This supporting information augments the discussion in the article on three key points. In the first section, we support the comparison of theoretical and experimental results by testing the robustness of the simulation method and the methods used to extract the data points shown in Fig. 2. In the second section, we explain the scaling of the number of distinct ancestors of the population as a function of time looking into the past; in the main article, this distribution is compared with experiment and is used to derive the scaling of diversity with habitat area. In the third section, we discuss in more detail how selection affects the results of the article.

Methods Used to Obtain the data and simulation results in Fig. 2

Data extraction. The data on the number of lineages $L(T)$ in *Pseudomonas* bacteria, shown in our Fig. 2, is derived from genetic data obtained by Cho and Tiedje (1). From their dendogram of 248 isolates, we obtained $L(r)$, where r represents BOX-PCR similarity coefficients. The r-values were then mapped onto T-values (2). In order to obtain $L(T)$, the r-values were first mapped onto DNA-DNA homology values h using a regression curve $h = \sqrt{1.046 - (1.15 - \frac{r}{1.07})^2}$ obtained from genetic sequences of *Xanthomonas* bacteria (3). Although the data used to obtain this curve are noisy, this noise does not significantly affect the comparison of the *Pseudomonas* data with theoretical results. Because measurements of multiple isolates are used to determine each value of $L(T)$, the impact of noise is reduced. Moreover, both noise in the values of r and h and variation in the transformation between them has little impact on the results described in the article, since the comparison is with the robust scaling behavior. To verify this, noise was added to the r-values of the original data with the same variance as the data in used to obtain the r-to-h transformation. Fig. 4 shows that the resulting $L(T)$ values are not substantially affected. We note that the only significant

sensitivity to noise is found in the region just above $T/T_0 = 1,000$ where the original points are least in agreement with theory and the size of the error is approximately the difference between data values and the theoretical curve, suggesting that even this disagreement may be due to noise. The effect on $L(T)$ of systematic changes in the shape of the fitted r-to-h curve were also tested and found insignificant.

The h-values were then mapped onto values proportional to T using the Tajima form (4) of the Jukes-Cantor correction. In the figures, this is normalized by dividing by T_0, the time to the smallest genetic difference considered[1] ($r = 0.95$).

Simulation of Sampled Population. The lineages of the sampled population were simulated directly as random walkers, with simulated time corresponding to time in the past (Fig. 5). Any walkers that moved onto the same site coalesced with probability $p_c = 0.15$. Initial random walkers were placed, one per isolate, on a 50×25 lattice on sites whose coordinates were proportional to the latitude and longitude of samples of the Cho and Tiedje data (Table 1).

In order to produce a curve that is smooth at long times, the average of 100 runs was taken. Error bars for the simulation of $L(T)$ are shown in Fig. 6. The theoretical simulations included a fitting of two parameters. These parameters do not set the key features of the comparison—the slopes of the long and short time behaviors—but rather only set the intercepts (offsets), since the shape of the curve is highly constrained by the theory. The fitting of the two parameters, shown in Fig. 7, is determined by the two intercepts. The slope of the long-time portion of the curve (the scaling regime) is determined by the scaling of the number of lineages, which has no adjustable parameters.

[1] Cho & Tiedje found r-values of individual samples to be reliable only for differences of 0.1; however, the average implied in the use of many such measurements to determine $L(r)$ allows a finer grid down to 0.05.

Explanation of Scaling of Number of Lineages as a Function of Time

The scaling of the number of distinct ancestors (lineages) at a time T before the present in well mixed populations (5, 6) can be understood as follows. In a genealogical tree, two lineages coalesce if two individuals have the same parent in the previous generation. In the Wright-Fisher model of a well mixed population, each lineage can be thought of as jumping to a random site on each time step (time being measured as generations in the past), with multiple lineages on the same site coalescing. The number of lineages $L(T+1)$ in generation $T+1$ is therefore the number of distinct sites obtained as a result of the independent selection of one of the N possible parents for each of the $L(T)$ offspring, i.e., when selecting a site $L(T)$ times from the set of N sites with replacement. Let $p(T) = L(T)/N$ be the fraction of individuals in generation T that have descendants in the present. For $p(T)$ sufficiently small (that is, when T is not small) we can neglect all but pairwise coalescence. The probability that two lineages coalesce into the previous generation is $p(T)^2$ when $p(T)$ is not large, with a resulting change in $p(T)$ of $\Delta p \cong -p^2/2$ (the factor of $1/2$ arises because only one lineage is lost for a pairwise coalescence[2]; still, the scaling behavior depends only on the proportionality of change to p^2 and not on the coefficient), which in the continuous limit is $\frac{dp}{dT} = -p^2/2$, giving $p(T) = \frac{2}{T}$, i.e. the scaling behavior (the dependence on T):

$$p(T) \sim \tfrac{1}{T}$$

This result, which we obtained from the Wright-Fisher model, applies more generally. Specifically, it applies for any model in which the probability of coalescence in a single generation is proportional to the square of the proportion of the rmeining lineages, $\Delta p \propto -p^2$. This holds when, for example, the meeting of a pair of lienages

[2]A more precise recursion relationship for the Wright-Fisher model is $p(T+1) = (1 - e^{p(T)})$.

is independent of the meeting of other pairs of lineages. As discussed below, it also applies for spatial models whose spatial dimension is > 2. In two dimensions there is a weak logarithmic correction, and in fewer than two dimensions there is a different behavior. The reason for the difference in low dimensions is that orgaisms that are near in space are highly likely to coalesce, which changes the way the loss of lineages depends on density.

For more than two spatial dimensions, explaining the scaling of $p(T)$ is straightforward, as it is for fewer than two. The two-dimensional case requires a more involved treatment.

For many dimensions, the density of walkers in space is essentially uniform even as they are coalescing. In this case we can treat the problem similarly to a chemical reaction, where two walkers have the probability of intersecting with each other given by the density squared of walkers. This gives a differential equation:

$$\frac{dp}{dT} \propto -p^2,$$

which has the solution

$$p \sim \frac{1}{T}.$$

This result, identical to the well mixed case, is valid for more than two dimensions.

For one dimension, we can assume that a walker coalesces with all of the others in an area of the typical distance traveled in a random walk (sweeping out the region of the walk). A random walk travels a characteristic length $l(T) \sim T^{1/2}$ in a time T. This results in a single lineage for each part of the habitat with an area proportional to $l(T)$. This gives a density of remaining walkers (inverse to the area of a single walker) of $p(T) \sim 1/l(T) \sim 1/T^{1/2}$.

The derivation in two dimensions is more difficult. Ref. 7 provides a formal derivation for all integral dimensions. A discusion of some key aspects of the derivation follows. If we consider the process of coalescence, we see that the likelihood that a walker is near another walker is lower than the probability that it is farther away due to previous coalescence events. We can approximate the probability distribution of walkers at a distance x from a particular walker as the solution of a diffusion equation with a sink (i.e. the Green's function):

$$\nabla^2 G(x) = -\delta(x).$$

The probability of finding a walker a distance x away from a particular walker is given by the Green's function. The derivation of $p(T)$ for more than two dimensions follows from the fact that the Green's function has a non-zero finite value at the origin, indicating that the rate at which coalescence occurs is reduced but still is proportional to the overall density. Since each walker coalesces with others at a rate that is proportional to the density, total coalescence is proportional to the density squared as we argued occurs for high dimensions for the well mixed case.

In two dimensions, however, the Green's function has a logarithmic singularity. This gives rise to a reduction of the rate at which walkers collide with each other. For each walker, the rate at which other walkers collide with it goes asymptotically as $p/\log(p)$ rather than p. This gives rise to the equation for the remaining density of walkers:

$$\frac{dp}{dT} \sim -\frac{p^2}{\log(p)},$$

which gives the scaling result

$$p \sim \frac{\log T}{T}.$$

This corresponds to the long-time scaling behavior seen in Fig. 2.

Scaling for Short Time Behavior of Sampled Populations. To understand the shape of the short-time behavior of $L(T)$ obtained from the experimental results, we studied the effect of taking samples at clustered locations in a spatial population. The short-time behavior of $L(T)$ for the sampled population describes the coalescence of random walks representing lineages that start from locations that are correlated in space. The scaling of the loss of walkers can be approximated by considering the aggregation of random walkers all starting from (approximately) the same location. The probability of coalescence is proportional to the coincidence of the positions of the two walkers. For convenience we calculate this in the continuum limit as:

$$\int \left(\tfrac{1}{4\pi^2 T^2}\right) \delta^2(x-y) e^{-x^2/2T} e^{-y^2/2T} d^2x d^2y \propto \tfrac{1}{T},$$

where x and y are both two-dimensional coordinates representing the locations of the two walkers, each with Gaussian distribution of probabilities as a result of diffusion from their point of origin. $\delta^2(x)$ is a two-dimensional (squared) Dirac delta function imposing coincidence on the walkers in order to coalesce. The result of the integral is proportional to $1/T$ giving the equation:

$$\tfrac{dp}{dT} \sim -\tfrac{1}{T},$$

which gives the scaling result for $p(T)$:

$$p(T) = p_0 - c\log(T/T_0).$$

This result is plotted in Fig. 8. The resulting curve is a reasonable approximation to the *Pseudomonas* data until the long-time behavior occurs. It does not match the result of the full simulation because it does not represent multiple locations of the samples. Thus we can understand both the short time and long time behaviors, and the simulation agrees with the data over the full range of times.

Effect of Selection

In this section we expand on the brief discussion of the effect of advantageous mutations on our results. A periodic selection event occurs when the descendants of a selected mutant take over the population, wiping out the population's existing diversity, before the mutant's descendants have had time to develop an equivalent amount of diversity. [This is also known as genetic hitchhiking (8, 9).] Thus, whether periodic selection affects the long-term average diversity depends on whether the steady-state diversity is recovered well before the next selection event reduces the diversity. The recovery time T_A is the expected time to the most recent common ancestor of a population in a steady state. If the time t_E between selection events is large compared to T_A, the population will be at its steady-state level of diversity most of the time, so periodic selection will not significantly decrease the long-term average diversity.

If t_E is shorter than the recovery time, the average diversity depends on how much diversity on average has been recovered by the next selection event. Assuming that the time for the selected mutant to take over the population is significantly shorter than the time to the common ancestor in a steady-state population, the genealogical tree of a population that has been subject to a selection event looks like that of a population in a steady state, except its more distant history is effectively "cut off" by the event. The fraction $F(t)$ of the diversity capacity reached by time t can thus be approximated as $\sqrt{t}/A$ in one dimension, $(\log(t)/\log(A))^2$ in two dimensions, and $\log(t)/\log(N)$ in a well mixed population. Thus, initially the increase proceeds rapidly, but it slows down with time. A first approximation for the average diversity under frequent selection events is $B(t = t_E)$, where $B(t)$ is the diversity at t generations after a founding event, giving $B \sim A[\log(\frac{1}{\mu_R A})]^2$ in two dimensions. More comprehensive results for well mixed populations are given in Ref. 8.

References

1. Cho, J. C. & Tiedje, J. M. (2000) *Appl. Env. Microbiol.* **66**, 5448-5456.

2. Rauch, E. M. & Bar-Yam, Y. (2004) *Nature* **431**, 449-452, supporting information.

3. Rademaker, J. L. W., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P. & de Bruijn, F. J. (2000) *Int. J. Systematic Evol. Microbiol.* **50**, 665-677.

4. Tajima, F. (1993) *Mol. Biol. Evol.* **10**, 677-88.

5. Fisher, R. A. (1930) *Proc. Roy. Soc. Edinburgh*, **50**, 205-220.

6. Wright, S. (1931) *Genetics* **16**, 97-159.

7. Bramson, M. & Griffeath, D. (1980) *Z. fuer Wahrscheinlichkeitstheorie* **53**, 183-196.

8. Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989) *Genetics* **123**, 887-899.

9. Majewski, J. & Cohan, F. M. (1999) *Genetics* **152**, 1459-1474.