution* (2021). URL `https://doi.org/10.1093/molbev/msab195`.

32. Mason, C. A. & Hamer, G. Cryptic growth in Klebsiella pneumoniae. *Applied Microbiology and Biotechnology* **25**, 577–584 (1987). URL `https://doi.org/10.1007/BF00252019`.

33. Banks, M. K. & Bryers, J. D. Cryptic growth within a binary microbial culture. *Applied Microbiology and Biotechnology* **33**, 596–601 (1990). URL `https://doi.org/10.1007/BF00172558`.

34. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011). URL `http://journal.embnet.org/index.php/embnetjournal/article/view/200`.

35. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017). URL `https://www.nature.com/articles/nature24287`.

36. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology (Clifton, N.J.)* **1151**, 165–188 (2014).

37. Transtrum, M. K. & Qiu, P. Bridging Mechanistic and Phenomenological Models of Complex Biological Systems. *PLoS Computational Biology* **12** (2016). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4871498/.

38. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference* (2010).

39. Patefield, W. M. Algorithm as 159: An efficient method of generating random r × c tables with given row and column totals. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **30**, 91–97 (1981). URL https://www.jstor.org/stable/2346669.

40. Baak, M., Koopman, R., Snoek, H. & Klous, S. A new correlation coefficient between categorical, ordinal and interval variables with pearson characteristics (2019). 1811.11440.

41. Shoemaker, W. R. & Lennon, J. T. Quantifying parallel evolution. *bioRxiv* 2020.05.13.070953 (2020). URL https://www.biorxiv.org/content/10.1101/2020.05.13.070953v1.

42. Good, B. H. & Desai, M. M. The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics* **199**, 177–190 (2015). URL http://www.genetics.org/lookup/doi/10.1534/genetics.114.172460.

43. Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences* **109**, 4950–4955 (2012). URL https://www.pnas.org/content/109/13/4950. Publisher: National Academy of Sciences Section: Biological Sciences.

44. Sung, W. *et al.* Asymmetric Context-Dependent Mutation Patterns Revealed through Mutation–Accumulation Experiments. *Molecular Biology and Evolution* **32**, 1672–1683 (2015). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4476155/.

45. Ewens, W. J. *Mathematical Population Genetics 1: Theoretical Introduction*. Interdisciplinary Applied Mathematics, Mathematical Population Genetics (Springer-Verlag, New York, 2004), 2 edn. URL http://www.springer.com/gp/book/9780387201917.

46. Dubnau, D. Genetic competence in Bacillus subtilis. *Microbiological Reviews* **55**, 395–424 (1991). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC372826/.

47. Hoch, J. A. Genetic analysis of pleiotropic negative sporulation mutants in Bacillus subtilis. *Journal of Bacteriology* **105**, 896–901 (1971).

48. Rozen, D. E., Philippe, N., Arjan de Visser, J., Lenski, R. E. & Schneider, D. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters* **12**, 34–44 (2009). URL http://doi.wiley.com/10.1111/j.1461-0248.2008.01257.x.

49. Bradley, J. A., Amend, J. P. & LaRowe, D. E. Necromass as a Limited Source of Energy for Microorganisms in Marine Sediments. *Journal of Geophysical Research: Biogeosciences* **123**, 577–590 (2018). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JG004186.

50. Bradshaw, W. E., Armbruster, P. A. & Holzapfel, C. M. Fitness consequences of hibernal diapause in the pitcher-plant mosquito, wyeomyia smithii. *Ecology* **79**, 1458–1462 (1998). URL https://www.jstor.org/stable/176758.

51. Cáceres, C. E. & Tessier, A. J. To sink or swim: Variable diapause strategies among daphnia species. *Limnology and Oceanography* **49**, 1333–1340 (2004).

52. Wetmore, K. M. *et al.* Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio* **6** (2015). URL https://mbio.asm.org/content/6/3/e00306-15. Publisher: American Society for Microbiology Section: Research Article.

53. Kinsler, G., Geiler-Samerotte, K. & Petrov, D. A. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife* **9**, e61271 (2020). URL https://doi.org/10.7554/eLife.61271. Publisher: eLife Sciences Publications, Ltd.

54. Reddy, G. & Desai, M. M. Global epistasis emerges from a generic model of a complex trait. *bioRxiv* (2020). URL http://biorxiv.org/lookup/doi/10.1101/2020.06.14.150946.

Tables

	Number unique enriched genes	
Transfer regime	*B. subtilis* WT	*B. subtilis* Δ*spo0A*
1-day	77	2
10-day	12	8
100-day	12	0

Figures

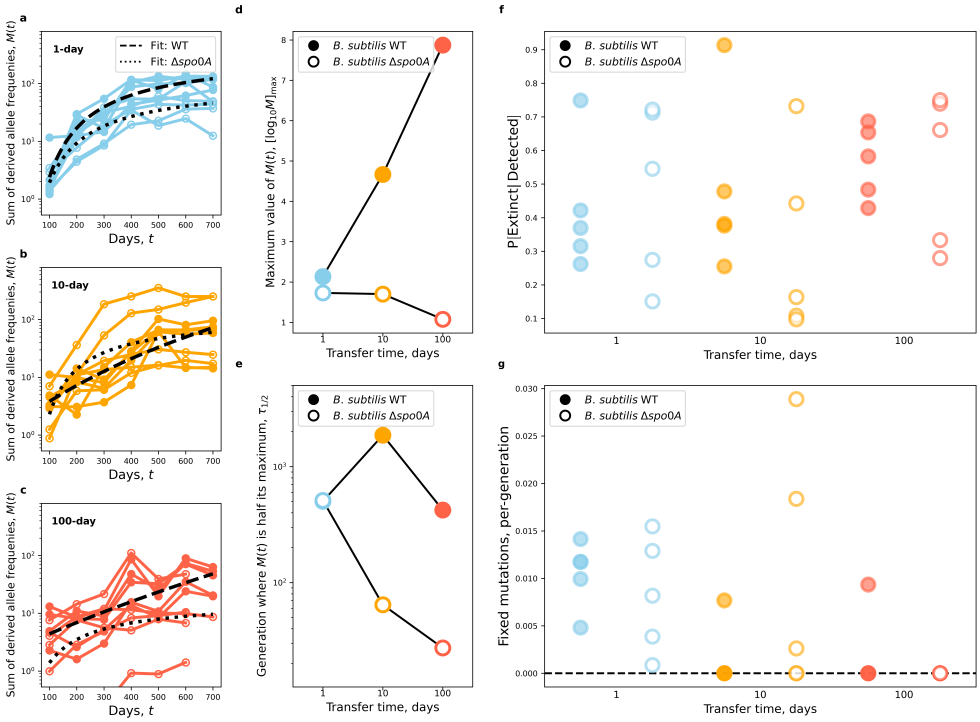

Figure 1. The presence of a seed bank altered the accumulation of genetic diversity. **a-c)** By examining the sum of derived allele frequencies ($M(t)$, Eq. 2), we were able to summarize the accumulation of *de novo* mutations over time for all strains and transfer regimes. The WT strain typically had higher estimates of $M(t)$ than Δ*spo0A* and the relationship between t and $M(t)$ became noticeably more linear as transfer time increased for the WT strain. **d,e**) To quantify the effect of seed bank formation on this empirical relationship we formulated a phenomenological model that allowed us to summarize the curve through two parameters: the maximum amount of genetic diversity that could accumulate ($[\log_{10} M]_{\max}$) and the number of generations until half of the maximum is reached ($\tau_{1/2}$; Eq. 3). Values of $[\log_{10} M]_{\max}$ for the WT strain steadily increased with transfer time while Δ*spo0A* remained consistent, consistent with the prediction that the presence of a seed banks increases the amount of genetic diversity that a population can maintain. Conversely, $\tau_{1/2}$ decreased for Δ*spo0A* but remained constant for the WT, consistent with the prediction that the rate of molecular evolution would increase in the absence of a seed bank. **f,g**) However, the effect of seed banks on the final states of alleles was less straightforward. While fixation events occurred across transfer regimes and strains, there was substantial variation across replicate populations that made it difficult to determine whether the presence of a seed bank affected the probability of fixation or the rate of molecular evolution (per-generation number of substitution rate).

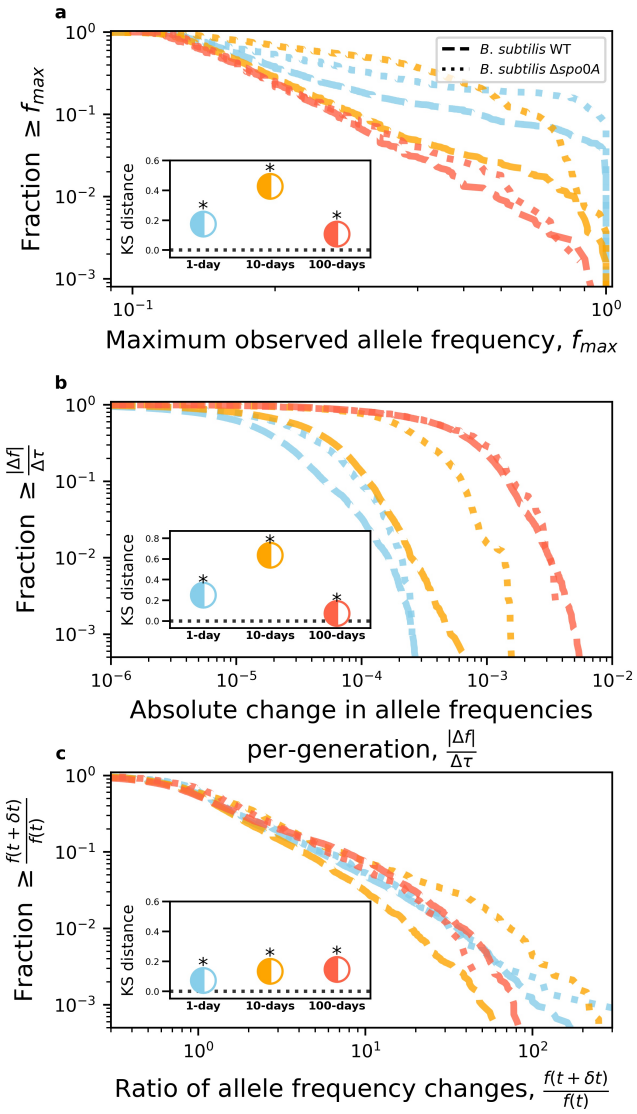

Figure 2. Due to the low number of fixation events, it was necessary to devise alternative measures of molecular evolution to evaluate the effect of seed banks. We examined three measures corresponding to **a)** the maximum frequency realized by an allele, **b)** the per-generation magnitude of change in allele frequency, and **c)** the change in the direction of allele frequencies between time points (Eq. 10). These measures were examined by calculating the empirical survival distribution (the complement of the empirical cumulative distribution function) for a given strain-transfer combination. The typical value of all three measures was higher for Δ*spo0a* than the WT across transfer regimes, results that are consistent with the predicted effect of a seed bank. The difference we observed between strains was confirmed via Kolmogorov–Smirnov tests for all transfer regimes ($P < 0.05$ marked by asterisk).

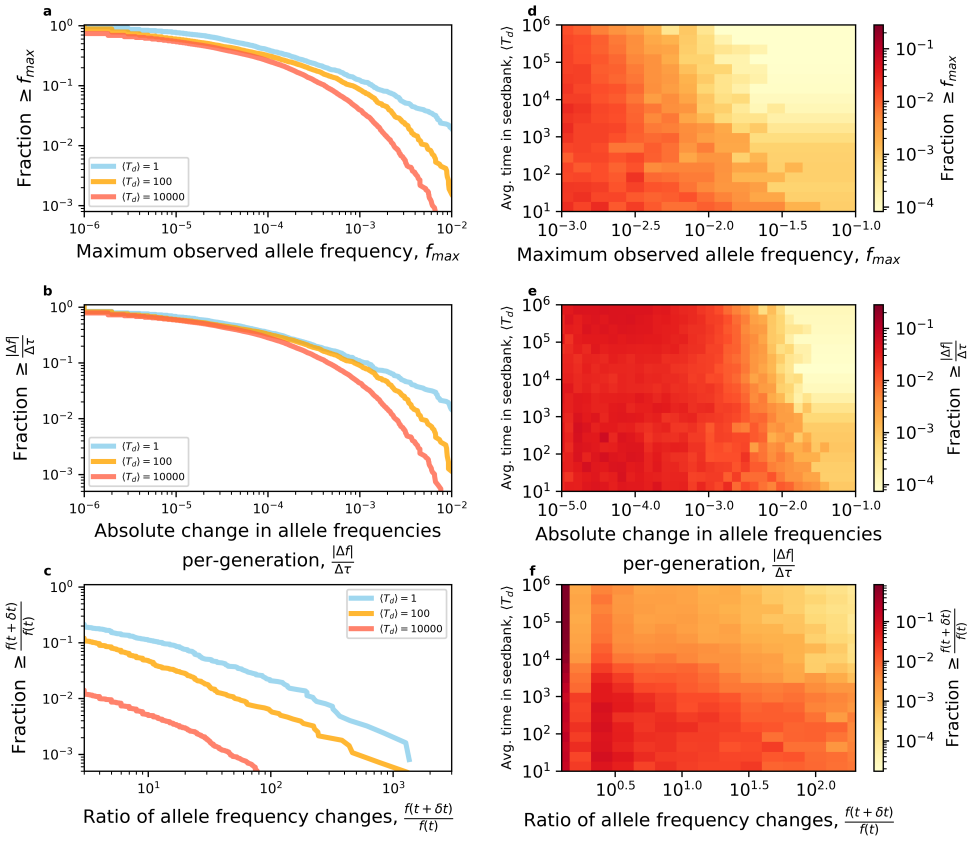

Figure 3. To validate our empirical observations we performed forward-time population genetic simulations of populations with seed banks of varying size, where the effect of seed banks can be summarized by the average number of generations that an individual spends in a dormant state ($\langle T_d \rangle$; Materials and Methods). **a)** The breadth of simulated survival distributions gradually decreased with $\langle T_d \rangle$ for all three measures of allele frequencies (Eq. 10), recapitulating the empirical results described in Fig. 2. **a)** The general form of all three CDFs changes as $\langle T_d \rangle$ increases in the manner we would expect in addition to what we observe in Fig. 2. **b)** The robustness of this pattern is made evident by examining a wide range of $\langle T_d \rangle$ values with greater resolution.

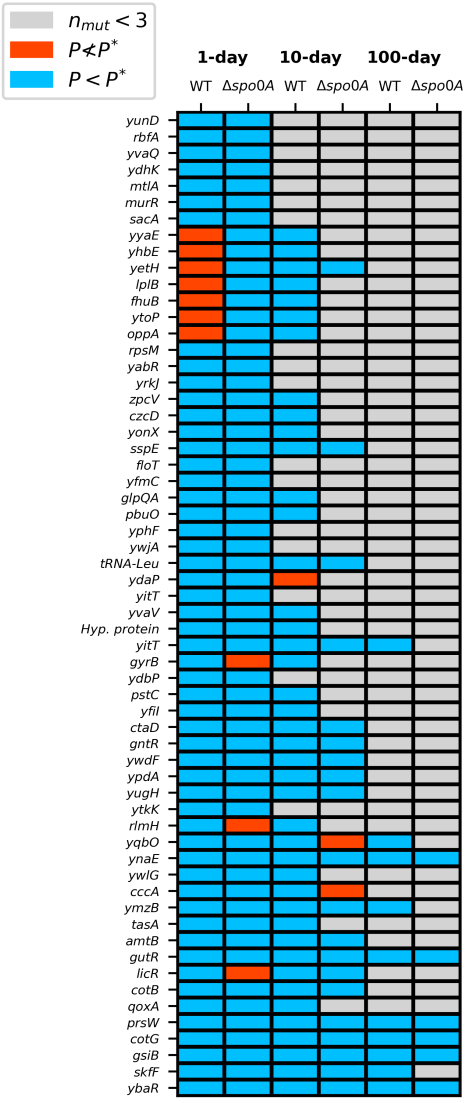

Figure 4. By comparing the set of genes that contributed towards parallel evolution with a given strain-transfer combination, we visualize patterns of convergent and divergent molecular evolution. While all strain-transfer combinations consistently acquired more nonsynonymous mutations than expected by chance at a large number of genes ($P < P^*$), very few genes were enriched exclusively within a given strain. Those genes that were significantly enriched within a given strain combination typically also acquired mutations at a non-significant level ($P \not< P^*$) in the remaining strain, suggesting that the removal of endospore formation did not generate evolutionary trajectories that were divergent in terms of gene identity. To increase statistical power we ignored all genes that acquires less than three mutations ($n_{mut} < 3$). Gene names are listed as provided in the annotated reference genome, all other names were acquired using RefSeq IDs (see File S1 for gene metadata).

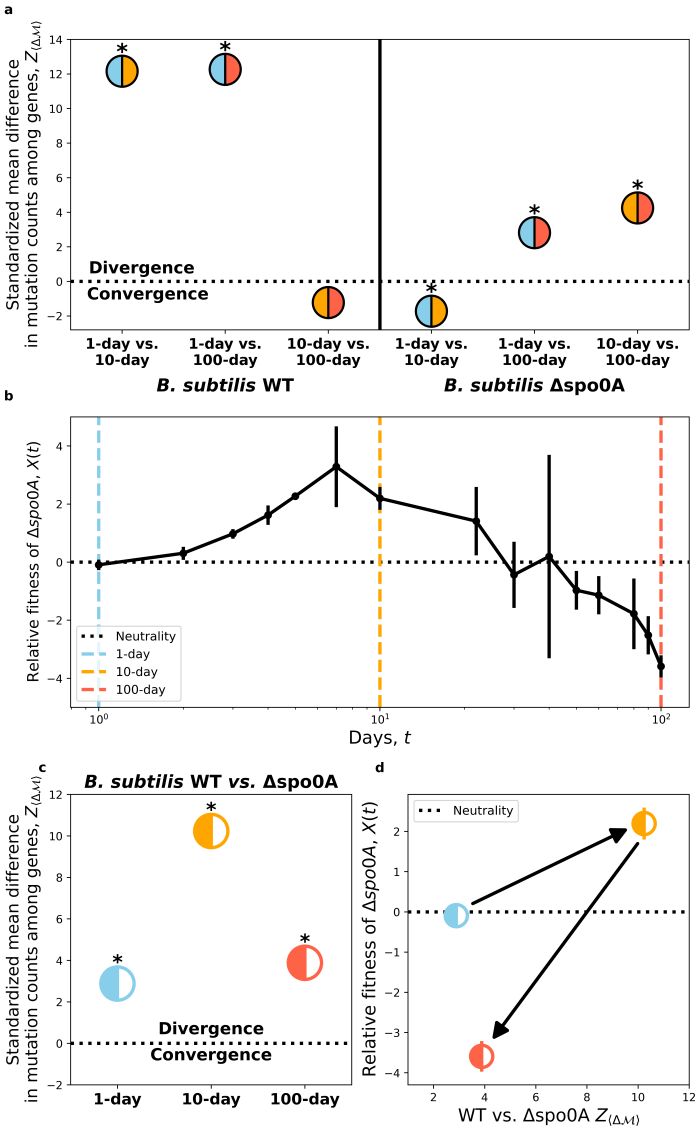

Figure 5. By examining the number of mutations within each gene, we can determine whether convergent or divergent evolution occurred between a given pair of transfer regimes or strains by calculating the mean absolute difference of mutation counts across genes ($\langle\Delta\mathcal{M}\rangle$; Eq. 9). **a)** A comparison between all transfer regime combinations within each strain reveals contrasting dynamics of divergence/convergence. Divergent evolution initially occurs among the WT background initially for 1 vs. 10 and 1 vs. 100-day comparisons, with 10 and 100-day transfers having a weak signal of convergence. This pattern is inverted for Δ*spo0A*, as 1 and 10-day transfers converged and remaining transfer regime combinations diverged. **b)** For comparisons between strains, we can compare signals of convergent/divergent evolution between WT and Δ*spo0A* strains with the transfer regime-dependent fitness effects of removing *spo0A*. **c)** Generally, we found that divergent evolution consistently occurred across transfer regimes, with the 10-day transfers harboring the strongest signal of divergent evolution. **d)** Mapping signals of divergent evolution to estimates fitness, we can see how the sign and magnitude of selection changes with the degree of divergent evolution. Asterisks denote $P < 0.05$.