Convergent Evolution During Local Adaptation to Patchy Landscapes

Peter L. Ralph[1]
email: pralph@usc.edu

Graham Coop[2]
email: gmcoop@ucdavis.edu

[1] Computational Biology and Bioinformatics
University of Southern California, Los Angeles, CA
[2]Center for Population Biology & Department of Evolution and Ecology
University of California – Davis, Davis, CA, 95616

Abstract

Species often encounter, and adapt to, many patches of locally similar environmental conditions across their range. Such adaptation can occur through convergent evolution if different alleles arise and spread in different patches, or through the spread of shared alleles by migration acting to synchronize adaptation across the species. The tension between the two reflects the degree of constraint imposed on evolution by the underlying genetic architecture versus how effectively selection and geographic isolation act to inhibit the geographic spread of locally adapted alleles. This paper studies a model of the balance between these two routes to adaptation in continuous environments with patchy selection pressures. We address the following questions: How long does it take for a novel, locally adapted allele to appear in a patch of habitat where it is favored through mutation? Or, through migration from another, already adapted patch? Which is more likely to occur, as a function of distance between the patches? How can we tell which has occurred, i.e. what population genetic signal is left by the spread of migrant alleles? To answer these questions we examine the family structure underlying migration–selection equilibrium surrounding an already adapted patch, in particular treating those rare families that reach new patches as spatial branching processes. This provides a way to understand the role of geographic separation between patches in promoting convergent adaptation and the genomic signals it leaves behind. We illustrate these ideas using the convergent evolution of cryptic coloration in the rock pocket mouse, *Chaetodipus intermedius*, as an empirical example.

Author Summary: Often, a large species range will include patches where the species differs because it has adapted to locally differing conditions. For instance, rock pocket mice are often found with a coat color that matches the rocks they live in, these color differences are controlled genetically, and mice that don't match the local rock color are more likely to be eaten by predators. Sometimes, similar genetic changes have occurred independently in different patches, suggesting that there were few accessible ways to evolve the locally adaptive form. However, the genetic basis could also be shared if migrants carry the locally beneficial genotypes between nearby patches, despite being at a disadvantage between the patches. We use a mathematical model of random migration to determine how quickly adaptation is expected to occur through new mutation and through migration from other patches, and study in more detail what we would expect successful migrations between patches to look like. The results are useful for determining whether similar adaptations in different locations are likely to have the same genetic basis or not, and more generally in understanding how species adapt to patchy, heterogeneous landscapes.

Introduction

The convergent evolution of similar phenotypes in response to similar selection pressures is a testament to the power that selection has to sculpt phenotypic variation. In some cases this convergence extends to the molecular level, with disparate taxa converging to the same phenotype through parallel genetic changes in the same pathway, genes, or even by precisely the same genetic changes [Stern, 2013, Zhen

et al., 2012, Martin and Orgogozo, 2013]. Convergent adaptation also occurs within species, if different individuals adapt to the same environment through different genetic changes. There are a growing number of examples of this in a range of well studied organisms and phenotypes [Arendt and Reznick, 2008]. Such evolution of convergent phenotypes is favored by higher mutational input, i.e. higher total mutational rate and/or population size [Pennings and Hermisson, 2006]. The geographic distribution of populations can also affect the probability of parallel mutation within a species: a widespread species is more likely to adapt by multiple, parallel mutations if dispersal is geographically limited, since subpopulations will adapt via new mutations before the adaptive allele arrives via migration [Ralph and Coop, 2010]. Standing variation for a trait can also increase the probability of convergence, as this increases the probability that the selective sweep will be *soft* (beginning from a base of multiple copies), which whether or not the copies derive from the same mutation, leads to genetic patterns similar to convergent adaptation [Orr and Betancourt, 2001, Hermisson and Pennings, 2005].

Intuitively, convergence is also more likely when geographically separated populations adapt to ecologically similar conditions. The probability that convergent adaptations arise independently before adaptations spread between the populations by migration will be larger if these adaptive alleles are maladapted in intervening environments, since such adverse conditions can strongly restrict the spread of locally adapted alleles [Slatkin, 1973].

One elegant set of such examples is provided by the assortment of plant species that have repeatedly adapted to patches of soil with high concentrations of heavy metals (e.g. serpentine outcrops and mine tailings) [Kruckeberg, 1951, Macnair, 1991, Schat et al., 1996, Turner et al., 2010]; the alleles conferring heavy metal tolerance are thought to be locally restricted because they incur a cost off of these patches. Similarly, across the American Southwest, a variety of species of animals have developed locally adaptive cryptic coloration to particular substrates, e.g. dark rock outcrops or white sand dunes [Benson, 1933]. One of the best-known examples is the rock pocket mouse (*Chaetodipus intermedius*), which on a number of black lava flows has much darker pelage than on intervening areas of lighter rock [Dice and Blossom, 1937]. Strong predator-mediated selection appears to favour such crypsis [Kaufman, 1974], and, perhaps as a result of this strong selection against migrants, at least two distinct genetic changes are responsible for the dark pigmentation found on different outcrops [Hoekstra and Nachman, 2003]. Similar situations have been demonstrated in other small rodent systems [Steiner et al., 2009, Kingsley et al., 2009, Dice, 1940] and in lizards [Rosenblum et al., 2010].

In this paper, we study this situation, namely, when a set of alleles provide an adaptive benefit in geographically localized patches, separated by inhabited regions where the alleles are deleterious. The main questions are: Under what conditions is it likely that each patch adapts in parallel, i.e. convergently through new mutation, and when is it likely that migration carries these alleles between patches? How easy will it be to detect adaptive alleles that are shared by migration, i.e. over what genomic scale will a population genetic signal be visible?

We work in a model of continuous geography, using a variety of known results and new methods. In the section *Establishment by Mutation* we develop a simple approximation, equation (2), for the rate at which patches become established by new mutations. The most important conceptual work of the paper occurs in the section *Establishment by Migration*, where we develop an understanding of the process by which alleles move from an existing patch to colonize a novel patch, culminating in equation (11) for the rate at which this happens. We combine these two results in the section *Probability of Parallel Adaptation* to discuss to the probability of parallel adaptation, equation (12). To understand the genomic signal left by adaptations shared by migration, in the section *Haplotypes Shared Between Patches*, we approximate the time it will take for an allele to transit between patches, (equation (17)), and use this to approximate the length of haplotype that hitchhikes with it (equation (19)). Finally, in the section *Applications* we apply this work to understand the geographic scale of convergent evolution in *Chaetodipus intermedius*.

Results

Model of a patchy landscape

Consider a population spread across a landscape to which it is generally well adapted, but within which are patches of habitat to which individuals are (initially) poorly adapted. (When we refer to "patches" it is to these pieces of poor habitat.) Suppose it takes only a single mutational change to create an allele (B) that adapts an individual to the poor habitat type. The required change could be as specific as a

single base change, or as general as a knockout of a gene or one of a set of genes. We assume that this mutational change occurs at a (low) rate of μ per chromosome per generation, and that this initially rare new allele has fitness advantage s_p relative to the unmutated type (b) in these "poor" habitat patches. Outside of these patches we assume that the new allele has a fitness disadvantage of s_m when rare in the intervening areas, with s_p and s_m both positive. (Here we define "fitness" to be the intrinsic growth rate of the type when rare.) We assume that the disadvantage s_m is sufficiently large that, on the relevant timescale, the allele is very unlikely to fix locally in these intervening regions. (The case where $s_m = 0$ requires a different approach, which we do not treat here.)

We are interested in the establishment of mutations in the "poor" patches by either migration or mutation, and so will focus on whether the allele can escape initial loss by drift when rare. Therefore, we need not specify the fitness of the homozygote; only that the dynamics of the allele when rare are determined by the fitness of the heterozygote. More general dominance will only make small corrections to the dynamics until initial fixation, with the exception of the recessive case, which we omit. In other words, we follow the literature in treating the diploid model as essentially haploid.

Also, to avoid introducing more parameters, we assume population density ρ is equal across the range (even in the "poor" patches). If this is not the case, in most equations below, the population density that should be used is the density on the poor patches; see Lenormand [2002] for a more thorough approach. We further assume that the variance in offspring number is ξ^2, and that the mean squared distance between parent and child is σ^2 (i.e. σ is the dispersal distance). We will deal with migration by immediately appealing to the central limit theorem, treating the total distance traveled across t dispersal events as Gaussian with variance $t\sigma^2$, and do not discuss the (interesting) cases where rare, long-distance dispersal events are more important (see e.g. Levin et al. [2003], Ralph and Coop [2010], Hallatschek and Fisher [2014] for discussion).

Past work on spatial models We will make use of several previous results from the literature on migration in models of spatially varying selection. Haldane [1948], Fisher [1950], and later Slatkin [1973] first analyzed deterministic models of of this sort. Nagylaki [1975] and Conley [1975] showed that in one dimension, if the physical width of the patch is less than $(\sigma/\sqrt{2s_p})\tan^{-1}(\sqrt{s_m/s_p})$, then there is no stable equilibrium with both b and B present, so that migrational swamping prevents B from establishing (see also Lenormand [2002] for a review). Barton [1987], using general theory of Pollak [1966], showed that patches must be at least this critical size so that a new mutation has positive probability of establishment (in the large population density limit). The same paper also showed that this probability of establishment of a new mutation decays exponentially with distance from the patch, with the scale given by $\sigma/\sqrt{2s_m}$. Mutations appearing within the patch have probability of establishment strictly less than the probability for a panmictic population, but approaching this as the size of the patch increases. This latter panmictic probability of establishment we denote p_e, and often approximate p_e by $2s_p/\xi^2$ [Haldane, 1927, Fisher, 1930]. This result holds quite generally for a geographically spread population that experiences a uniform selection pressure [Maruyama, 1970, Cherry and Wakeley, 2003]. More general work on the mathematically equivalent problem of density-dependent population growth has obtained much more general criteria under which a habitat configuration will support a stable polymorphic equilibrium [Cantrell and Cosner, 1989], and determined the habitat configuration that maximizes the probability of establishment [Lou and Yanagida, 2006].

Establishment of a locally adaptive allele due to mutational influx

Consider first a single, isolated poor habitat patch in which no B allele has yet become established. We first compute the time scale on which a new B mutations will appear and fix in such a patch, if the rate of mutation from b to B is μ per individual per generation. As we are interested in patches where local adaptation can occur, we will assume that our patch is larger than the cutoff for local establishment mentioned above.

Let $p(x)$ be the probability that a new mutant B allele that arises at location x relative to the center of the patch fixes within the poor habitat patch. Under various assumptions, precise expressions can be found for $p(x)$ [Barton, 1987], but results will be more interpretable if we proceed with a simple approximation. The total influx per generation of mutations that successfully establish is given by the

integral of $p(x)$ over the entire species range:

$$\lambda_{\text{mut}} = \rho\mu \int p(x)dx \tag{1}$$

and hence depends in a complicated way on the patch geometry and selection coefficients, but still scales linearly with the mutational influx density $\rho\mu$. If we consider a patch of area A, whose width is large relative to $\sigma/\sqrt{2s_m}$, then a reasonable approximation is to ignore migration, so that $p(x) = p_e \approx 2s_p/\xi^2$ within the patch, and $p(x) = 0$ otherwise. This approximates the integrand $p(x)$ in (1) by a step function, which will be a good approximation if the patch is large relative the scale over which $p(x)$ goes from 0 to p_e, or if $p(x)$ is close to p_e at some point and is symmetric about the edge of the patch. We examine this more generally via exact calculation of $p(x)$ in the section *Probability of Establishment*.

The rate at which mutations arise and colonize a patch of area A is then

$$\lambda_{\text{mut}} \approx 2s_p\rho A\mu/\xi^2, \tag{2}$$

i.e. the product of mutational influx in the patch ($\rho A\mu$) and the probability of establishment ($p_e \approx 2s_p/\xi^2$). If this rate is low, then the time (in generations) until a mutation arises that will become locally established within the patch is exponentially distributed with mean $1/\lambda_{\text{mut}}$. Assuming that once a mutation becomes established it quickly reaches its equilibrium frequency across the patch, the time scale on which new patches become colonized by the B allele from new mutation is therefore approximately $1/\lambda_{\text{mut}}$.

Establishment of a locally adaptive allele due to migrational influx

Now suppose that there are two patches, each of area A and separated by distance R. If the B allele has arisen and become established in the first patch, but has not yet appeared in the second, we would like to know the time scale on which copies of B move between patches by migration (supposing that no B allele arises independently by mutation in the meantime). To determine this, we study the fine-scale genealogy of alleles that transit between patches, obtaining along the way other useful information about the genealogy of B alleles. Doing this we arrive at (11) for the rate at which an allele established in one patch spreads to a neighboring one by migration.

Migration–selection balance ensures that there will be some number of B alleles present in the regions where they are disadvantageous, but only very few, and rarely, far away from the patch where B is established. Denote the expected frequency of allele B at displacement x relative to the patch by $q(x)$, and assume that the shape of the patch is at least roughly circular. Following Haldane [1948] or Slatkin [1973], one can write down a differential equation to which $q(x)$ is the solution, and show that $q(x)$ decays exponentially for large $|x|$, with a square-root correction in two dimensions:

$$q(x) \approx C\left(|x|\sqrt{2s_m}/\sigma\right)^{-(d-1)/2} \exp(-|x|\sqrt{2s_m}/\sigma) \qquad \text{for large } |x|, \tag{3}$$

where d is the dimension ($d = 1$ or 2), and C is a constant depending on the geographic shape of the populations and the selection coefficients. In applications we fit C to the data; to get concrete numbers from theory we take $C = 1$ if necessary.

From simulations, C seems to be not too far from 1 in practice (see below). As J. Hermisson pointed out in comments on an earlier draft, this functional form has a simple intuition: local migration leads to the exponential decay, since if each migrant at distance x produces an average of c descendants that make it to distance $x + 1$, then in one dimension, the number of migrants must decay as $\exp(-cx)$; and the square-root term in two dimensions enters because the available area grows with x. The "renewal" aspect of this same argument suggests that for equation (3) to hold, $|x|$ must be larger than a few multiples of σ.

These asymptotics fit simulations quite well, as shown in Figure 1. To be clear about the assumptions (especially regarding scaling) implicit here, below provide a derivation of (3) in section *Equilibrium Frequency*, as well as justification for the asymptotics below in section *Asymptotics*. In one dimension, the equation can be solved to give $q(x)$ in terms of Jacobi elliptic functions; see appendix A.

This expected frequency $q(x)$ is the *time-averaged* occupation frequency, i.e. the total number of B alleles found across T generations near location x, per unit area, divided by T. If, for instance, $q(x) = .01$ and the population density is $\rho = 10$ individuals per unit area, then in a survey tract of

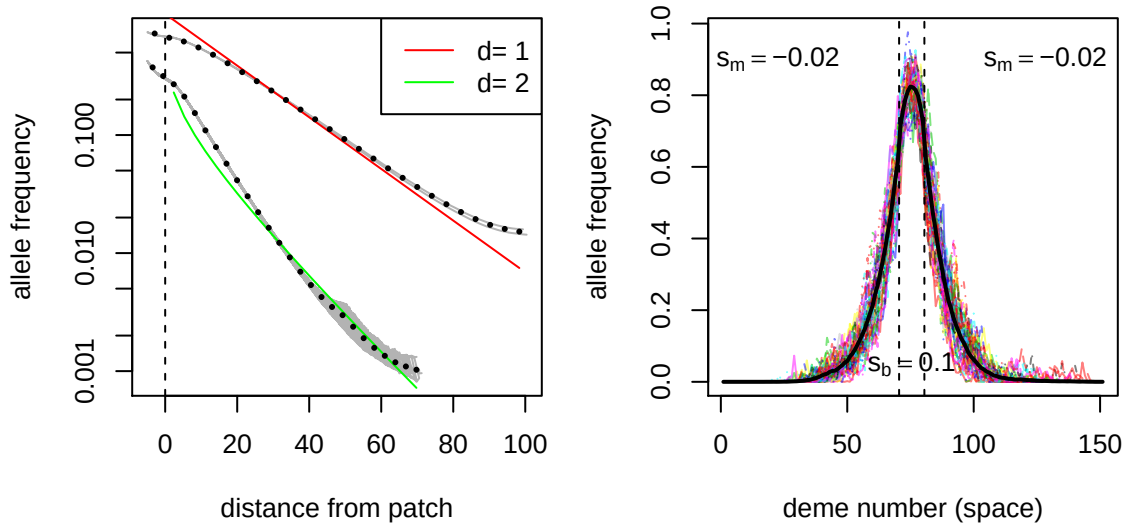

Figure 1. **(A)** Mean occupation frequencies of individual demes (grey dots) and averaged over demes at similar distances (black circles) from simulations, on a one-dimensional grid of 201 demes, and a two-dimensional grid of 101×101 demes. Superimposed in color is the decay predicted by (3), with C chosen arbitrarily to give a reasonable fit: 0.85 and 0.11 respectively. (The disagreement at long distances is due to boundary effects near the edge of the range.) **(B)** Fluctuations: colored lines are snapshots of allele frequencies at evenly spaced times from a one-dimensional grid of 151 demes; the solid line shows the cumulative mean occupation frequency across all generations. For both plots, the simulation was run for 1000 generations; the focal allele is beneficial in the central 10 demes, with $s_p = 0.1$, and deleterious elsewhere, with $s_m = -0.02$; demes had 1000 individuals in each; dispersal was to nearby demes with mean dispersal distances of $\sigma = 0.95$ and 0.74 times the deme spacing, respectively. More details are given below in *Simulation methods*.

unit area around x we expect to find one individual every 10 generations – or, perhaps more likely, 10 individuals every 100 generations. This points to an important fact: close to the original patch, the "equilibrium frequency" $q(x)$ describes well the density of the B allele at most times, but far away from the patch, the equilibrium is composed of rare excursions of families of B alleles, interspersed by long periods of absence. An example of one of these rare excursions is shown in Figure 2. The relevant time scale on which B alleles migrate between patches is given by the rate of appearance of such families. To compute the rate of establishment by such migration, we argue that these families (suitably defined) can be approximated by independent branching processes whose spatial motion is described by Brownian motion. Properties of these, combined with equation (3), will lead us eventually to the expression (11) for the rate of effective migration, an approximation that is valid if the distance R is large (at least several times $\sigma/\sqrt{2s_m}$).

To make the argument, we decompose the genealogy of B alleles into families in the following way. First, consider the genealogy of all B alleles that were alive at any time outside of the original patch. This is a collection of trees, each rooted at an allele living outside the patch whose parent was born inside the patch. Next, fix some intermediate distance r_0 from the established patch, and erase from the genealogy every allele that has no ancestors living further away than r_0 to the patch. This fragments the original set of trees into a collection of smaller trees that relate to each other all B alleles living outside of r_0 at any time, and some living inside of r_0; we refer to these trees as "families". If r_0 is large enough that there are not too many families and interactions between family members are weak, then these families can be approximated by a collection of independent spatial branching processes whose members are ignored if they enter the original patch. (This can be made formal in the limit of large

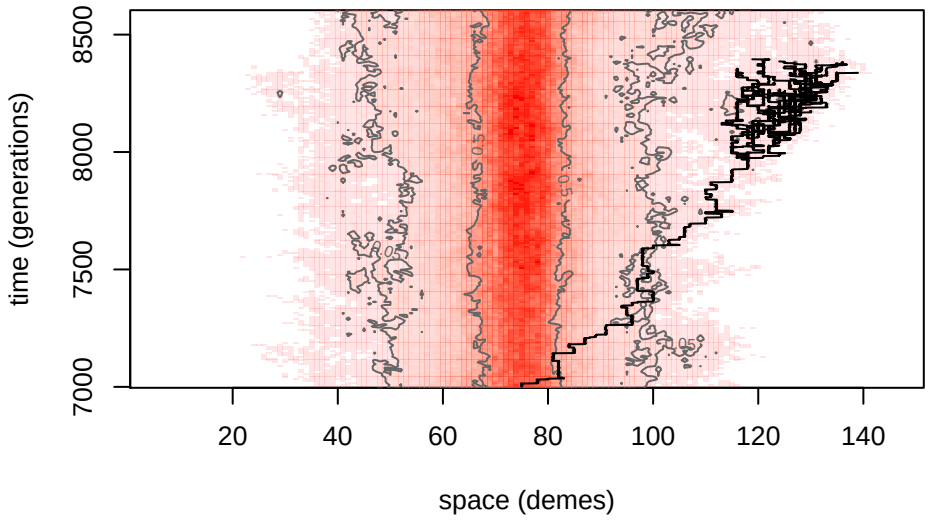

Figure 2. A segment of the temporal dynamics of the same simulation as Figure 1B. Space is again shown across the horizontal axis, with time on the vertical axis; demes are colored darker red the greater their allele frequency at the corresponding time. The 50% and 5% frequency contour lines are in grey, and the genealogy of 50 individuals in a rare long-distance excursion is traced back with black lines.

population density, also taking r_0 large enough that the chance of reaching the original patch is small.) This decomposition is illustrated in figure 3. In terms of these families, the equilibrium frequency above is then

$$\rho q(x) = \text{ (outflux of families) } \times \text{ (mean occupation density of a family at } x \text{) }, \tag{4}$$

and the quantity we wish to compute, the effective rate at which *families of* migrant B alleles establish in a new patch at displacement x from the original, is

$$\lambda_{\text{mig}}(x) = \text{ (outflux of families) } \times \text{ (probability a family establishes in patch at } x \text{) } \tag{5}$$

The quantities on the right-hand side depend implicitly on r_0, but their product does not. All terms except the "outflux of families" are calculations with spatial branching processes; and we can solve for this outflux using the established form for $q(x)$.

The genealogy of migrant families

To understand the process of adaptation by migration, we therefore need to look more closely at the rare families of B alleles far from the original patch where it is already established. Consider one such family, that lives far from the original patch. As noted above, we will model this as a branching process, i.e. by assuming that each individual migrates, then reproduces, independently of all other family members. Migration takes the individual in a random direction for a random distance whose mean square is σ^2, and if the individual arrives to a patch where the B allele is favored, we drop the individual from the family. Since the B allele is uniformly deleterious elsewhere, we can decouple reproduction and migration by first forming a tree in which each individual has a random number of offspring, then recursively assigning spatial locations to each individual in the tree, and finally pruning any branches that happen to wander into a patch. By our definition of fitness, in this tree the mean number of offspring per individual is $e^{-s_m} < 1$. This pruning removes both branches leading back into the already established patch (which are irrelevant) and branches leading into the new patch (we will need to count these). Since we are concerned with the rare migrant families that transit quickly between patches, back migrants to the

already established patch will contribute little, since they are going the wrong direction; so to simplify the analysis we ignore this detail.

Since each individual has on average e^{-s_m} offspring, the expected family size after time t is e^{-ts_m}. Therefore, if we define $k_e(t)$ to be the chance that the family has gone extinct by time t, and let K_t be the (random) family size conditioned on nonextinction, then by Bayes' rule, $e^{-s_m t} = (1 - k_e(t))\mathbb{E}[K_t]$. It turns out that if the distribution of offspring numbers is not too heavy-tailed (see Jagers [1975] for details), then K_t has a limiting distribution: $K_t \xrightarrow{d} K$, and

$$e^{-s_m t}/(1 - k_e(t)) \to \mathbb{E}[K] \quad \text{as} \quad t \to \infty. \tag{6}$$

In other words, the chance of survival to time t is asymptotically a constant multiplied by $e^{-s_m t}$, and conditional on survival, the family size is given by a fixed distribution.

This suggests that families can be well-described by motion of a central "trunk", which dies at a random time whose distribution is given by $k_e(t)$, and conditional on survival of this trunk, the family is composed of side "branches" off of this trunk, as depicted in Figure 3. We can understand this quite concretely, by a method described rigorously by Geiger [1999] (see also Chauvin et al. [1991]). When we condition on survival until t, we condition on existence of at least one trunk lineage surviving from time 0 to time t (the long red lineage in Figure 3). Given this lineage, nothing distinguishes any of the other reproduction events – each individual along the trunk lineage must give birth to at least one offspring, and all other reproduction events are unconditioned. In other words, the genealogy of a family that survives until t can be decomposed into a trunk that lasts until t and that sprouts independent branches that may or may not survive until t. Specifically, if we write Z_t for the number of offspring in the branching process alive at time t, then this gives the following decomposition:

$$(Z_t \mid Z_t > 0) \stackrel{d}{=} 1 + \sum_{s=0}^{t} \sum_{k=1}^{W_s - 1} Z_{t-s}^{(s,k)}, \tag{7}$$

where each $Z^{(s,k)}$ is an independent copy of Z that arose from the main branch in generation s, and each W_s is an independent copy of the offspring distribution W, conditioned on $W_s \geq 1$. Since each subfamily $Z^{(s,k)}$ is a branching process with mean offspring number less than one, most of these die out quickly, and so Z_t is composed of a cluster of recently split subfamilies. If we take the expectation of (7), since $\mathbb{E}[Z_t] = e^{-s_m t}$, we get the limiting mean family size conditioned on nonextinction:

$$\mathbb{E}[K] = 1 + \mathbb{E}[W - 1 \,|\, W \geq 1]/(1 - e^{-s_m}) \approx \mathbb{E}[W - 1 \,|\, W \geq 1]/s_m. \tag{8}$$

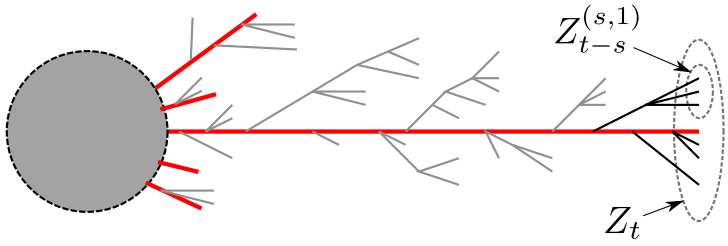

Figure 3. Cartoon of the decomposition described in the text. Migrants families that get further than r_0 from the already-established patch (grey circle) are each composed of a "trunk" lineage (in red), whose spatial motion we follow, embellished by transient branches. For the depicted long-lived family, these are further distinguished by those still alive at time t (in black) and those that did not survive until t (in grey). Figure 2 has a depiction of such a family, from a simulation. Note that the spatial motions of lineages are independent random walks (and hence should be much wigglier than in this cartoon, which confounds space and time).

Heuristics

The parallel form suggested by (4) and (5) suggests that we can find $\lambda_{\text{mig}}(x)$ by multiplying the mean density $q(x)$ by the ratio of the second terms in the two equations, i.e.

$$\lambda_{\text{mig}}(x) = \rho q(x) \times \frac{(\text{ probability a family establishes in patch at } x\ )}{(\text{ mean occupation density of a family at } x\ )}. \tag{9}$$

To make this precise we need to understand what we mean by "outflux of families"; we do this below in section *Hitting and Occupation*, and provide a heuristic argument here.

Both numerator and denominator in (9) count only families that have arrived at the new patch, so we can consider a family conditioned on having successfully transited between the patches. To obtain simply interpretable expressions, we assume that once any member of the family arrives at the patch, all other members are close by, and so have roughly equal opportunity to establish in the new patch.

Suppose the family arrives with K individuals. The probability that at least one individual leaves a lineage that establishes in the patch is $1-(1-p_e)^K$. On the other hand, in the absence of the patch of new habitat, each of the K newly arrived individuals would leave behind a lineage with mean total size $\sum_{t=0}^{\infty} e^{-ts_m} \approx 1/s_m$, roughly half of which would stay in the new patch. Therefore, the expected contribution of the family to the occupation density is $K/2s_m$. (Readers who feel uncomfortable about the geometry of the patch, or the distinction between occupation density and number of individuals are encouraged to take a detour through section *Hitting and Occupation*, below.)

This suggests that

$$\lambda_{\text{mig}}(x) = \rho q(x) \times \frac{\mathbb{E}[1-(1-p_e)^K]}{\mathbb{E}[K]/2s_m}. \tag{10}$$

There are now two extremes to consider. If $\mathbb{P}\{K > 1/p_e\}$ is small, then each family has a small chance of establishment, and the numerator is approximately $p_e\mathbb{E}[K]$. On the other hand, if p_e is not too small, then the numerator would be close to 1, as any family that arrives will likely establish. Combining these, and absorbing the term $\mathbb{E}[W|W>1]$ from equation (8) into the constant C,

$$\lambda_{\text{mig}}(x) \approx C\rho s_m \min(s_m, p_e) \left(|y|\sqrt{2s_m}/\sigma\right)^{-(d-1)/2} \exp(-|y|\sqrt{2s_m}/\sigma) \qquad \text{for } |y| > \sigma/\sqrt{2s_m} \tag{11}$$

When we use this below, we set $p_e = 2s_p/\xi^2$. Note that the approximation has the undesirable property that it is non-monotonic in s_m, when we expect the rate of adaptation via migration to be a decreasing function of s_m. However, it only increases for small s_m, and is a decreasing function of s_m for $s_m > |y|^2/2\sigma^2$, which is the parameter region that we are restricted to by other assumptions.

A review, and a test, of the assumptions

Now that we have expressions for the mean rates of adaptation by new mutation, (2), and by migration from an already colonized patch, (11), it seems helpful to step back and review the assumptions underlying the asymptotic results we have used, or will use below. Our results should hold exactly in the limit of large, circular patches that are much further apart than they are wide, large population density, and small selection coefficients of equal magnitude.

To be more precise, the mutation rate results will hold best if $s_m \approx s_p$, if the patch diameter is large relative to $\sigma/\sqrt{\min(s_m, s_p)}$, and the patch is not too far from circular. If $s_m \ll s_p$ then (2) will underestimate the rate, and will overestimate if $s_m \gg s_p$. As for the migration rate, we assume that each patch is large enough to support a stable population of B alleles ($A^{1/d} > (\sigma/\sqrt{2s_p})\tan^{-1}(\sqrt{s_m/s_p})$). The geometry and size of the patch will also affect the approximation of equation (50). Next, in using equation (3), we assume that the inter-patch distance is large relative to the characteristic length ($R > \sigma/\sqrt{s_m}$), and that local drift is not so strong that the equilibrium frequency is at least approximately attained (using Wright's effective neighborhood size, $s_m \gg 1/(\rho\sigma^d)$). The last requirement is necessary because if the B allele fixes in large neighborhoods where it is deleterious, we cannot approximate its dynamics via a branching process. We also neglect the time for migration-selection equilibrium to be reached. As discussed above, we also assume that migration to a new patch takes place over a number of generations; if there are sufficient rare, long-distance migration events that would move between patches in a single hop, this would require a different analysis.

To test the robustness of our results to the various approximations we used, we used individual-based simulations on a one-dimensional lattice of 501 demes, with ρ haploid individuals per deme, dispersal to nearby demes with $\sigma = 0.95$, and run for 25,000 generations. More details are given below in section *Simulation methods*, and the number of simulations used and parameter values are given in supplementary tables 1 and 2. To estimate the mean time until adaptation by mutation, we used one centrally located patch of 99 demes and a mutation rate of $\mu = 10^{-5}$, and to estimate the mean time until adaptation by migration, we used two centrally located patches of 99 demes separated by a varying number of demes, with one patch initialized to have the B allele at frequency 0.8. In each case, we measured the "time to adaptation" as the amount of time until at least 100 B alleles were present in the previously unadapted patch. Figure 4 summarizes how the results compare to theory, excluding parameter combinations where a majority of simulations did not adapt by 25,000 generations or the patch was not large enough to support a stable population of B alleles. Supplementary figures S1 and S2 show all times, and depictions of typical simulation runs are shown in supplementary figures S3, S4, S5, S6, S7, S8, S9, and S10.

The agreement is reasonable for both. The theoretical value $1/\lambda_{\text{mut}}$ underestimates the mean time to mutation by a factor of around 2 that depends increases with s_m. This is to be expected for two reasons: First, we compute the time to reach 100 B alleles, while theory predicts the time until the progenitor of those 100 B alleles arose. Second, the expression for λ_{mut} neglects effects near the boundary of the patch, and the larger s_m is, the harder it is for mutations that arise near the edge of the patch to establish. The theoretical values $1/\lambda_{\text{mig}}$ again has a margin of error of a factor of about 2, with the exception of the weakest disadvantage off the patch, $s_m = 10^{-4}$, which adapt much more quickly than predicted by theory. However, these simulations fall outside the region where we expect the model to do well, since $R < \sigma/\sqrt{s_m} \approx 100$. A glance at supplementary figure S10 shows more: since s_m is so small, the second patch is reached by migrants before the migration-selection equilibrium is reached, and at migration-selection equilibrium, the two patches act as a single meta-patch.

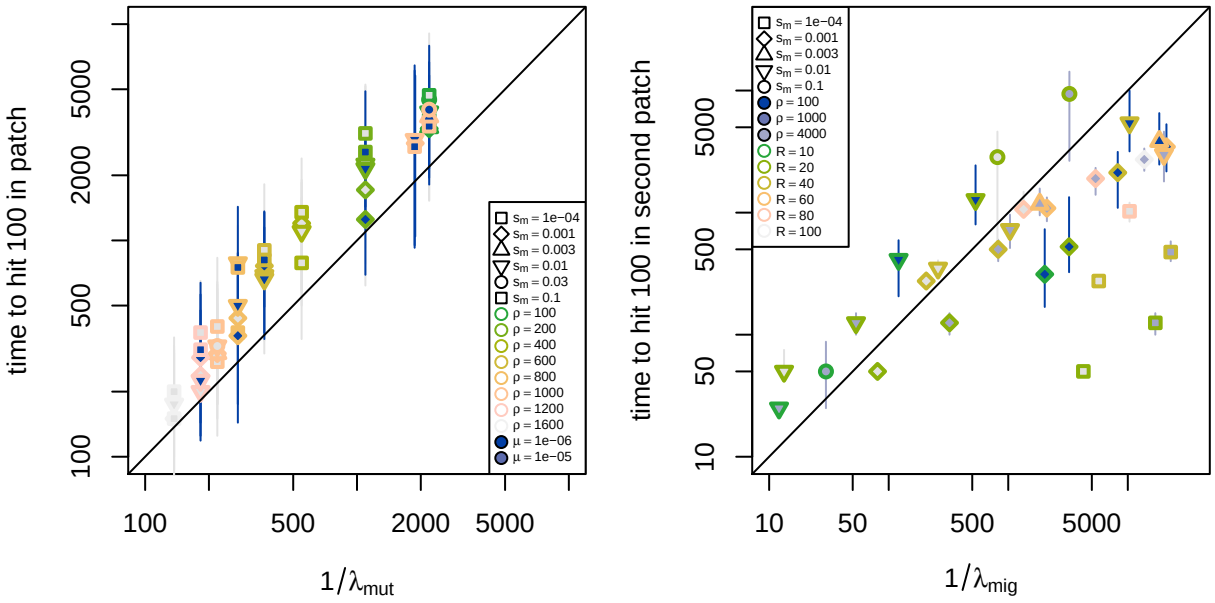

Figure 4. (left) Time until adaptation by mutation in an isolated patch for different parameter combinations, as observed in simulations (vertical axis), and as predicted from equation (2) (horizontal axis). **(right)** Time until adaptation by migration from an already occupied patch to a new patch, as observed in simulations (vertical axis), and as predicted from equation (11), with $C = 5$ (horizontal axis). For each, parameter combinations and numbers of simulations are in supplemental tables 1 and 2; shown are the median (points) and interquartile range (lines) of the time until the patch had at least 100 adapted individuals, for each unique parameter combination.

The probability of parallel adaptation between patches.

We now turn to the question of whether the separated patches adapt by parallel genetic changes or by the spread of a migrant allele between the patches. As only a single mutation is required for individuals to adapt to the poor habitat patch, subsequent mutations that arise after an allele becomes established in the patch gain no selective benefit. Similarly, an allele introduced into a patch by migration will not be favored by selection to spread, if the patch has already been colonized. Therefore, mutations selectively exclude each other from patches, over short time scales, and will only slowly mix together over longer time scales by drift and migration, an assumption we also made in [Ralph and Coop, 2010]. In reality, different mutations will likely not be truly selectively equivalent, in which case this mixing occurs over a time-scale dictated by the difference in selection coefficients.

We will assume that once a B allele is introduced (by migration or mutation) it becomes established in the poor habitat patch rapidly if it escapes loss by demographic stochasticity. Under this approximation, and the "selective exclusion" just discussed, after one of the patches becomes colonized by a single B allele, the other patch will become colonized by migration or mutation, but not by both. As such, the question of how the second patch adapts simply comes down to whether a new mutation or a migrant allele is the first to become established in the second patch. To work out how adaptation is likely to proceed, we can compare expressions (2) and (11) above for the effective rates of adaptation by new mutation and by migration. We work in one dimension, as the square root term appearing for two dimensions is relatively unimportant (as seen in figure 1 by the fact that the $d = 2$ curve is nearly straight).

We first consider the order of magnitude that our parameters need to be on in order for adaptation via mutation or migration to dominate. Let $\gamma = \min(1, s_m/p_e)$. Effective migration and mutation rates will be on the same order if $A\mu \approx 2\gamma s_m \exp(-R\sqrt{2s_m}/\sigma)$, where R is the distance between the patches, and A is the area of the not-yet-adapted patch. In other words, the migrational analogue of "mutational influx" ($\mu\rho A$) is $2\rho\gamma s_m \exp(-R\sqrt{2s_m}/\sigma)$, which depends fairly strongly on the selective disadvantage s_m between patches. Equivalently, the rates are roughly equal if $R/\sigma = \log(2\gamma s_m/(A\mu))/\sqrt{2s_m}$, which gives the critical gap size past which adaptation will be mostly parallel in terms of selection coefficients, patch size, and mutation rate.

If we take the mutational influx in a patch to be one per 10,000 generations ($\rho A\mu = .0001$), $s_m > p_e$ so that $\gamma = 1$, and the population density to be 1 ($\rho = 1$, so $-\log(A\mu) \approx 9.2$) then adaptation is mostly parallel for patches separated by gaps larger than $R/\sigma > (9.2+\log(2s_m))/\sqrt{2s_m}$. If the selective pressure is fairly strong (say, $s_m = .05$), then for convergence to be likely, the distance between patches must be at least 21.8 times the dispersal distance (i.e. $R/\sigma > 21.8$). If s_m is much smaller, say $s_m = .001$, the required distance increases to $R/\sigma > 67$. On the other hand, if the separation is much smaller than this transitional value, then adaptations are likely to be shared between patches.

If the mutational influx is higher – say, $\rho A\mu = .01$ – the required separation between patches is only reduced to $R/\sigma > 7.3$ with $s_m = .05$. If $s_m = .001$, then with this higher mutational influx this calculation predicts that mutation will *always* be faster than migration – however, this should be taken with caution, since as discussed above, this model does not hold if R is of the same order as σ or if s_m is small enough that local drift is more important.

We can go beyond these orders of magnitude calculations to find the probability of parallel adaptation if we are willing to take our approximations at face value. Doing so, the time until the first of the two patches is colonized by the B allele will be approximately exponentially distributed with mean $1/(2\lambda_{\text{mut}})$. Following this, the time until the second patch is subsequently colonized (via either migration or new mutation) will be exponentially distributed with mean $1/(\lambda_{\text{mut}} + \lambda_{\text{mig}})$. Both these rates scale linearly with population density (ρ) and the probability of establishment of a rare allele ($p_e \approx 2s_b/\xi^2$, for $p_e > s_m$), so that the overall time to adaptation is robustly predicted to increase with offspring variance (ξ^2) and decrease with population density and beneficial selection coefficient. Furthermore, the probability that the second adaptation is a new mutation, i.e. the probability of parallel adaptation, with $\gamma = \min(1, s_m/p_e)$, is

$$\frac{\lambda_{\text{mut}}}{\lambda_{\text{mut}} + \lambda_{\text{mig}}(R)} = \frac{A\mu/(2s_m)}{A\mu/(2s_m) + C\gamma\left(R\sqrt{2s_m}/\sigma\right)^{-(d-1)/2}\exp\left(-\sqrt{2s_m}R/\sigma\right)}, \tag{12}$$

so that the probability of parallel mutation should increase approximately logistically with the distance between the patches, on the spatial scale $\sigma/\sqrt{s_m}$.

We tested this using the same simulations as figure 4, by using the empirical distributions of the respective times to adaptation to estimate the probability, for each parameter combination, that a new, successful mutation appears before a successful migrant arrives from another, already adapted patch. The results are compared to (12) in figure 5.

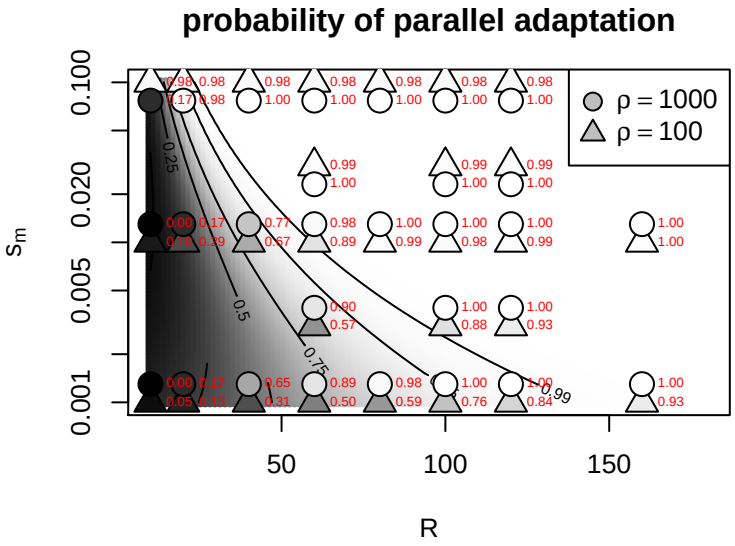

Figure 5. Probability that two patches evolve by independent mutations (i.e. convergently), as a function of their geographic separation (R) and the strength of selection between patches (s_m). The background shades of grey and contours give the probability predicted by theory (equation (12), with $C = 5$). Points correspond to parameter combinations of the simulations in figure 4, with the shade of grey and the red numeric label indicating the same probability estimated from these simulations (see text for details). The two densities, $\rho = 100$ and $\rho = 1000$ would lie directly on top of each other, so the $\rho = 1000$ points have been shifted vertically.

Multiple patches

Above we worked with two patches separated by distance R; but these calculations apply as well to multiple patches in which the B allele can facilitate local adaptation. Under our assumptions of selective exclusion and rapid establishment, each patch will adapt through establishment of a single copy of the B allele, either by migration or mutation. If we imagine that the times between successive establishments are the result of many nearly-independent attempts with small probability of success, the process of occupation is well approximated by a Markov chain, whose state records the labeling of patches according to the colonizing allele (or none, for those that haven't yet adapted).

If not yet adapted, a focal patch of area A will acquire the adaptation by new mutation at rate $2s_p\mu\rho A/\xi^2$ (equation (2)). Suppose that patches $1,\ldots,k$ are already adapted, and that these patches are at distances $R_1,\ldots,R_k$ away from this unadapted patch. The total rate of adaptation through migration, from equation (11) if patches do not interfere with each other, is

$$\sum_{i=1}^{k}\lambda_{\mathrm{mig}}(R_i) = C\rho s_m \min(s_m,p_e)\sum_{i=1}^{k}\left(R_i\sqrt{2s_m}/\sigma\right)^{-(d-1)/2}\exp\left(-R_i\sqrt{2s_m}/\sigma\right). \tag{13}$$

If the patch adapts through migration, the probability the adapted allele comes from patch i is

$$\frac{(R_i)^{-(d-1)/2}\exp\left(-R_i\sqrt{2s_m}/\sigma\right)}{\sum_{j=1}^{k}(R_j)^{-(d-1)/2}\exp\left(-R_j\sqrt{2s_m}/\sigma\right)}. \tag{14}$$

Equations (13) and (14) specify the transition rates of the process.

Since the compound parameter $s_p\rho$ is common to both migration and mutation rates, it functions only as a time scaling of the process, and therefore has no effect on the final configuration, namely, the sets of patches sharing the genetic basis of their adaptive response. Also note that we can rescale time by the typical patch size, and introduce a parameter, say $a_k = A_k/\bar{A}$, making the properties (other than the time scaling) of the discrete model independent of the *numerical sizes* of the demes themselves. This is complementary to the results of Pennings and Hermisson [2006], who showed that multiple mutations are likely to arise *within* a panmictic deme if the population-scaled mutation rate $2N\mu$ is greater than 1.

The resulting spatially discrete process is quite general, but if we oversimplify back to the island model, it is tractable: suppose the probability of adaptation by mutation is the same for each patch and the probability of migration is the same between each pair of patches. The resulting process is a continuous-time version of the "Chinese restaurant process" described in Aldous [1985] and Pitman [1995], and so the final partition of types across demes has the Ewens sampling distribution with parameter $A\mu/(4s_m \exp(-\sqrt{2s_m}R/\sigma))$. Now we can compute most properties we might want about the process. For instance, the proportion of demes that shares the same origin as a randomly sampled deme is approximately Beta distributed – see Donnelly and Joyce [1989]. The high connectedness of the discrete deme island model means that the expected number of distinct alleles grows with the log of the number of demes. This strongly contrasts with the continuous spatial model where the local nature of dispersal means that doubling the species range will double the number of mutations expected.

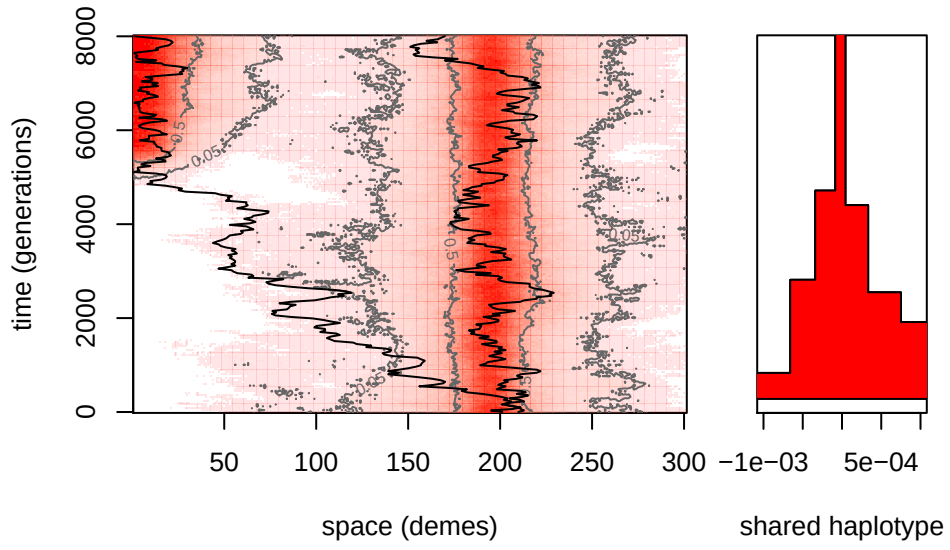

Figure 6. Adaptation by migration: An initially unadapted patch (10 demes at the left end of the range) is colonized around generation 5000; a representative lineage is shown in the new patch, which follows the original migrant "trunk" lineage between the patches and coalesces with a lineage in the unadapted patch. In the left panel, as in Figure 2, color and contours show allele frequency across space and time; other simulation parameters are as in Figures 1B and 2. The right panel shows (in red) the extent of the shared haplotype between the two lineages around the selected locus (at position 0), which decreases as time progresses.

Length of the hitchhiking haplotype

If a patch adapts through new mutation or a rare migrant lineage, the genomic region surrounding the selected site will hitchhike along with it [Maynard Smith and Haigh, 1974], so that at least initially, all adapted individuals within a patch share a fairly long stretch of haplotype. Pairs of adapted individuals

within a patch will initially share a haplotype of mean genetic length of about $\log(\rho A s_p)/s_p$ around the selected site, as long as the patch is reasonably well mixed by dispersal (otherwise see Barton et al. [2013]). This association gets slowly whittled down by recombination during interbreeding at the edge of the patch, but there will always be longer LD nearby to the selected site [Barton, 1979].

When an already adapted patch colonizes another through migration, the newly colonized patch will inherit a long piece of haplotype around the selected site from the originating patch. The genetic length of this haplotype is roughly inversely proportional to the number of generations the allele spends "in transit", because while very rare, the allele will be mostly in heterozygotes, and so each generation provides another opportunity for a different haplotype to recombine closer to the transiting B allele. A large, linked haplotype may still arrive and fix in the new patch, in which case the haplotype has literally hitchhiked across geography! Figure 6 shows a simulation of such an event, including the lineage that founds an adaptive population on a second patch, and the length of the haplotype shared between this lineage and one in the original patch. The entire time that the lineage is outside the region where the B allele is common (dark red in Figure 6), the haplotype that accompanies it is broken down rapidly. After the lineage establishes on the patch, the rate of decay of the haplotype is slowed significantly, since most others with which it recombines have similar haplotypic backgrounds. We first discuss the transit of the lineage, and its accompanying haplotype, and return to how haplotypes are whittled down within established patches in the next section.

Above we argued that a good model for this transit is a single Brownian trunk lineage surrounded by a cloud of close relatives of random size K whose chance of surviving until t generations is $1 - k_e(t)$. Consider a single such family, and let τ be the (random) time at which it first hits the new patch, with $\tau = \infty$ if it never does. We assume that the length of the hitchhiking segment depends only on the time spent by the trunk lineage of the first successful migrant "in transit" between the patches, i.e. τ conditioned on $\tau < \infty$. In so doing, we neglect recombination between B alleles (since they are at low density in transit), and the possibility that more than one successful migrant family is in transit at once (so that faster migrants would be more likely to have arrived first). These factors should have a minor impact if the rest of our framework fits.

Our first obstacle is that we do not know $1 - k_e(t)$, the probability a family survives until t, exactly. However, we do know that $1 - k_e(t) \simeq e^{-s_m t}/\mathbb{E}[K]$, i.e. that it has exponential tails. This suggests one further approximation, that the random lifetime of a family is exponentially distributed with mean $1/s_m$; equation (8) shows this introduces a constant into the equations below. Once we do this, we can use standard results on hitting times of d-dimensional Brownian motion that is killed at rate s_m (see Borodin and Salminen [2002] 2.2.0.1 and 4.2.0.1). In particular, if the patch is circular with radius w and lies at distance R from the already adapted patch, then we know that the Laplace transform of the transit time τ is

$$\mathbb{E}[e^{-\ell\tau}] = \begin{cases} e^{-R\sqrt{2(s_m+\ell)}/\sigma} & d = 1 \\ \frac{K_0((R+w)\sqrt{2(s_m+\ell)}/\sigma)}{K_0(w\sqrt{2(s_m+\ell)}/\sigma)} & d = 2, \end{cases} \tag{15}$$

where K_0 is a modified Bessel function of the second kind. We are interested in lineages that manage to reach the patch before being killed, i.e. having $\tau < \infty$, which occurs with probability $\mathbb{P}\{\tau < \infty\} = \lim_{\ell \to 0} \mathbb{E}\left[\exp(-\ell\tau)\right]$.

To keep the expressions simple, in the remainder of this section we only compute quantities for one dimension. By Bayes' rule, the Laplace transform of τ for a successful lineage is

$$\mathbb{E}[e^{-\ell\tau} | \tau < \infty] = \exp\left\{-\frac{R}{\sigma}\left(\sqrt{2(\ell + s_m)} - \sqrt{2s_m}\right)\right\}. \tag{16}$$

This can be differentiated to find that the expected transit time is

$$\mathbb{E}[\tau] = (R\sqrt{2s_m}/\sigma) \times 1/(2s_m) \tag{17}$$

and that $\mathrm{Var}[\tau] = (R\sqrt{2s_m}/\sigma) \times 1/(2s_m)^2$.

Since each generation spent in transit provides an opportunity for recombination, if recombination is Poisson, the length of the haplotype (in Morgans) initially shared between populations on each side of the selected locus is exponentially distributed with mean τ, so that if L is the length of hitchhiking segment on, say, the right of the selected locus, then

$$\mathbb{P}\{L > \ell\} = \mathbb{E}[e^{-\ell\tau} | \tau < \infty], \tag{18}$$

and so (one minus) equation (16) in fact gives the cumulative distribution function of L. The form of this distribution function implies that if Y is an exponential random variable with rate $R\sqrt{2}/\sigma$, then L has the same distribution as $(Y + \sqrt{s_m})^2 - s_m$. Furthermore, the expected length of shared hitchhiking haplotype is

$$\mathbb{E}[L] = \sigma\sqrt{2s_m}/R + \sigma^2/R^2 \tag{19}$$

and $\mathrm{Var}[L] = 2s_m\sigma^2/R^2 + 4\sigma^3\sqrt{2s_m}/R^3 + 5\sigma^4/R^4$. For two dimensions, asymptotics of Bessel functions show that L has the same tail behavior, but how other properties might change is unclear.

As a general rule of thumb, equation (17) means that families who successfully establish move at speed $\sqrt{s_m}\sigma$ towards the new patch – if s_m, the strength of the selection against them, is smaller, the need to move quickly is less imperative. Then, equation (19) means that the haplotype length is roughly the length one would expect given the mean transit time. Note that we have become used to seeing σ divided by $\sqrt{2s_m}$, rather than multiplied; it appears here because $\sigma/\sqrt{2s_m}$ is a length scale, and to convert it to a speed this is divided by $1/2s_m$.

In conclusion, the length of shared haplotype drops off quickly with the distance between the patches. When s_m is small we expect shorter haplotypes as the allele is only weakly deleterious and so successful migrants traverse the distance between patches relatively slowly compared to successfully migrating strongly deleterious alleles. Therefore, when adaptation by migration is most likely (small s_m) we expect shorter shared haplotypes.

Applications

Coat color in the rock pocket mouse *Chaetodipus intermedius* is a classic example of local adaptation [Benson, 1933, Dice and Blossom, 1937]. These mice live on rocky outcrops throughout the southwestern United States and northern Mexico, and generally have a light pelage similar to the predominant rock color. However, in some regions these mice live on darker substrates (e.g. old lava flows), and in these patches have much darker pigmentation, presumably to avoid visual predators. Nachman et al. [2003] demonstrated that on one of these flows (Pinacate), much of the change in coloration is due to an allele at the MC1R locus. This dark allele differs from the light allele by four amino acid changes, and has a dominant or partially dominant effect depending on the measure of coat color. These areas of dark rock are separated from each other by a range of distances, with some being separated from any other by hundreds of kilometers of light colored substrate. The Pinacate allele is not present in a number of other populations of dark colored rock pocket mouse, suggesting these populations have adapted in parallel [Nachman et al., 2003, Hoekstra and Nachman, 2003]. However, Hoekstra et al. [2005] reasoned that, elsewhere in the range, multiple close dark outcrops may share a dark phenotype whose genetic basis has been spread by migration despite intervening light habitat. We apply our theory to this example, in particular investigating how the probability of parallel mutation depends on the distance between dark outcrops.

A key parameter above was the dispersal distance divided by the square root of strength of selection against the focal allele between patches, $\sigma/\sqrt{s_m}$. Hoekstra et al. [2004] studied the frequency of the dark MC1R allele and coat color phenotypes, at sites across the (dark) Pinacate lava flow and at two nearby sites in light-colored rock. On the lava flow the dark MC1R allele is at 86% frequency, while at a site with light substrate 12km to the west (Tule), the frequency is 3%. The dark allele was entirely absent from Christmas pass, a site with light substrate 7km north of Tule, and 3km further from the lava flow. In the other direction, the dark MC1R allele was at 34% at a site with light substrate 10km to the east of the flow (O'Neill). Note that such apparent asymmetry is expected, since as noted above the migration–selection equilibrium can be highly stochastic. These numbers give us a sense of a plausible range of the parameters. Assuming the dark allele is at 50% frequency at the edge of the lava flow, we can fit formula (3) to these frequencies (similar to Hoekstra et al. [2004]). Doing this for Tule we obtain $\sigma/\sqrt{s_m} \approx 3$km, and for O'Neill, $\sigma/\sqrt{s_m} \approx 30$km, giving us a range of cline widths. We also need estimate of s_m. Using $\sigma \approx 1$km [French et al., 1968, Allred and Beck, 1963], these cline widths imply that $s_m = 1/9$ and $1/900$ are reasonable values.

The mutational target size μ for the trait is unclear. While the Pinacate dark haplotype differs from the light haplotype at four amino acid residues, it is likely that not all of these changes are needed for a population to begin to adapt. Also there a number of genes besides MC1R at which adaptive changes affecting pigmentation have been identified in closely related species and more broadly across vertebrates [Hoekstra, 2006]. To span a range of plausible values, we use a low mutation rate of $\mu = 10^{-8}$, where the

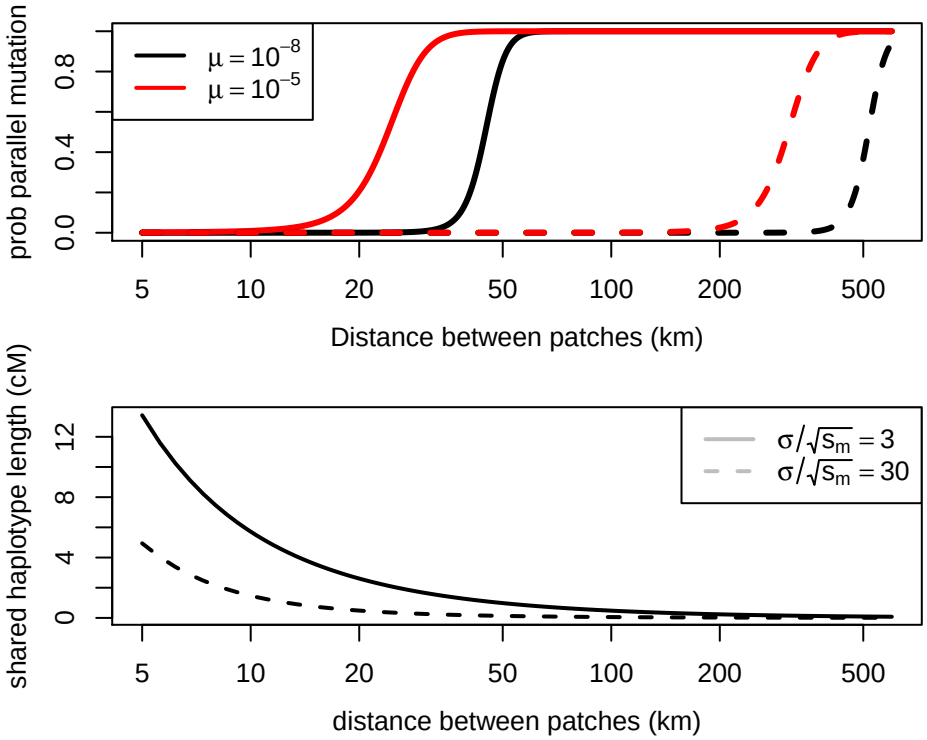

Figure 7. Top panel: The probability of parallel mutation as a function of the distance between environmental patches for two different cline widths and two different mutation rates (using equation (12)). The parameters were chosen to match the example of *Chaetodipus intermedius*; see text. Bottom panel: The initial genetic length of the haplotype shared between patches due to adaptation via migration as a function of the distance between environmental patches. As a rough rule of thumb 1cM is approximately 10^6 bases in a number of mammalian species.

adaptive mutation has to occur at a single base pair; and a high mutation rate $\mu = 10^{-5}$, equivalent to the mutation being able to arise at any of a thousand base pairs. Finally, we set $A = 100\text{km}^2$ (roughly the size of the Pinacate patch). In Figure 7 we show the dependence of the probability of parallel mutation on the distance between lava flow patches using these parameters, showing that parallel mutation should become likely over a scale of tens to a few hundred kilometers between patches.

Given the large selection coefficient associated with the dark allele on the dark substrate, we expect the initial haplotype associated with either a new mutation or migrant allele to be large. We can use equation (19) to calculate how long the founding haplotype shared between populations is expected to be, as shown in Figure 7. The initial length can be quite long between geographically close patches (tens of kilometers). However, for the wider cline width ($\sigma/\sqrt{s_m} = 30\text{km}$), adaptation by migration can still be likely for patches 100km apart, but the shared basis may be hard to detect, as the length of shared haplotype can be quite short.

Discussion

This paper is an investigation into the basic question: What is the spatial resolution of convergent local adaptation? In other words, over what spatial scale of environmental patchiness will the process of adaptation develop independent solutions to evolutionary problems? The answer to this depends most strongly on $\sigma/\sqrt{s_m}$, the dispersal distance divided by the square root of the strength of selection

against the allele between patches. It depends much more weakly on the selective benefit within the patches or (perhaps surprisingly) the population density, although these two factors will determine the time-scale over which adaptation will occur (and note that population density could affect the selection coefficients). This is in contrast with models of panmictic populations [Hermisson and Pennings, 2005, Messer and Petrov, 2013, Wilson et al., 2014] and geographically spread populations adapting to homogeneous selection pressures [Ralph and Coop, 2010], where the probability of multiple, independently arising adaptive alleles increases with the population size/density. However, in all of these models the dependence on the beneficial selection coefficient is absent or weak, due to the fact that selection both aids establishment of new alleles and the spread of existing alleles (but see Wilson et al. [2014] for the complications of varying population sizes).

We have also shown that while weaker selection against alleles will make sharing of adaptations between patches easier, it will also makes it harder to spot the sharing, since the lucky alleles that manage to colonize new patches move slower, and thus carry a shorter shared haplotype. This issue is amplified by the fact that the length of haplotype shared within patches decays over time, potentially making the identification of shared adaptations to old selection pressures difficult.

Perhaps the most useful rule-of-thumb quantities we found were the following. The effective rate of migration into an as-yet-unadapted patch from an already-adapted patch distance R away – the analogue of the mutational influx $\mu\rho A$ – is $2\rho s_m \exp(-R\sqrt{2s_m}/\sigma)$. Equivalently, the critical gap size between patches past which adaptation is likely independent is $R = (\sigma/\sqrt{2s_m})\log(2s_m/(A\mu))$. Finally, successfully transiting migrant lineages move at rate $\sigma\sqrt{2s_m}$, and so shared haplotype lengths between patches will be of order $R/(\sigma\sqrt{2s_m})$.

In developing the set of approximations in the paper we have ignored a number of complicating factors. We now briefly discuss these.

Standing variation We have focused on relative rates of adaptation, since we are thinking of applications in which we know that adaptation has occurred, and the question is whether or not adaptations in distinct patches have appeared independently or not. However, especially absolute if rates of adaptation by new mutation are low, any adaptation that does occur may have to make use of standing variation. The case of a panmictic population was studied by Hermisson and Pennings [2005], and we study the case of a continuous, spatial population in an upcoming paper [Ralph and Coop, 2014]. If parallelism in local adaptation of the sort we study here is due to standing variation rather than new mutation, then the dynamics of adaptation should not depend strongly on migration patterns (but the initial spatial distribution of standing variation may).

Dominance We have mostly ignored the issue of dominance by dealing with essentially haploid models, and appealing to the fact that the dynamics we study occur where the mutation is rare, and hence mostly present only in heterozygotes. Our results should hold as a good approximation to dominant and partially dominant alleles (with s_m the selection against heterozygotes). If, however, the mutation is recessive, then it is essentially neutral where rare, and so would encounter much less resistance to spreading between patches. The shape of the cline obtained is given by Haldane [1948]. This is counteracted, however, by the increased difficulty with which the mutation would establish once arriving in a new patch, if the beneficial effect is also recessive. As such it is not clear what our intuition should be about the contribution of recessive alleles to adaptation via migration. Further work is needed to put empirical observations of local adaptation by recessive alleles in a theoretical context. It is similarly unclear how the model should extend to a polygenic trait.

Decay of shared haplotypes To provide context for the results on shared haplotype length in section *Haplotypes Shared Between Patches* it is important to also understand the process by which haplotypes are whittled down within patches. The initial haplotype that sweeps within a patch will be dispersed over time by recombination. Likewise, the haplotype that is shared between patches coadapted by migration (equation 19) will also break down. However, a long time after the initial sweep, we may still expect to find individuals within the patch sharing longer haplotypes around the selected locus than with individuals elsewhere, since selection against migrants decreases mean coalescence times within the patch near the selected locus. The literature on clines (e.g. Barton [1979]) has important information, but more work is needed to provide robust estimates for these processes. Questions about the genomic length-scale of signals of sweeps shared by migration have also been addressed in discrete population

settings [Slatkin and Wiehe, 1998, Kim and Maruki, 2011], reviewed in Barton [2000]. This work has shown that the length of the shared swept haplotype is often significantly shorter than the sweep with each patch, resulting in a pattern of shoulders of elevated F_{ST} between adapted populations some distance away from the shared selected allele. It would be of interest to see how similar patterns can arise in a continuous population setting, as a way of uniting these results.

Long distance migration We have also ignored the possibility of very long distance migration, instead focusing on local dispersal (and using Gaussian and appealing to the central limit theorem). However, dispersal distributions can be very heavy tailed, with a small fraction of individuals moving very long distances indeed [Levin et al., 2003, Reynolds and Rhodes, 2009]. In addition, over long time-scales, very rare chance events (mice carried off by hurricanes and the like; Censky et al. [1998], Nathan et al. [2008]) could play a role in spreading migrant alleles if adaptation by other means is sufficiently unlikely. Such tail events could greatly increase the probability of shared adaptation above that predicted by our model. Furthermore, if adaptive alleles do move between distant patches via rare, long distance migration then they will be associated with a much longer shared haplotype than predicted by equation (19). As such, we view our results as a null model by which the contribution of long distribution migrants to adaptation could be empirically judged.

Other geographic models We have studied circular patches of habitat at long distances from each other. Real habitat geometry can be much more complex, e.g. with archipelagos of patches of varying sizes, or patches connected by long, skinny corridors, for instance. The work of Cantrell and Cosner [1991] comes closest to a general theory of balanced polymorphisms in such habitats. It is possible that our techniques could be applied in their much more general setting, as both are based, fundamentally, on branching process approximations. It is also interesting to think about the probability of convergent adaptation to continuously varying environments, e.g. replicated environmental clines.

Concluding thoughts The falling cost of population genomic sequencing means that we will soon have the opportunity to study the interplay of adaptation with geography and ecology across many populations within a species. Our work suggests that even quite geographically close populations may be forced to locally adapt by repeated, convergent, *de novo* mutation when migration is geographically limited and selective pressures are divergent. Thus, systems where populations have been repeatedly subject to strong local selection pressures may offer the opportunity to study highly replicated convergent adaptation within a similar genetic background [Stern, 2013]. Such empirical work will also strongly inform our understanding of the ability of gene flow to keep ecologically similar populations evolving in concert [Sexton et al., 2013]. Our results suggest that adaptation to shared environments is certainly no guarantee of a shared genetic basis to adaptation, suggesting that rapid adaptation to a shared environment could potentially drive speciation if the alleles that spread in each population fail to work well together [Kondrashov, 2003].

Acknowledgements

We would like to heartily thank Joachim Hermisson for a number of insightful suggestions, and catching the algebra error present in the first version of this paper on the bioRxiv, as well as the two anonymous reviewers. We would also like to acknowledge Michael Nachman, Steve Evans, and Yaniv Brandvain for useful discussions, as well as Gideon Bradburd, and Simon Aeschbacher and the rest of the Coop lab for helpful comments and discussion of the manuscript.

Materials & Methods

Simulation methods

First we briefly describe the simulations we used for illustration and validation (the `R` code used is provided as a supplement). We simulated forward-time dynamics of the number of alleles of each type in a rectangular grid (either one- or two-dimensional) of demes with fixed size N. Each generation, each individual independently chose to reproduce or not with a probability depending on her type and

location in the grid; locally beneficial alleles were more likely to reproduce. Each then produced a random, Poisson distributed number of offspring that each, independently, either remained in the same location with probability $1-m$ or migrated a random number of steps with probability m in a uniformly chosen cardinal direction; for most simulations $m = 0.2$ and the probability of migrating k steps was proportional to 2^{-k} for $1 \leq k \leq 5$. Once migrants were distributed, each deme was uniformly resampled back down to N individuals. (Although we described the simulation in terms of individuals, we kept track only of total numbers in an equivalent way.)

The probability of reproduction in each generation in simulations for type b alleles was 30%; this was then multiplied by $1+s$ to get the probability of reproduction for type B, where the value of s is either s_m or s_p depending on the individual's location. This determines the values of s_m and s_p reported in the figures, and do not depend on the basic rate of reproduction. However, to obtain values for s_p and s_m when comparing theory to simulation, we computed the rate of intrinsic growth, i.e. the s so that the numbers of B alleles when rare would change by e^{st} after t generations in the absence of migration. (The resulting values are close to the first notion of s, but give better agreement with theory, which uses the second definition.)

To sample lineages, we first simulated forwards in time the population dynamics, then sampled lineages moving back through time by, in each generation, moving each lineage to a new deme with probability proportional to the reverse migration probability weighted by the number of B alleles in that deme in the previous time step. If more than one lineage was found in a deme with n alleles of type B, then each lineage picked a label uniformly from $1 \ldots n$, and those picking the same label coalesced. Since reproduction is Poisson, this correctly samples from the distribution of lineages given the population dynamics.

Numerical calculation of the probability of establishment

When rare, under at least some demographic models, copies of a new mutant allele are approximately independent and experience a uniform selective benefit; and can therefore be treated as a branching process. Furthermore, whether or not a new, beneficial mutation establishes or is lost to demographic stochasticity is determined by this initial phase where it is rare. Fortunately, the probability that a branching process does not establish (i.e. dies out) can be found as a fixed point of the generating function of the process [Jagers, 1975]. Therefore, we calculated explicitly the generating function for a spatial branching process with nearest-neighbor migration on a one-dimensional lattice and a Poisson number of offspring with mean $1+s$, where s could vary by location, and iterated this forward to convergence to obtain $1-p(x)$, the probability a single mutation appearing at x would fail to establish. We considered two situations: where s is a step function, and where it has a linear transition. These solutions are shown, and parameters described, in figure 8.

Here (and at other parameter choices) we see that the probability of establishment $p(x)$ goes to the equilibrium value (approximately $p_e = 2s/\xi^2$) within the patch; the transition is fairly symmetrical about the edge of the patch, even if the edge of the patch is not sharp. The fit is equally good for other parameter values, even if migration can move further than one deme and offspring numbers are not Poisson (results not shown). This lends credence to our approximation that the integral of $p(x)$ over the entire range is close to p_e multiplied by the area of the patch.

The equilibrium frequency

For completeness, and clarity as to the scalings on the relevant parameters, here we provide a derivation of the differential equations referred to above equation (3), and establish the asymptotics given in that equation. One route to the "equilibrium frequency" of the allele outside the range where it is advantageous is as follows; see Slatkin [1973] or Barton [1987] (or Kolmogorov et al. [1991] or Fisher [1937] or Haldane [1948]) for other arguments in equivalent models, and see Etheridge [2000] and/or Dawson [1993] for a general framework for the stochastic processes below. Following is a description of a discrete, haploid model (which is nearly the one we simulate from). Suppose that the population is composed of a finite number of small demes of equal size N arranged in a regular grid, and that selection (for or against) the allele is given by the function $s(x)$, with x denoting the spatial location. Each individual at location x reproduces at random, exponentially distributed intervals, producing a random number of offspring with distribution given by X who then all migrates to a new location chosen randomly from the distribution

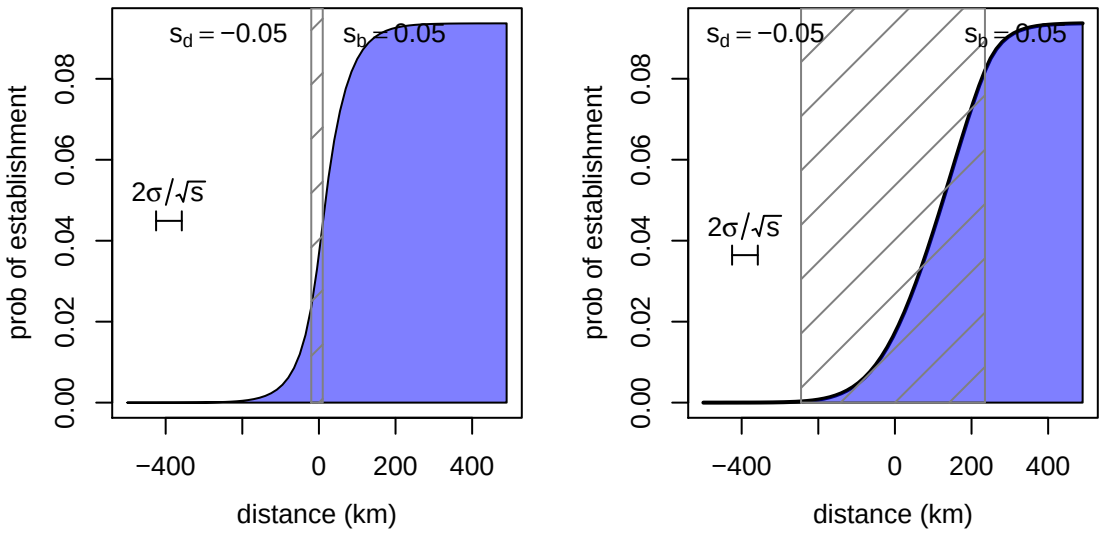

Figure 8. The probability of establishment of a new mutant allele as a function of distance from the edge of the region where it is beneficial, in an abrupt transition (left) and a gradual transition (right). The allele is deleterious to the left and beneficial to the right (with selection coefficients ∓ 0.05 respectively); and the selection coefficient interpolates linearly between these in the central hatched region. The number of offspring is Poisson. The underlying landscape has demes every 15km and a migration rate of 0.5 between neighboring demes; the probability was found by numerically solving the equation for the probability of establishment of a multi-type branching process. Note that the approximation used in the text is here $p_e \approx 2s_p/\xi^2 = 0.1$, agreeing with the asymptotic value on the right.

given by $x + R$, where they replace randomly chosen individuals. If $x + R$ is outside of the range, then they perish. Each individual's time until reproduction is exponentially distributed: the reproduction rate is 1 if it carries the original allele, or with rate $1 + s(x)$ if it carries the mutant allele. Suppose that the number of offspring X has mean μ; the variance of X will not enter into the formula (but assume X is well-behaved). Also suppose that the migration displacement R has mean zero and variance σ^2; if we are in more than one dimension, we mean that the components of the dispersal distance are uncorrelated and each have variance σ^2. We also assume that the distribution of R makes the associated random walk irreducible and aperiodic.

Let $\Phi_t^N(x)$ be the proportion of mutant alleles present at location x at time t, and $\Phi_t(x)$ the process obtained by taking $N \to \infty$ (which we assume exists). Denote by δ_x a single unit at location x, so that e.g. $\Phi_t^N + \delta_x/N$ is the configuration after a mutant allele has been added to location x. For $0 \le \phi \le 1$, we also denote by $\bar{X}_\phi$ the random number of mutant alleles added if X new offspring carrying mutant alleles replace randomly chosen individuals in a deme where the mutant allele is at frequency ϕ (i.e. hypergeometric with parameters $(X, 1-\phi)$); similarly, $\widetilde{X}_\phi$ is the number lost if the new offspring do not carry the allele (i.e. hypergeometric with parameters (X, ϕ)). (We like to think of Φ_t^N as a measure, but it does not hurt to think of Φ^N as a vector; we aren't providing the rigorous justification here.) Then

the above description implies that for any sufficiently nice function $f(\Phi)$ that

$$\begin{aligned}\frac{\partial}{\partial t}\mathbb{E}\left[f(\Phi_t^N)\right] &= N\sum_x \left\{\mathbb{E}\left[(1+s(x+R))\Phi_t^N(x+R)\left(f\left(\Phi_t^N + \frac{\bar{X}_{\Phi_t(x+R)}}{N}\delta_x\right) - f(\Phi_t^N)\right)\right]\right. \\ &\left. + \mathbb{E}\left[\left(1-\Phi_t^N(x+R)\right)\left(f\left(\Phi_t^N - \frac{\widetilde{X}_{\Phi_t(x+R)}}{N}\delta_x\right) - f(\Phi_t^N)\right)\right]\right\} \qquad (20) \\ &= \mu\sum_x \mathbb{E}\left[\left(\partial_{\phi(x)}f(\Phi_t)\right)\left\{\Phi_t(x+R) - \Phi_t(x) + s(x+R)\Phi_t(x+R)(1-\Phi_t(x))\right\}\right] + O\left(\frac{1}{N}\right). \qquad (21)\end{aligned}$$

In the final expectation, R and Φ are independent. This follows by taking first-order terms in $1/N$ in the Taylor series for f, and the fact that $\mathbb{E}[\bar{X}_\phi] = \phi\mu$ and $\mathbb{E}[\widetilde{X}_\phi] = (1-\phi)\mu$. We can see two things from this: First, since this is a first-order differential operator, the limiting stochastic process Φ obtained as $N \to \infty$ is in fact deterministic (check by applying to $f(\Phi) = \Phi(x)^2$ to find the variance). Second, if we want to rescale space as well to get the usual differential equation, we need to choose $\mathrm{Var}[R] = \sigma^2$ and $s(x)$ to be of the same, small, order; this is another way of seeing that $\sigma/\sqrt{s}$ is the relevant length scale (as noted by Slatkin [1973]). More concretely, suppose that the grid size is $\epsilon \to 0$, that $\mathrm{Var}[R] = (\sigma\epsilon)^2$, and that the strength of selection is $s(x)\epsilon$, suppose that $\Phi_t(x)$ is deterministic and twice differentiable, and let $\xi(t,x) = \Phi_{t/\epsilon}(x)$; then the previous equation with $f(\Phi) = \Phi(x)$ converges to the familiar form:

$$\partial_t\xi(t,x) = \mu\left(\frac{\sigma^2}{2}\sum_{k=1}^d \partial_{x_k}^2\xi(t,x) + s(x)\xi(t,x)(1-\xi(t,x))\right). \qquad (22)$$

Here we have taken first the population size $N \to \infty$ and then the grid size $\epsilon \to 0$; we could alternatively take both limits together, but not if ϵ goes to zero too much faster than N grows. One reason for this is that at finite N, the process Φ_t is an irreducible, aperiodic finite-state Markov chain with absorbing states at 0 and 1; therefore, the inevitable outcome is extinction of one type or another, which is not the regime we want to study.

In one dimension, we are done (and discuss exact solutions in Appendix A); in higher dimensions, we are more interested in the mean frequency at a given distance r from a patch. If we take a radially symmetric patch centered at the origin (so s only depends on r), and let $\xi(t,r)$ denote the mean occupation frequency at distance r, then the polar form of the Laplacian in d dimensions gives us that (22) is

$$\partial_t\xi(t,r) = \mu\left(\frac{\sigma^2}{2}\partial_r^2\xi(t,r) + \sigma^2\frac{d-1}{2r}\partial_r\xi(t,r) + s(r)\xi(t,r)(1-\xi(t,r))\right). \qquad (23)$$

Asymptotic solution for the equilibrium frequency

If $q(r)$ is a radially symmetric equilibrium frequency and $\xi(t,x) = q(|x|)$, then $\partial_t\xi(t,x) = 0$. So, using (23), a radially symmetric solution with $s(r) = -s < 0$ for all $r > r_0$ solves, for $r_0 < r < \infty$,

$$\partial_r^2 q(r) + \frac{d-1}{r}\partial_r q(r) - \frac{2s}{\sigma^2}q(r)(1-q(r)) = 0 \qquad (24)$$

$$\lim_{r\to\infty} q(r) = \lim_{r\to\infty}\partial_r q(r) = 0 \qquad 0 < q(r) < 1 \qquad (25)$$

Since $q(r) \to 0$ as $r \to \infty$, so $q(r)(1-q(r)) \approx q(r)$, it can be shown that the true equilibrium frequency q is close, for large r, to the solution to

$$\partial_r^2 u(r) + \frac{d-1}{r}\partial_r u(r) - \frac{2s}{\sigma^2}u(r) = 0 \qquad (26)$$

$$\lim_{r\to\infty} u(r) = \lim_{r\to\infty}\partial_r u(r) = 0 \qquad . \qquad (27)$$

This has general solution given by a modified Bessel function: using Gradshteyn and Ryzhik [2007] 8.494.9, the general solution is

$$u(r) = C'(r-r_1)^{(2-d)/2}K_{(2-d)/2}\left((r-r_1)\sqrt{2s}/\sigma\right), \qquad (28)$$

where C' and r_1 are chosen to match boundary conditions. Asymptotics of Bessel functions (Gradshteyn and Ryzhik [2007], 8.451.6) then imply that

$$q(r) \approx Cr^{(1-d)/2} \exp\left(-r\sqrt{2s}/\sigma\right) + O(1/r), \tag{29}$$

where C is a different constant.

Hitting and occupation

Here we make a more precise argument to back up expression (11) for the migration rate. The argument made above in section *Heuristics* applies to general migration mechanisms, since it relies only on a decomposition of the migrant families upon hitting the new patch; but it is also imprecise in a number of subtle ways that are difficult to formalize. Here we take a somewhat different tack, supposing that it suffices to model the spatial movement of a migrant family by following only the motion of the "trunk" (i.e. the red line in figure 3), and supposing this motion is Brownian, with variance σ^2 per generation. (Recall σ is the dispersal distance.) We then use facts about Brownian motion and branching processes, to compute more precise versions of equations (4) and (5).

We are approximating the dynamics of the focal allele in the region further away than r_0 from the patch as the sum of independent migrant families, each of whose dynamics are given by a spatial branching process (as depicted in Figure 3). Call $B(r_0)$ the region closer than r_0 to the patch, and $\partial B(r_0)$ its boundary. Denote by $\gamma(x)$ the mean rate of outflux of migrant families from a point $x \in \partial B(r_0)$, i.e. the time-averaged proportion of individuals near a point x in $\partial B(r_0)$ that are the founders of new migrant families. Now note that expressions (4) and (5) are a simple product of the "outflux of families", when in fact they should be an integral of $\gamma(x)$ over possible locations. However, examining the constituent terms, it will turn out that equation (5) is well-approximated by a constant multiple of (4).

First consider equation (4) for the equilibrium frequency. Suppose that Z is one such spatial branching process as above in section *The genealogy of migrant families*, started at time 0 with a single individual at x, and denote by $Z_t(A)$ is the number of descendant individuals alive at time t living in the spatial region A. The *mean occupation density* of Z at location y can be thought of informally as the expected total number of offspring of a family beginning at x that ever live near y. It is defined to be the density of $\mathbb{E}[\int_0^t Z_t(\cdot)dt]$, i.e. the function $u(x, y)$ that, integrated over a set of locations S, gives the expected total number of individuals ever living in S:

$$\mathbb{E}\left[\int_0^\infty Z_t(S)dt\right] = \int_S u(x, y)dy. \tag{30}$$

Since reproduction and spatial motion are independent, $u(x, y)$ is the sum across all generations of the expected number of individuals alive in that generation, multiplied by the probability density that spatial motion takes a lineage from x to y, i.e.

$$u(x, y) = \int_0^\infty e^{-s_m t} p_t(x, y)dt. \tag{31}$$

The equilibrium frequency as decomposed in expression (4) is the sum of mean occupation densities of a constant outflux of branching processes from $\partial B(r_0)$:

$$q(y) = \int_{\partial B(r_0)} \gamma(x)u(x, y)dx. \tag{32}$$

This form we can now compare to the expression (5) for the outflux of successful migrants. Consider the probability that a migrant family beginning with a single individual at x will ever establish in the new patch. We assume this new patch is circular, has relatively small area A and is centered at position y. It would be possible to analyze this probability directly, as in Barton [1987]; but here we take a simpler route, approximating this by the chance that the trunk hits the new patch, multiplied by the chance that at least one member of the family escapes demographic stochasticity and successfully establishes in the new patch. Write $h_A(x, y)$ for the probability that the Brownian trunk hits the new patch centered at y before the family dies out, and $f(A)$ for the chance that the family manages to establish in the new

patch, given that it successfully arrives. As for $q(y)$ above, expression (5) is properly an integral against the outflux $\gamma(x)$:

$$\lambda_{\text{mig}}(y) = \rho \int_{\partial B(r_0)} \gamma(x) h_A(x,y) f(A) dx. \tag{33}$$

Recall that ρ appears here but not in (32) because both $q(x)$ and $\gamma(x)$ are local *proportions* of B alleles.

Intuitively, $h_A(x,y)$ counts each family once if it hits the patch, while the occupation density $u(x,y)$ counts the total amount of time that the family spends in the patch. Since the expected total time spent in the patch does not depend on the distance at which the family started if we know that it survived to hit the patch, it is intuitively reasonable that $u(x,y)$ and $h_A(x,y)$ are simply related. Following this intuition, and using $(1-k_e(t)) \approx e^{-s_m t}/\mathbb{E}[K]$, we show below that

$$h_A(x,y) \approx \frac{u(x,y)}{\mathbb{E}[K] g(A)}, \tag{34}$$

where $g(A)$ is the total expected occupation time of the patch given the family arrives. Recall that $\mathbb{E}[K]$ is the expected family size after a large number of generations, given survival. Since the integrands in expressions (32) and (33) only differ by a factor that (at least asymptotically) does not depend on the distance between x and y, we can obtain the migration rate by multiplying the equilibrium frequency by this factor.

The result will not depend on A, because although $f(A)$ and $g(A)$ in principle depend on the patch size (and geometry), the dependence is very weak. For instance, the width of a circular patch only very weakly affects the chance of establishment of a new migrant that appears on its edge, as long as the patch is large enough.

Consider a single migrant family Z beginning at x, and write S for the new patch, which has area A and is centered at y. Let B_t be a Brownian motion with variance σ^2, and $\tau_\dagger$ an independent Exponential(s_m) time. The mean occupation time of Z spent in a region S is, since the marginal distribution of a single lineage is Brownian, letting $u(x,S) = \int_S u(x,y)dy$,

$$\begin{aligned}
u(x,S) &= \mathbb{E}\left[\int_0^\infty Z_t(S) dt\right] && (35)\\
&= \int_0^\infty \mathbb{E}[Z_t] p_t(x,S) dt && (36)\\
&= \int_0^\infty e^{-s_m t} p_t(x,S) dt && (37)\\
&= \int_0^\infty \mathbb{P}^x\{\tau_\dagger > t \;\&\; B_t \in S\} dt && (38)\\
&= \mathbb{E}^x\left[\int_0^{\tau_\dagger} \mathbf{1}_S(B_t) dt\right], && (39)
\end{aligned}$$

where $\mathbb{P}^x$ gives probabilities for Brownian motion begun at x (i.e. $B_0 = x$), and likewise $\mathbb{E}^x$. If we define τ_S to be the hitting time of S by the Brownian motion B, and $\mu_S(x)$ to be the hitting distribution of ∂S by B_{τ_S} conditioned on $\tau_S < \tau_\dagger$, by the strong Markov property this is equal to

$$u(x,S) = \mathbb{P}^x\{\tau_S < \tau_\dagger\} \mathbb{E}^{\mu_S(x)}\left[\int_0^{\tau_\dagger} \mathbf{1}_S(B_t) dt\right], \tag{40}$$

where now $\mathbb{E}^\mu$ denotes expectations for Brownian motion for which the distribution of B_0 is μ, provided S is not pathological. If S is circular with area A, the latter expectation will not depend on x; call this $g(A)$.

Now consider the chance that Z ever hits the new patch, which we call $h(x,S)$. (Before, this was $h_A(x,y)$, here modified to parallel $u(x,S)$.) If we make the approximation that Z hits the new patch only if the trunk of Z does, and recall that $1-k_e(t)$ is the chance that Z survives for t generations, then

this is equal to

$$h(x,S) \approx \int_0^\infty (1-k_e(t))\mathbb{P}^x\{\tau_S \in dt\} \tag{41}$$

$$= \int_0^\infty e^{s_m t}(1-k_e(t))e^{-s_m t}\mathbb{P}^x\{\tau_S \in dt\} \tag{42}$$

$$= \int_0^\infty e^{s_m t}(1-k_e(t))\mathbb{P}^x\{\tau_S \in dt \,\&\, t < \tau_\dagger\} \tag{43}$$

$$\approx \frac{1}{\mathbb{E}[K]} \int_0^\infty \mathbb{P}^x\{\tau_S \in dt \,\&\, t < \tau_\dagger\} \tag{44}$$

$$= \frac{1}{\mathbb{E}[K]}\mathbb{P}^x[\tau_S < \tau_\dagger]. \tag{45}$$

Therefore, we have that

$$h(x,S)\mathbb{E}[K] \approx \mathbb{P}^x[\tau_S < \tau_\dagger] = u(x,S)/g(A). \tag{46}$$

If we let $q(S) = \int_S q(y)dy$ be the total equilibrium occupation of S, then by equations (32) and (33),

$$q(S) = \int_{\partial B(r_0)} \gamma(x)u(x,S)dx \tag{47}$$

$$\approx g(A)\mathbb{E}[K] \int_{\partial B(r_0)} \gamma(x)h(x,S)dx \tag{48}$$

$$= \frac{g(A)\mathbb{E}[K]}{\rho f(A)}\lambda_{\text{mig}}(y). \tag{49}$$

Once we show that $g(A) \approx 1/(2s_m)$, we will have arrived at the result, equation (10).

The function $g(A)$ is the expected amount of time that a Brownian motion begun on the edge of a disk of area A is expected to spend inside the disk before $\tau_\dagger$. This is integral of the Green function for the Bessel process of the appropriate order, so using Borodin and Salminen [2002], and letting w be the width of the patch, in $d = 1$,

$$\begin{aligned} g(A) &= \int_0^{2w/\sigma} \frac{e^{-y\sqrt{2s_m}/\sigma}}{\sqrt{2s_m}}dy \\ &= (1-e^{-2w\sqrt{2s_m}/\sigma})/(2s_m) \end{aligned} \tag{50}$$

and in $d = 2$, by Gradshteyn and Ryzhik [2007] 5.56.2,

$$g(A) = \int_0^{2w/\sigma} 2yK_0(y\sqrt{2s_m})dy \tag{51}$$

$$= \frac{1}{s_m}\int_0^{2w\sqrt{2s_m}/\sigma} yK_0(y)dy \tag{52}$$

$$= \frac{1}{2s_m}\left(1 - \frac{2w\sqrt{2s_m}}{\sigma}K_1(2w\sqrt{2s_m}/\sigma)\right), \tag{53}$$

where K_0 and K_1 are modified Bessel functions of the second kind. In either case, $g(A) \approx 1/(2s_m)$, which is the approximation we use in the main text.

The outflux of families

It is of interest to have an expression for the "outflux of families", $\gamma(r_0)$, defined to be the mean proportion of B alleles at distance r_0 whose lineage traces back to the original patch without venturing further away than r_0. By expression (32),

$$\gamma(r_0) = \frac{q(y)}{\int_{B(r_0)} u(x,y)dx}$$

for any y. Take $\hat{x} = (1, 0)$ and $y = r\hat{x}$. Then

$$\frac{u(x,y)}{q(y)} \simeq C' \frac{\|y-x\|^{-(d-1)/2} e^{-\|y-x\|\sqrt{2s_m}/\sigma}}{\|y\|^{-(d-1)/2} e^{-\|y\|\sqrt{2s_m}/\sigma}} \tag{54}$$

$$= C' \|\hat{x} - x/r\|^{-(d-1)/2} e^{-r\sqrt{2s_m}/\sigma(\|\hat{x}-x/r\|-1)}. \tag{55}$$

Now let $x = (r_0 \cos\theta, r_0 \sin\theta)$, so that the expression in the exponent is to, first order in $1/r$,

$$r(\|\hat{x} - x/r\| - 1) = r\left(\sqrt{\left(1 - \frac{r_0}{r}\cos\theta\right)^2 + \frac{r_0^2}{r^2}\sin^2\theta} - 1\right) \tag{56}$$

$$= -2r_0 \cos\theta + O(1/r). \tag{57}$$

This implies that, still with r, r_0, and θ defined as above,

$$\frac{u(x,y)}{q(y)} \longrightarrow_{r\to\infty} C' e^{-2r_0 \frac{\sqrt{2s_m}}{\sigma}\cos\theta}. \tag{58}$$

In one dimension we are done. In two dimensions, by bounded convergence and Gradshteyn and Ryzhik [2007] (3.339),

$$\int_{B(r_0)} \frac{u(x,y)}{q(y)} dx \to \gamma(r_0) \quad \text{as } y \to \infty \tag{59}$$

$$= C' \int_0^{2\pi} e^{-2r_0 \frac{\sqrt{2s_m}}{\sigma}\cos\theta} d\theta \tag{60}$$

$$= 2\pi C' I_0(-2r_0\sqrt{2s_m}/\sigma), \tag{61}$$

where I_0 is a modified Bessel function of order zero of the first kind.

References

D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985. URL `http://www.springerlink.com/content/c31v17440871210x/fulltext.pdf`.

D. M. Allred and D. E. Beck. Range of movement and dispersal of some rodents at the Nevada atomic test site. *Journal of Mammalogy*, 44(2):pp. 190–200, 1963. ISSN 00222372. URL `http://www.jstor.org/stable/1377452`.

J. Arendt and D. Reznick. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol*, 23(1):26–32, Jan. 2008. doi: 10.1016/j.tree.2007.09.011.

N. Barton. Gene flow past a cline. *Heredity*, 43(3):333–339, Dec. 1979. ISSN 0018-067X. URL `http://dx.doi.org/10.1038/hdy.1979.86`.

N. H. Barton. The probability of establishment of an advantageous mutant in a subdivided population. *Genetics Research*, 50(1):35–40, 1987. doi: 10.1017/S0016672300023314. URL `http://dx.doi.org/10.1017/S0016672300023314`.

N. H. Barton. Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci*, 355(1403):1553–1562, Nov. 2000. doi: 10.1098/rstb.2000.0716. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1692896/`.

N. H. Barton, A. M. Etheridge, J. Kelleher, and A. Véber. Genetic hitchhiking in spatially extended populations. *Theoretical Population Biology*, (0):–, 2013. ISSN 0040-5809. doi: 10.1016/j.tpb.2012.12.001. URL `http://www.sciencedirect.com/science/article/pii/S0040580912001359`.

S. B. Benson. *Concealing coloration among some desert rodents of the southwestern United States*. Number v. 40 in University of California publications in zoology. University of California Press, 1933. URL `http://books.google.com/books?id=Lis-AQAAIAAJ`.

A. N. Borodin and P. Salminen. *Handbook of Brownian motion: facts and formulae*. Springer, 2002.

R. S. Cantrell and C. Cosner. Diffusive logistic equations with indefinite weights: population models in disrupted environments. *Proceedings of the Royal Society of Edinburgh, Section: A Mathematics*, 112 (3-4):293–318, 0 1989. ISSN 1473-7124. doi: 10.1017/S030821050001876X. URL `http://journals.cambridge.org/article_S030821050001876X`.

R. S. Cantrell and C. Cosner. Diffusive logistic equations with indefinite weights: population models in disrupted environments. II. *SIAM J. Math. Anal.*, 22(4):1043–1064, 1991. ISSN 0036-1410. doi: 10.1137/0522068. URL `http://dx.doi.org/10.1137/0522068`.

E. J. Censky, K. Hodge, and J. Dudley. Over-water dispersal of lizards due to hurricanes. *Nature*, 395 (6702):556–556, Oct. 1998. ISSN 00280836. URL `http://dx.doi.org/10.1038/26886`.

B. Chauvin, A. Rouault, and A. Wakolbinger. Growing conditioned trees. *Stochastic Process. Appl.*, 39 (1):117–130, 1991. ISSN 0304-4149. doi: 10.1016/0304-4149(91)90036-C. URL `http://dx.doi.org/10.1016/0304-4149(91)90036-C`.

J. L. Cherry and J. Wakeley. A diffusion approximation for selection and drift in a subdivided population. *Genetics*, 163(1):421–428, Jan. 2003. URL `http://www.ncbi.nlm.nih.gov/pubmed/12586727`.

C. Conley. An application of Wazewski's method to a non-linear boundary value problem which arises in population genetics. *Journal of Mathematical Biology*, 2(3):241–249, 1975. ISSN 0303-6812. doi: 10.1007/BF00277153. URL `http://dx.doi.org/10.1007/BF00277153`.

D. A. Dawson. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Math.*, pages 1–260. Springer, Berlin, 1993.

L. R. Dice. Ecologic and genetic variability within species of *Peromyscus*. *The American Naturalist*, 74 (752):pp. 212–221, 1940. ISSN 00030147. URL `http://www.jstor.org/stable/2457573`.

L. R. Dice and P. M. Blossom. *Studies of Mammalian Ecology in Southwestern North America With Special Attention to the Colors of Desert Mammals*. Carnegie Institution, 1937.

P. Donnelly and P. Joyce. Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Process. Appl.*, 31(1):89–103, 1989. ISSN 0304-4149. URL `http://www.sciencedirect.com/science/article/pii/030441498990104X`.

A. M. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2000. ISBN 0-8218-2706-5.

R. Fisher. The wave of advance of advantageous genes. *Ann. Eugenics*, 7:353–369, 1937. URL `http://digital.library.adelaide.edu.au/coll/special/fisher/152.pdf`.

R. A. Fisher. *The genetical theory of natural selection*. Oxford University Press, Oxford, 1930. ISBN 0-19-850440-3. URL `http://www.archive.org/details/geneticaltheoryo031631mbp`.

R. A. Fisher. Gene frequencies in a cline determined by selection and diffusion. *Biometrics*, 6(4):pp. 353–361, 1950. ISSN 0006341X. URL `http://www.jstor.org/stable/3001780`.

N. R. French, T. Y. Tagami, and P. Hayden. Dispersal in a population of desert rodents. *Journal of Mammalogy*, 49(2):pp. 272–280, 1968. ISSN 00222372. URL `http://www.jstor.org/stable/1377984`.

J. Geiger. Elementary new proofs of classical limit theorems for Galton-Watson processes. *Journal of Applied Probability*, 36(2):pp. 301–309, 1999. URL `http://www.jstor.org/stable/3215457`.

I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition, 2007. ISBN 978-0-12-373637-6; 0-12-373637-4. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger.

J. B. S. Haldane. A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(07):838–844, 7 1927. ISSN 1469-8064. doi: 10.1017/S0305004100015644. URL http://journals.cambridge.org/article_S0305004100015644.

J. B. S. Haldane. The theory of a cline. *J Genet*, 48(3):277–284, Jan. 1948. URL http://www.ncbi.nlm.nih.gov/pubmed/18905075.

O. Hallatschek and D. S. Fisher. Acceleration of evolutionary spread by long-range dispersal. *Proc Natl Acad Sci U S A*, 111(46):4911–4919, Nov. 2014. doi: 10.1073/pnas.1404663111. URL http://www.ncbi.nlm.nih.gov/pubmed/25368183.

J. Hermisson and P. S. Pennings. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352, Apr. 2005. doi: 10.1534/genetics.104.036947. URL http://www.ncbi.nlm.nih.gov/pubmed/15716498.

H. E. Hoekstra. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity (Edinb)*, 97(3):222–234, Sep 2006. URL http://www.ncbi.nlm.nih.gov/pubmed/16823403.

H. E. Hoekstra and M. W. Nachman. Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol. Ecol.*, 12:1185–1194, May 2003. URL http://www3.interscience.wiley.com/journal/118890526/abstract.

H. E. Hoekstra, K. E. Drumm, and M. W. Nachman. Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, 58(6):1329–1341, June 2004. URL http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2004.tb01711.x/abstract.

H. E. Hoekstra, J. G. Krenz, and M. W. Nachman. Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity*, 94(2):217–228, Nov. 2005. ISSN 0018067X. URL http://dx.doi.org/10.1038/sj.hdy.6800600.

P. Jagers. *Branching processes with biological applications*. Wiley Series in Probability and Statistics: Applied Probability and Statistics Section Series. Wiley, 1975. ISBN 9780471436522.

D. W. Kaufman. Adaptive coloration in *Peromyscus polionotus*: Experimental selection by owls. *Journal of Mammalogy*, 55(2):pp. 271–283, 1974. ISSN 00222372. URL http://www.jstor.org/stable/1378997.

Y. Kim and T. Maruki. Hitchhiking effect of a beneficial mutation spreading in a subdivided population. *Genetics*, 189(1):213–226, Sept. 2011. doi: 10.1534/genetics.111.130203. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3176130/.

E. P. Kingsley, M. Manceau, C. D. Wiley, and H. E. Hoekstra. Melanism in *Peromyscus* is caused by independent mutations in agouti. *PLoS ONE*, 4:e6435, 2009. URL http://dx.doi.org/10.1371%2Fjournal.pone.0006435.

A. Kolmogorov, I. Petrovskii, and N. Piscunov. A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. In *Selected Works of A.N. Kolmogorov: Mathematics and mechanics*, volume 25 of *Mathematics and its Applications (Soviet Series)*, pages 1–25. Kluwer Academic Publishers Group, Dordrecht, 1991.

A. S. Kondrashov. Accumulation of Dobzhansky–Muller incompatibilities within a spatially structured population. *Evolution*, 57(1):151–153, Jan 2003.

A. R. Kruckeberg. Intraspecific variability in the response of certain native plant species to serpentine soil. *American Journal of Botany*, 38(6):pp. 408–419, 1951. URL http://www.jstor.org/stable/2438248.

T. Lenormand. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4):183 – 189, 2002. ISSN 0169-5347. doi: DOI:10.1016/S0169-5347(02)02497-7. URL http://www.sciencedirect.com/science/article/pii/S0169534702024977.

S. A. Levin, H. C. Muller-Landau, R. Nathan, and J. Chave. The ecology and evolution of seed dispersal: a theoretical perspective. *Annu. Rev. Ecol. Evol. Syst.*, 34:575–604, 2003. URL http://arjournals.annualreviews.org/doi/pdf/10.1146/annurev.ecolsys.34.011802.132428.

Y. Lou and E. Yanagida. Minimization of the principal eigenvalue for an elliptic boundary value problem with indefinite weight, and applications to population dynamics. *Japan J. Indust. Appl. Math.*, 23(3): 275–292, 2006. ISSN 0916-7005. URL http://projecteuclid.org/euclid.jjiam/1197390801.

M. R. Macnair. Why the evolution of resistance to anthropogenic toxins normally involves major gene changes: the limits to natural selection. *Genetica*, 84(3):213–219, 1991. ISSN 0016-6707. doi: 10.1007/BF00127250. URL http://dx.doi.org/10.1007/BF00127250.

A. Martin and V. Orgogozo. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*, 67(5):1235–1250, 2013. ISSN 1558-5646. doi: 10.1111/evo.12081. URL http://dx.doi.org/10.1111/evo.12081.

T. Maruyama. On the fixation probability of mutant genes in a subdivided population. *Genetics Research*, 15(02):221–225, 1970. doi: 10.1017/S0016672300001543. URL http://dx.doi.org/10.1017/S0016672300001543.

J. Maynard Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genet Res*, 23(1):23–35, Feb. 1974. URL http://www.ncbi.nlm.nih.gov/pubmed/4407212.

P. W. Messer and D. A. Petrov. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol. (Amst.)*, 28(11):659–669, Nov 2013.

M. W. Nachman, H. E. Hoekstra, and S. L. D'Agostino. The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci U S A*, 100(9):5268–5273, Apr. 2003. doi: 10.1073/pnas.0431157100. URL http://www.ncbi.nlm.nih.gov/pubmed/12704245.

T. Nagylaki. Conditions for the existence of clines. *Genetics*, 80(3):595–615, 1975. URL http://www.genetics.org/content/80/3/595.abstract.

R. Nathan, F. M. Schurr, O. Spiegel, O. Steinitz, A. Trakhtenbrot, and A. Tsoar. Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution*, 23(11):638 – 647, 2008. ISSN 0169-5347. doi: http://dx.doi.org/10.1016/j.tree.2008.08.003. URL http://www.sciencedirect.com/science/article/pii/S0169534708002723.

NEQwiki. NEQwiki, the nonlinear equations encyclopedia, 2013. URL http://www.primat.mephi.ru/wiki/ow.asp?Korteweg-de_Vries_equation. Accessed December 17, 2014.

H. A. Orr and A. J. Betancourt. Haldane's sieve and adaptation from the standing genetic variation. *Genetics*, 157(2):875–884, Feb. 2001. URL http://www.ncbi.nlm.nih.gov/pubmed/11157004.

P. S. Pennings and J. Hermisson. Soft sweeps II - molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol*, page msj117, 2006. doi: 10.1093/molbev/msj117. URL http://mbe.oxfordjournals.org/cgi/content/abstract/msj117v1.

J. Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields*, 102(2):145–158, 1995. ISSN 0178-8051. doi: 10.1007/BF01213386. URL http://dx.doi.org/10.1007/BF01213386.

E. Pollak. On the survival of a gene in a subdivided population. *Journal of Applied Probability*, 3(1): 142–155, 1966. ISSN 00219002. URL http://www.jstor.org/stable/3212043.

P. Ralph and G. Coop. Parallel adaptation: One or many waves of advance of an advantageous allele? *Genetics*, 186(2):647–668, 2010. doi: 10.1534/genetics.110.119594. URL http://www.genetics.org/cgi/content/abstract/186/2/647.

P. L. Ralph and G. Coop. The role of standing variation in geographic convergent adaptation. *bioRxiv*, 2014. doi: 10.1101/009803.

A. M. Reynolds and C. J. Rhodes. The Lévy flight paradigm: random search patterns and mechanisms. *Ecology*, 90(4):877–887, Apr. 2009. URL http://www.esajournals.org/doi/abs/10.1890/08-0153.1.

E. B. Rosenblum, H. Römpler, T. Schöneberg, and H. E. Hoekstra. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences*, 107(5):2113–2117, 2010. doi: 10.1073/pnas.0911042107. URL http://www.pnas.org/content/107/5/2113.abstract.

H. Schat, R. Vooijs, and E. Kuiper. Identical major gene loci for heavy metal tolerances that have independently evolved in different local populations and subspecies of *Silene vulgaris*. *Evolution*, 50 (5):pp. 1888–1895, 1996. URL http://www.jstor.org/stable/2410747.

J. P. Sexton, S. B. Hangartner, and A. A. Hoffmann. Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, 2013. ISSN 1558-5646. doi: 10.1111/evo.12258. URL http://dx.doi.org/10.1111/evo.12258.

M. Slatkin. Gene flow and selection in a cline. *Genetics*, 75(4):733–756, 1973. URL http://www.genetics.org/cgi/content/abstract/75/4/733.

M. Slatkin and T. Wiehe. Genetic hitch-hiking in a subdivided population. *Genet Res*, 71(2):155–160, Apr. 1998. URL http://www.ncbi.nlm.nih.gov/pubmed/9717437.

C. C. Steiner, H. Rompler, L. M. Boettger, T. Schoneberg, and H. E. Hoekstra. The genetic basis of phenotypic convergence in beach mice: Similar pigment patterns but different genes. *Mol Biol Evol*, 26 (1):35–45, 2009. doi: 10.1093/molbev/msn218. URL http://mbe.oxfordjournals.org/cgi/content/abstract/26/1/35.

D. L. Stern. The genetic causes of convergent evolution. *Nat Rev Genet*, 14(11):751–764, Nov. 2013. doi: 10.1038/nrg3483. URL http://www.ncbi.nlm.nih.gov/pubmed/24105273.

T. L. Turner, E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.*, 42(3):260–3, Jan. 2010. URL http://www.ncbi.nlm.nih.gov/pubmed/20101244.

B. A. Wilson, D. A. Petrov, and P. W. Messer. Soft selective sweeps in complex demographic scenarios. *Genetics*, 198(2):669–684, Oct. 2014. doi: 10.1534/genetics.114.165571. URL http://www.ncbi.nlm.nih.gov/pubmed/25060100.

Y. Zhen, M. L. Aardema, E. M. Medina, M. Schumer, and P. Andolfatto. Parallel molecular evolution in an herbivore community. *Science*, 337(6102):1634–1637, 2012. doi: 10.1126/science.1226630. URL http://www.sciencemag.org/content/337/6102/1634.abstract.

A The exact solution to the equilibrium frequency

Here we describe how to solve (24) exactly in one dimension, i.e. the Fisher-KPP equation with piecewise constant selection. Let $p(t,x)$ be the proportion of the allele under spatially varying selection $s(x)$ in one dimension, so that

$$\partial_t p(t,x) = \frac{\mu}{2}\sigma^2 \partial_x^2 p(t,x) + \mu s(x) p(t,x)(1-p(t,x)).$$

The stable distribution $\phi(x) = \lim_{t\to\infty} p(t,x)$ then solves

$$\partial_x^2 \phi(x) = -2s(x)\phi(x)(1-\phi(x))/\sigma^2, \tag{62}$$

with appropriate boundary conditions. First rescale space by $\sigma/\sqrt{2}$ so the $\sigma^2/2$ term disappears. If we now assume that $s(x)$ is piecewise constant, $s(x) = s_i$ for $x \in [x_i, x_{i+1})$, with $x_0 = -\infty$ and $x_{n+2} = \infty$,

then the equation is integrable: if we multiply through by $2\partial_x\phi(x)$ and integrate, then we get that

$$(\partial_x\phi(x))^2 = -\int^x 2s(x)\phi(x)(1-\phi(x))\partial_x\phi(x)dx \tag{63}$$

$$= -s_i\phi(x)^2\left(1-\frac{2}{3}\phi(x)\right) - K_i \quad \text{for } x \in [x_i, x_{i+1}), \tag{64}$$

if we define

$$K_i = -s_i\phi(x_i)^2\left(1-\frac{2}{3}\phi(x_i)\right) - (\partial_x\phi(x_i))^2. \tag{65}$$

For ease of reference, we define

$$V_i(\phi) = s_i\phi^2\left(1-\frac{2}{3}\phi\right) + K_i.$$

Note that $V_i'(0) = V_i'(1) = 0$, that $V_i(0) = K_i$ and $V_i(1) = K_i + s_i/3$. We will always have that $V(\phi) \leq 0$. (We have then that $\partial_x^2\phi = -\partial_\phi V(\phi)$, the equation of motion of a particle in potential V. Also note that this implies "conservation of energy", i.e. $(\partial_x\phi(x))^2 + V(\phi(x))$ is constant.) Rearranging, we get that $dx = d\phi/\sqrt{-V(\phi)}$, so where $\phi(x)$ is monotone, the inverse is

$$x(\phi) = x(\phi^*) \pm \int_{\phi^*}^{\phi} \frac{d\psi}{\sqrt{-V(\psi)}}.$$

For each i then define the elliptic function

$$F_i(\phi) = \int_{\phi_i^*}^{\phi} \frac{d\psi}{\sqrt{-V_i(\psi)}}, \tag{66}$$

where take the positive branch of the square root, and ϕ_i^* will be chosen later. Then we have that $x(\phi) - x(\phi_0) = \pm(F(\phi) - F(\phi_0))$, or for an appropriate x_0,

$$\phi(x) = F^{-1}\left(F(\phi(x_0)) \pm (x - x_0)\right).$$

We clearly want $\lim_{x\to\infty}\partial_x\phi(\pm x) = 0$, and $\lim_{x\to\infty}\phi(\pm x)$ to be zero or one depending on the sign of s_0 and s_{n+1}. Since $(\partial_x\phi(x))^2 = -V(x)$, this implies that if $s_0 < 0$, then $K_0 = 0$, while if $s_0 > 0$ then $K_0 = s_0/6$; and likewise for K_{n+1}.

We also require that $\phi(x)$ and $\phi'(x)$ are continuous. Continuity of $\phi'(x)$ is equivalent to $V_i(x_{i+1}) = V_{i+1}(x_{i+1})$, which we can rearrange to find an equation for K_{i+1} in terms of K_i and $\phi(x_{i+1})$:

$$K_{i+1} - K_i = (s_i - s_{i+1})\phi(x_{i+1})^2(1 - 2\phi(x_{i+1})/3).$$

What about the frequency at the points the selection changes, $\phi(x_i)$? Well, if $\phi(x)$ is monotone on $[x_i, x_{i+1})$ then we can without loss of generality take $\phi_i^* = 0$ or 1 depending on the sign of s_i. Otherwise, let ϕ_i^* be the (unique) root of V_i in $[x_i, x_{i+1})$, so $V_i(\phi_i^*) = 0$. In this case, ϕ_i^* is the maximum or minimum of ϕ in the interval: if $s > 0$, then $\phi_i = \max\{\phi(x) : x \in [x_i, x_{i+1})\}$. Recall we defined F_i using ϕ_i^*; now using the fact that ϕ is monotone with the opposite sign on either side of ϕ_i^*, $x_{i+1} - x_i = F_i(\phi(x_{i+1})) - F_i(\phi_i^*) + F_i(\phi(x_i)) - F_i(\phi_i^*)$, and that $F(\phi_i^*) = 0$, we know that the length of the ith stretch is

$$x_{i+1} - x_i = \pm\left(F_i(\phi(x_i)) + F_i(\phi(x_{i+1}))\right).$$

Note that all the $\pm$'s are easily relatable to the signs of s_i. If we knew $\phi(x_1)$ and $\phi'(x_1)$, then we'd be able to solve the equations for $\phi(x_i)$ and K_i recursively upwards. In some cases, such as Slatkin [1973], we can infer $\phi(0)$ and $\partial_x\phi(0)$ by spatial symmetry. In other cases, we are only given $\phi(-\infty)$ and $\phi(\infty)$, and have to work inwards from the ends.

A.1 Doing the integrals

An important ingredient in the above method is the integral (66), to which a method for solving the Korteweg-de Vries equation can be applied [NEQwiki, 2013]. Recall that $V_i(\phi) = s_i\phi^2(1-2\phi/3) + K_i$, with $K_i = V_i(0)$ chosen to match V at the boundaries; since we're just working within an interval with s constant, we can rescale space by $\sqrt{2s_i}/\sigma$, so that s and σ drop from the equation. We then want to integrate

$$\int \frac{d\phi}{\sqrt{-V(\phi)}} = \int \frac{d\phi}{\sqrt{\phi^2(1-2\phi/3)+K}},$$

over a domain where V is always negative. Let $\phi^2(1-2\phi/3)+K = (\alpha-\phi)(\phi-\beta)(\phi-\gamma)$, and change variables first to $y^2 = (\alpha-\phi)$, and then to $x = y/\sqrt{\alpha-\beta}$, so that

$$\begin{aligned}\frac{d\phi}{\sqrt{\phi^2(1-2\phi/3)+K}} &= \frac{-2dy}{\sqrt{(\alpha-\beta-y^2)(\alpha-\gamma-y^2)}} \\ &= \frac{-2dx}{\sqrt{\alpha-\gamma}\sqrt{(1-x^2)(1-S^2x^2)}},\end{aligned}$$

with $S^2 = (\alpha-\beta)/(\alpha-\gamma)$. Now Jacobi's incomplete elliptic integral of the first kind is defined by

$$F(x;k) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}},$$

and the Jacobian elliptic function $\mathrm{sn}(x;k)$ is the inverse: $F(\mathrm{sn}(x;k);k) = x$. As $k \to 0$, $\mathrm{sn}(x;k) \to \sin(x)$, while as $k \to 1$, $\mathrm{sn}(x;k) \to \sinh(x)$.

B More information about the simulations

Here we provide additional details regarding the simulations we used to validate the theory in figures 4 and 5.

μ	ρ	s_p	s_m	T_{adapt}	p_{adapted}	n	μ	ρ	s_p	s_m	T_{adapt}	p_{adapted}	n
1e-05	25	0.01	-0.10	19630	0.06	100	1e-05	600	0.01	-0.10	1082	1.00	40
1e-05	25	0.01	-0.01	15533	0.55	100	1e-05	600	0.01	-0.01	924	1.00	40
1e-05	25	0.01	-0.00	14353	0.72	100	1e-05	600	0.01	-0.01	828	1.00	40
1e-05	25	0.01	-0.00	13282	0.74	100	1e-05	600	0.01	-0.00	846	1.00	40
1e-05	50	0.01	-0.10	11587	0.33	40	1e-05	600	0.01	-0.00	966	1.00	40
1e-05	50	0.01	-0.10	11586	0.38	40	1e-05	800	0.01	-0.10	802	1.00	40
1e-05	50	0.01	-0.01	9419	0.88	40	1e-05	800	0.01	-0.10	1182	1.00	40
1e-05	50	0.01	-0.01	8848	0.82	40	1e-05	800	0.01	-0.01	846	1.00	40
1e-05	50	0.01	-0.00	6529	0.95	40	1e-05	800	0.01	-0.01	607	1.00	40
1e-05	50	0.01	-0.00	9986	0.90	40	1e-05	800	0.01	-0.00	646	1.00	40
1e-05	100	0.01	-0.10	7425	0.91	316	1e-05	800	0.01	-0.00	507	1.00	40
1e-06	100	0.01	-0.10	21010	0.23	217	1e-05	1000	0.01	-0.10	615	1.00	256
1e-05	100	0.01	-0.03	6036	0.98	195	2.4e-06	1000	0.01	-0.10	5309	0.96	257
1e-06	100	0.01	-0.03	19859	0.31	194	1e-05	1000	0.01	-0.03	523	1.00	194
1e-05	100	0.01	-0.01	5408	0.98	315	1e-06	1000	0.01	-0.03	6425	0.95	195
1e-06	100	0.01	-0.01	20606	0.28	216	1e-05	1000	0.01	-0.01	449	1.00	254
1e-05	100	0.01	-0.00	4745	0.99	195	2.4e-06	1000	0.01	-0.01	4806	0.96	256
1e-06	100	0.01	-0.00	19564	0.33	195	1e-05	1000	0.01	-0.00	460	1.00	193
1e-05	100	0.01	-0.00	4894	0.99	315	1e-06	1000	0.01	-0.00	5203	0.98	194
1e-06	100	0.01	-0.00	19417	0.33	216	1e-05	1000	0.01	-0.00	444	1.00	255
1e-05	100	0.01	-0.00	4940	0.99	316	2.4e-06	1000	0.01	-0.00	4037	1.00	257
1e-06	100	0.01	-0.00	18025	0.40	217	1e-05	1000	0.01	-0.00	404	1.00	216
1e-05	200	0.01	-0.10	4296	0.97	40	1e-06	1000	0.01	-0.00	4567	1.00	216
1e-05	200	0.01	-0.10	3736	1.00	40	1e-05	1200	0.01	-0.10	501	1.00	40
1e-05	200	0.01	-0.01	2943	1.00	40	1e-05	1200	0.01	-0.10	460	1.00	40
1e-05	200	0.01	-0.01	2402	1.00	40	1e-05	1200	0.01	-0.01	336	1.00	40
1e-05	200	0.01	-0.00	1899	1.00	40	1e-05	1200	0.01	-0.01	431	1.00	40
1e-05	200	0.01	-0.00	1992	1.00	40	1e-05	1200	0.01	-0.00	390	1.00	40
1e-05	400	0.01	-0.10	1740	1.00	100	1e-05	1200	0.01	-0.00	372	1.00	40
1e-05	400	0.01	-0.01	1471	1.00	100	1e-05	1600	0.01	-0.10	299	1.00	100
1e-05	400	0.01	-0.00	1446	1.00	100	1e-05	1600	0.01	-0.01	462	0.99	100
1e-05	400	0.01	-0.00	1224	1.00	100	1e-05	1600	0.01	-0.00	241	1.00	100
1e-05	600	0.01	-0.10	1276	1.00	40	1e-05	1600	0.01	-0.00	274	1.00	100

Table 1. Parameter values used in estimates of mean time to adaptation by mutation of figure 4. All simulations also used a linear grid of 501 demes with a patch of 99 demes in the center, the migration model described in Simulation methods, $\mu = 10^{-5}$, and $s_p = .0023$ (calculated as the growth rate as described in the text). T_{adapt} is the mean time until 100 B alleles were present in the patch, and p_{adapted} is the proportion of the simulations that adapted by the 25,000 generations. The simulation began with no B alleles.

R	ρ	s_p	s_m	T_{adapt}	p_{adapted}	n	R	ρ	s_p	s_m	T_{adapt}	p_{adapted}	n
100	100	0.01	-1e-01	24975	0.00	300	60	1000	0.01	-3e-02	22224	0.21	200
10	100	0.01	-1e-01	24975	0.00	40	100	1000	0.01	-1e-02	18098	0.55	299
120	100	0.01	-1e-01	24975	0.00	295	10	1000	0.01	-1e-02	14	1.00	100
160	100	0.01	-1e-01	24975	0.00	40	120	1000	0.01	-1e-02	23242	0.14	297
20	100	0.01	-1e-01	24975	0.00	40	160	1000	0.01	-1e-02	24762	0.01	100
40	100	0.01	-1e-01	24975	0.00	40	20	1000	0.01	-1e-02	136	1.00	100
60	100	0.01	-1e-01	24975	0.00	300	40	1000	0.01	-1e-02	874	1.00	99
80	100	0.01	-1e-01	24975	0.00	40	60	1000	0.01	-1e-02	3485	1.00	300
100	100	0.01	-3e-02	24975	0.00	200	80	1000	0.01	-1e-02	9204	0.96	100
120	100	0.01	-3e-02	24975	0.00	199	100	1000	0.01	-3e-03	4168	1.00	200
60	100	0.01	-3e-02	24903	0.01	200	120	1000	0.01	-3e-03	6304	1.00	200
100	100	0.01	-1e-02	24354	0.04	300	60	1000	0.01	-3e-03	1273	1.00	199
10	100	0.01	-1e-02	1705	0.95	40	100	1000	0.01	-1e-03	2813	1.00	299
120	100	0.01	-1e-02	24802	0.01	299	10	1000	0.01	-1e-03	12	1.00	100
160	100	0.01	-1e-02	24975	0.00	40	120	1000	0.01	-1e-03	4002	1.00	300
20	100	0.01	-1e-02	3601	0.95	40	160	1000	0.01	-1e-03	7386	1.00	100
40	100	0.01	-1e-02	9463	0.88	40	20	1000	0.01	-1e-03	121	1.00	100
60	100	0.01	-1e-02	18674	0.45	299	40	1000	0.01	-1e-03	507	1.00	100
80	100	0.01	-1e-02	23899	0.12	40	60	1000	0.01	-1e-03	1115	1.00	300
100	100	0.01	-3e-03	14601	0.80	200	80	1000	0.01	-1e-03	1941	1.00	100
120	100	0.01	-3e-03	18645	0.53	199	100	1000	0.01	-1e-04	2528	1.00	300
60	100	0.01	-3e-03	5424	0.98	200	10	1000	0.01	-1e-04	11	1.00	99
100	100	0.01	-1e-03	8992	0.96	299	120	1000	0.01	-1e-04	3590	1.00	298
10	100	0.01	-1e-03	522	1.00	40	160	1000	0.01	-1e-04	5648	1.00	100
120	100	0.01	-1e-03	11375	0.93	300	20	1000	0.01	-1e-04	122	1.00	100
160	100	0.01	-1e-03	16930	0.68	40	40	1000	0.01	-1e-04	498	1.00	100
20	100	0.01	-1e-03	1518	0.97	40	60	1000	0.01	-1e-04	1051	1.00	298
40	100	0.01	-1e-03	2234	1.00	40	80	1000	0.01	-1e-04	1796	1.00	100
60	100	0.01	-1e-03	4277	0.98	300	10	4000	0.01	-1e-01	0	1.00	20
80	100	0.01	-1e-03	5328	1.00	40	160	4000	0.01	-1e-01	24975	0.00	20
100	100	0.01	-1e-04	6867	0.99	299	20	4000	0.01	-1e-01	4406	1.00	20
10	100	0.01	-1e-04	461	1.00	40	40	4000	0.01	-1e-01	24975	0.00	20
120	100	0.01	-1e-04	9347	0.99	300	80	4000	0.01	-1e-01	24975	0.00	20
160	100	0.01	-1e-04	12731	0.90	40	10	4000	0.01	-1e-02	3746	0.85	20
20	100	0.01	-1e-04	860	1.00	40	160	4000	0.01	-1e-02	24975	0.00	20
40	100	0.01	-1e-04	2031	1.00	40	20	4000	0.01	-1e-02	58	1.00	20
60	100	0.01	-1e-04	3529	0.99	300	40	4000	0.01	-1e-02	362	1.00	20
80	100	0.01	-1e-04	4839	1.00	40	80	4000	0.01	-1e-02	4025	1.00	20
100	1000	0.01	-1e-01	24975	0.00	300	10	4000	0.01	-1e-03	6244	0.75	20
10	1000	0.01	-1e-01	1004	0.99	100	160	4000	0.01	-1e-03	4399	1.00	20
120	1000	0.01	-1e-01	24975	0.00	298	20	4000	0.01	-1e-03	52	1.00	20
160	1000	0.01	-1e-01	24975	0.00	100	40	4000	0.01	-1e-03	281	1.00	20
20	1000	0.01	-1e-01	18403	0.42	100	80	4000	0.01	-1e-03	1209	1.00	20
40	1000	0.01	-1e-01	24965	0.00	100	10	4000	0.01	-1e-04	3746	0.85	20
60	1000	0.01	-1e-01	24975	0.00	300	160	4000	0.01	-1e-04	3798	1.00	20
80	1000	0.01	-1e-01	24975	0.00	100	20	4000	0.01	-1e-04	52	1.00	20
100	1000	0.01	-3e-02	24975	0.00	200	40	4000	0.01	-1e-04	279	1.00	20
120	1000	0.01	-3e-02	24975	0.00	200	80	4000	0.01	-1e-04	1039	1.00	20

Table 2. Parameter values used in estimates of mean time to adaptation by migration of figure 4. All simulations also used a linear grid of 501 demes with two patches of 99 demes each, separated by R demes in the center; the migration model described in Simulation methods, $\mu = 10^{-5}$, and $s_p = .0023$ (calculated as the growth rate as described in the text). T_{adapt} is the mean time until 100 B alleles were present in the patch, and p_{adapted} is the proportion of the simulations that adapted by the 25,000 generations. At the start of the simulation, one patch was initialized with a frequency of 0.8 B alleles, which were absent elsewhere.

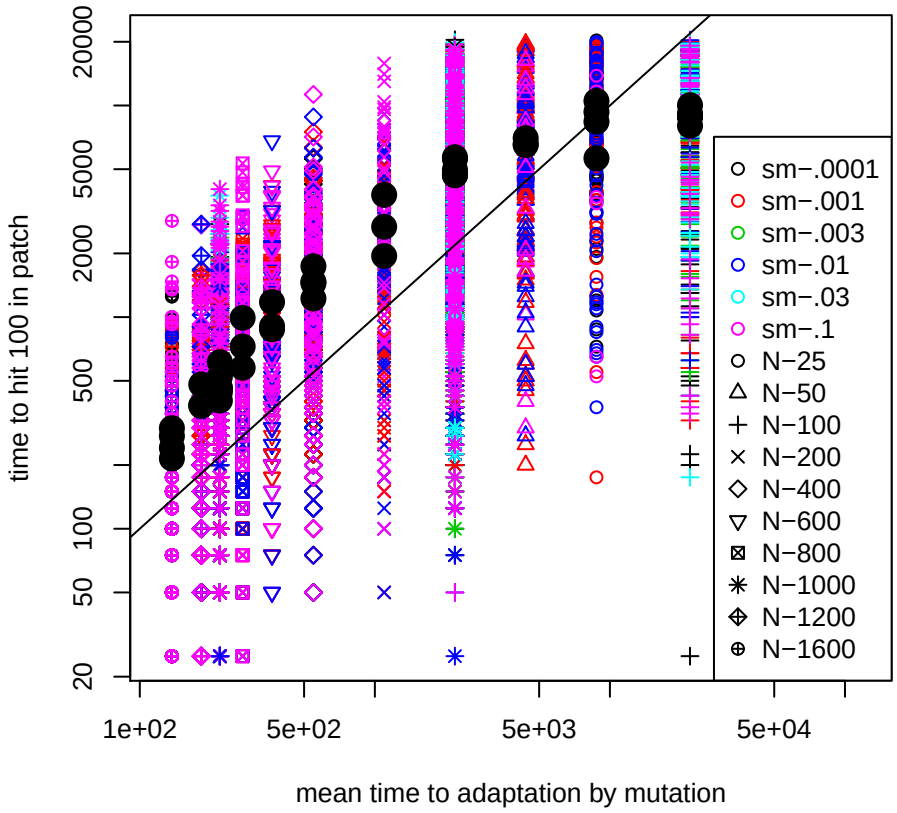

Figure S1. The same data shown in the left panel of figure 4, but all times shown (not just the interquartile ranges), and including those parameter values at which most of the simulations did not adapt by 25,000 generations.

Supplemental material

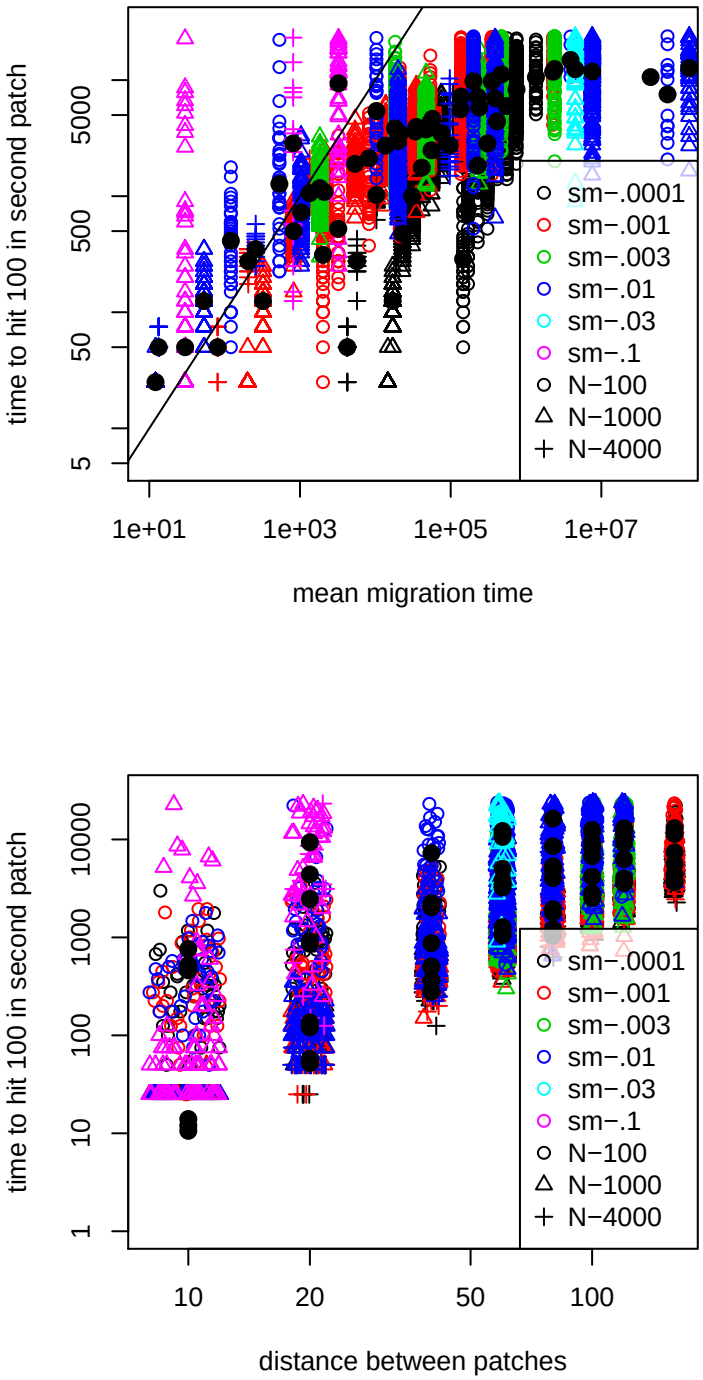

Figure S2. The same data shown in the right panel of figure 4, but all times shown (not just the interquartile ranges), and including those parameter values at which most of the simulations did not adapt by 25,000 generations. The upper panel has the predicted time to adaptation on the horizontal axis as in figure 4, and the lower panel has, for comparison, the raw distance between patches (which predicts time to adaptation, but not as well).

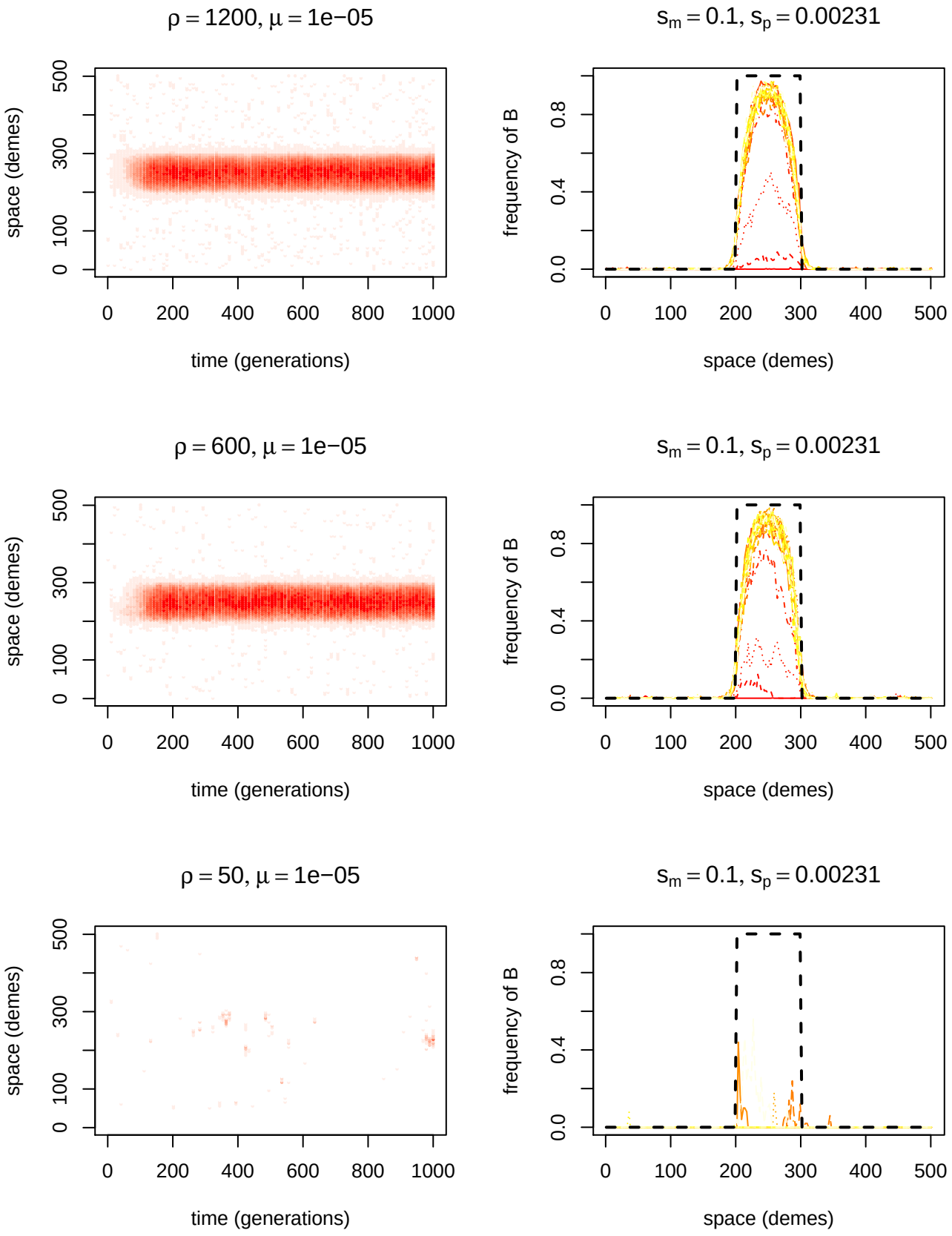

Figure S3. Randomly chosen simulations of adaptation by new mutation with $s_m = 0.1$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

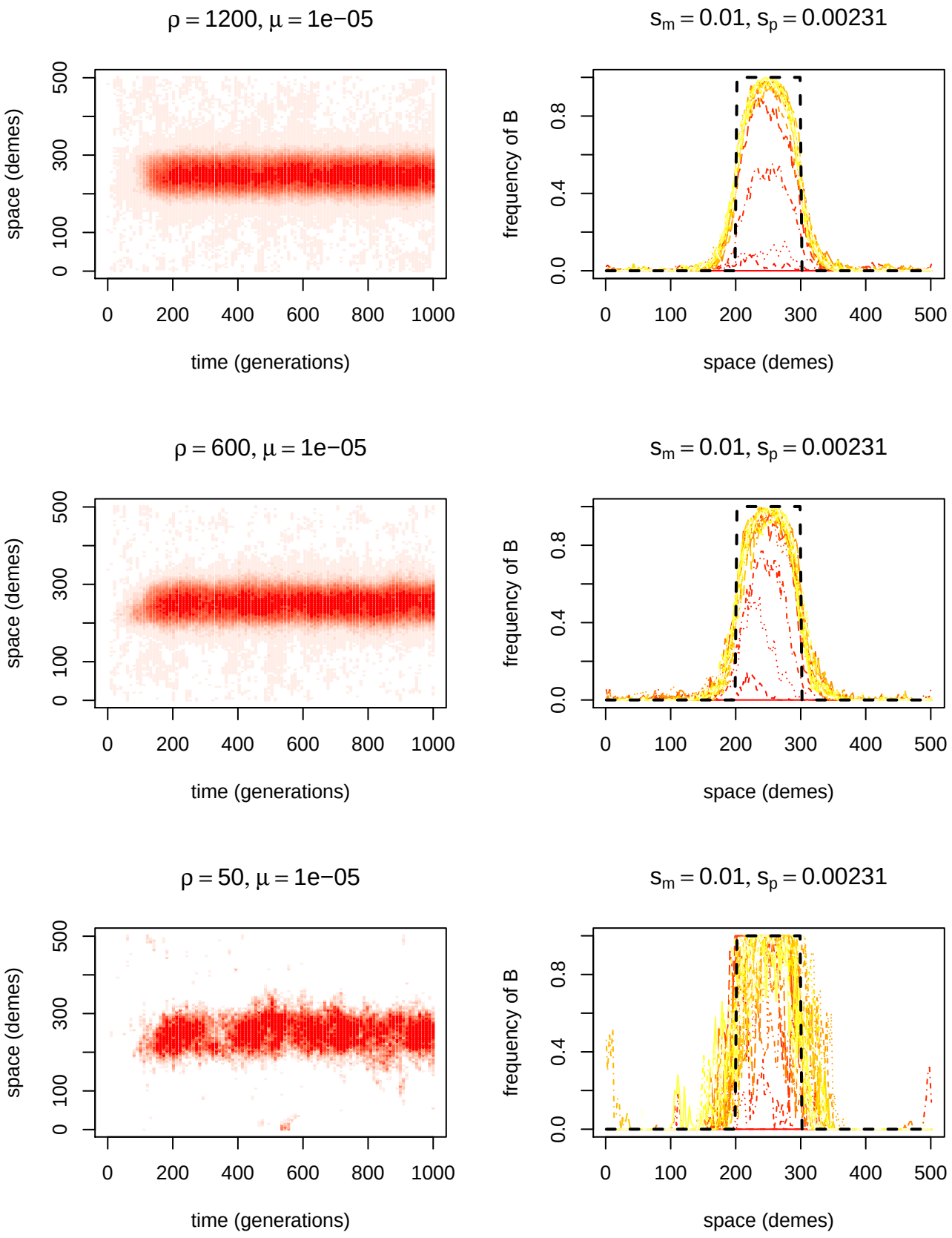

Figure S4. Randomly chosen simulations of adaptation by new mutation with $s_m = 0.01$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

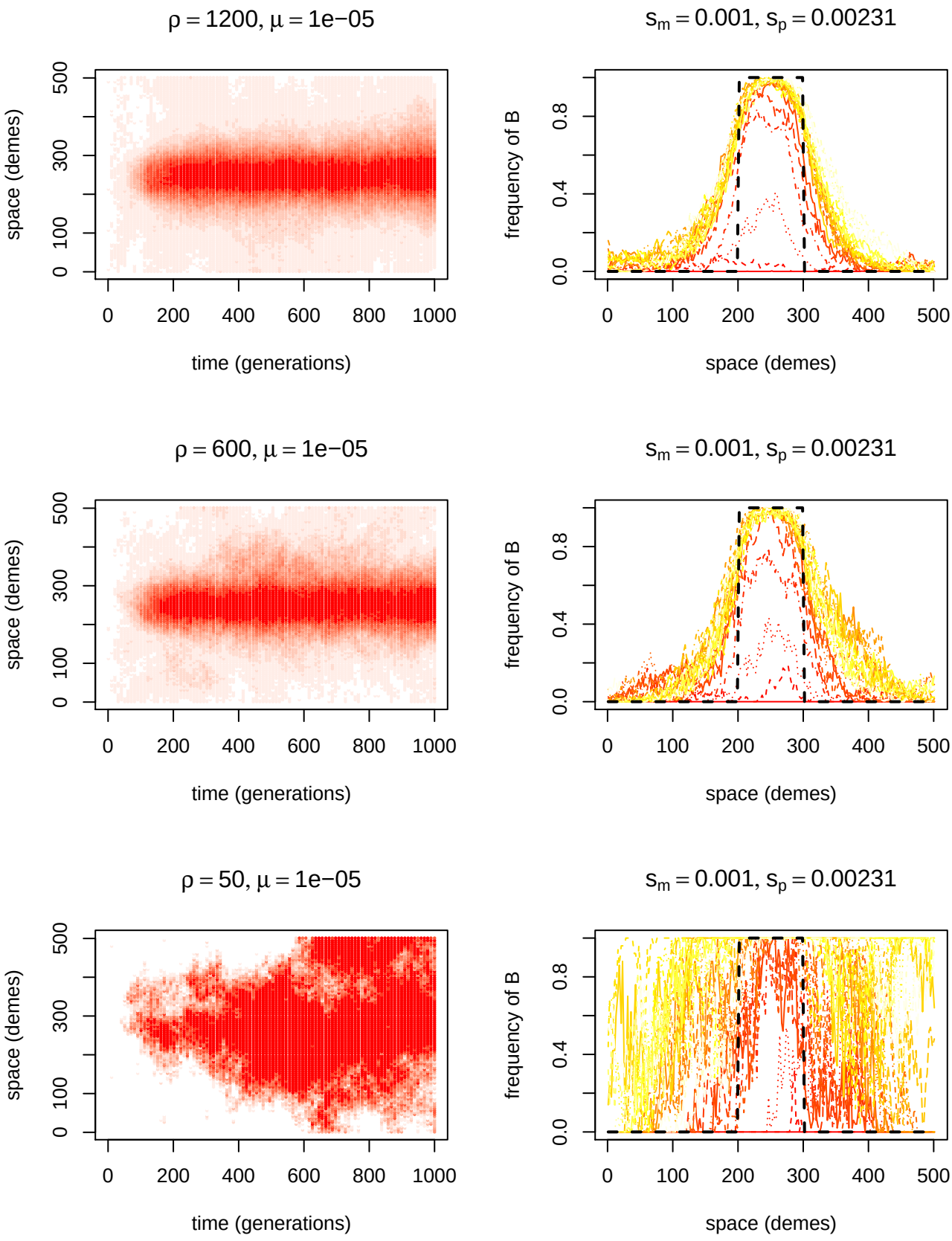

Figure S5. Randomly chosen simulations of adaptation by new mutation with $s_m = 0.001$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

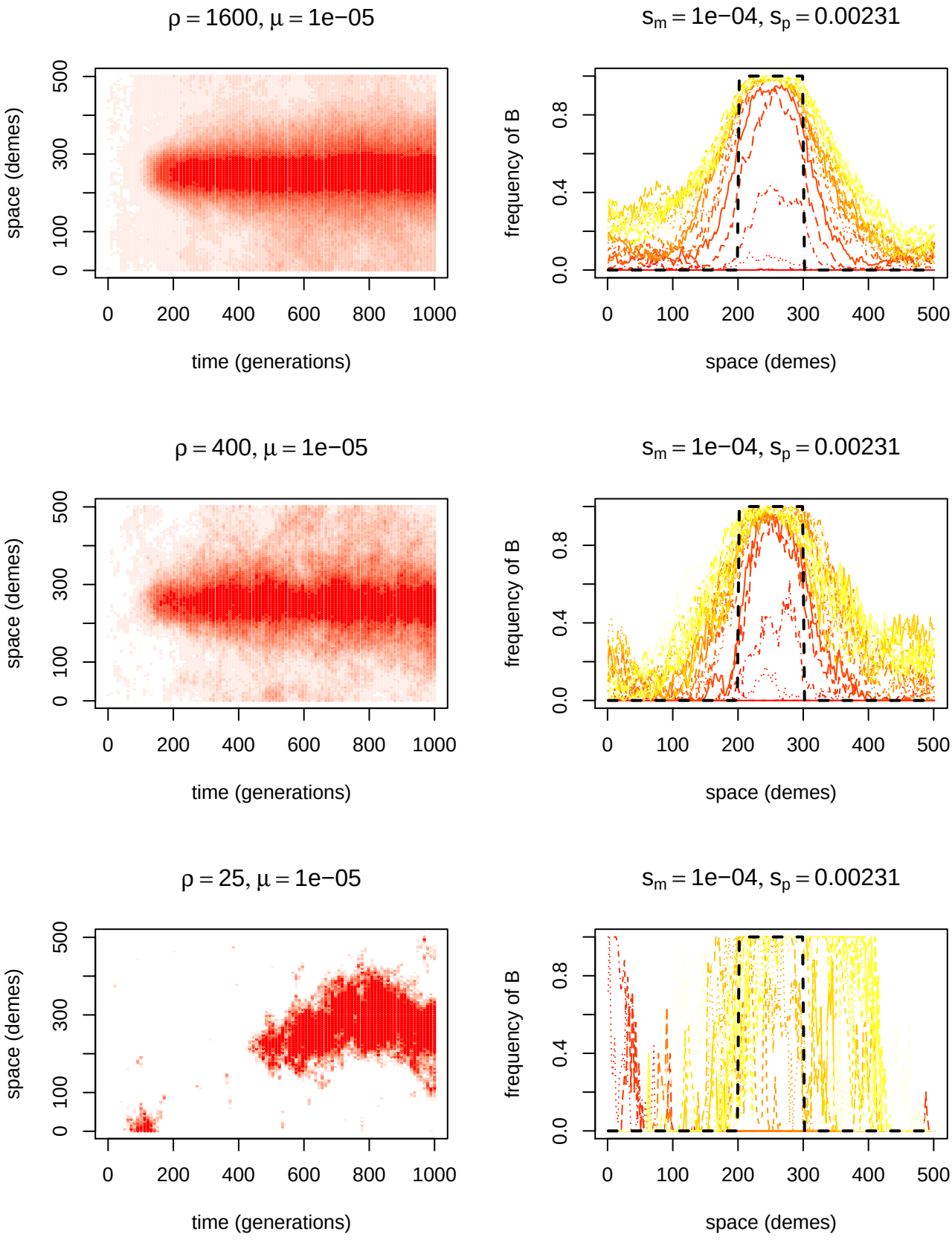

Figure S6. Randomly chosen simulations of adaptation by new mutation with $s_m = 0.001$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

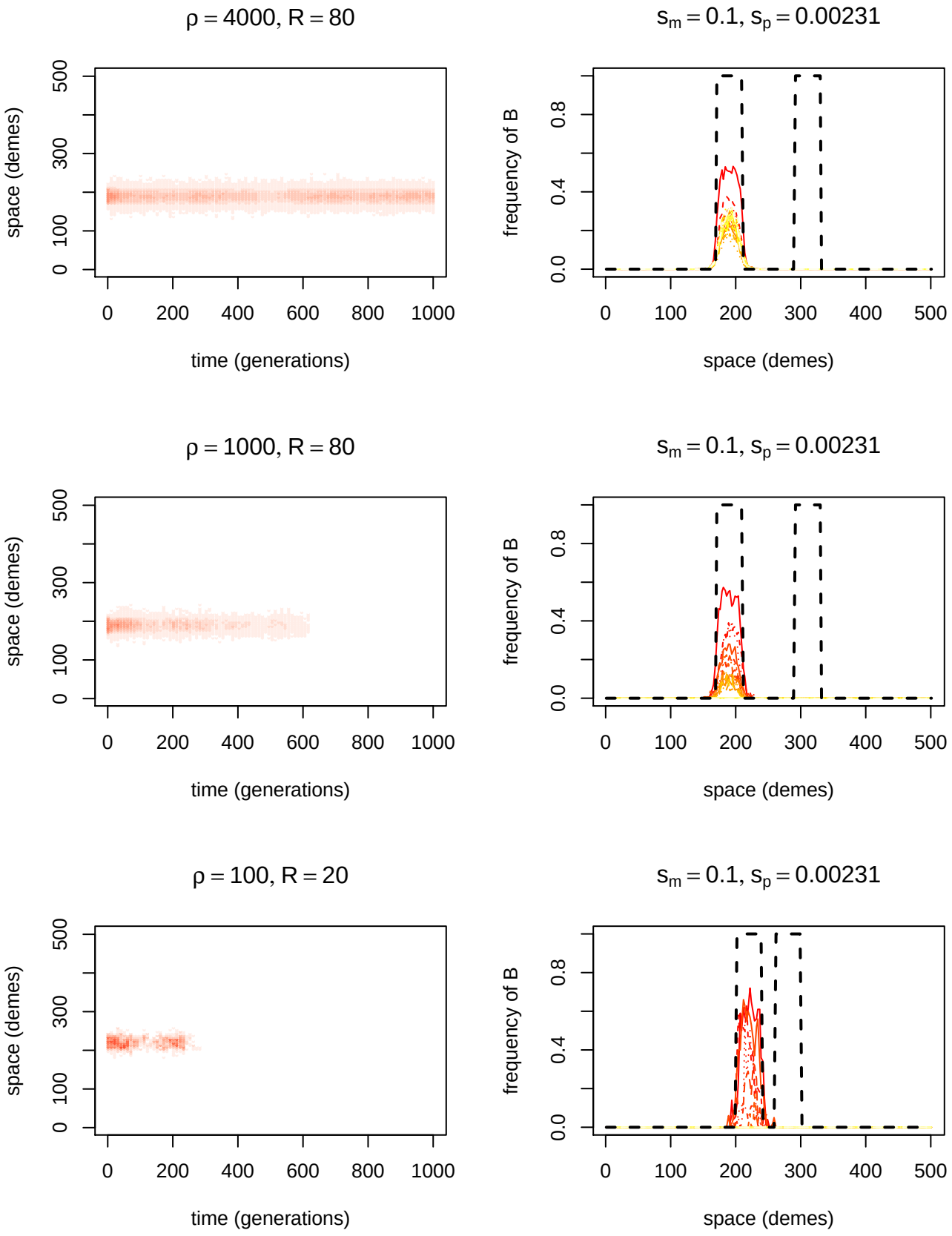

Figure S7. Randomly chosen simulations of adaptation by migration with $s_m = 0.1$, $R = 80$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

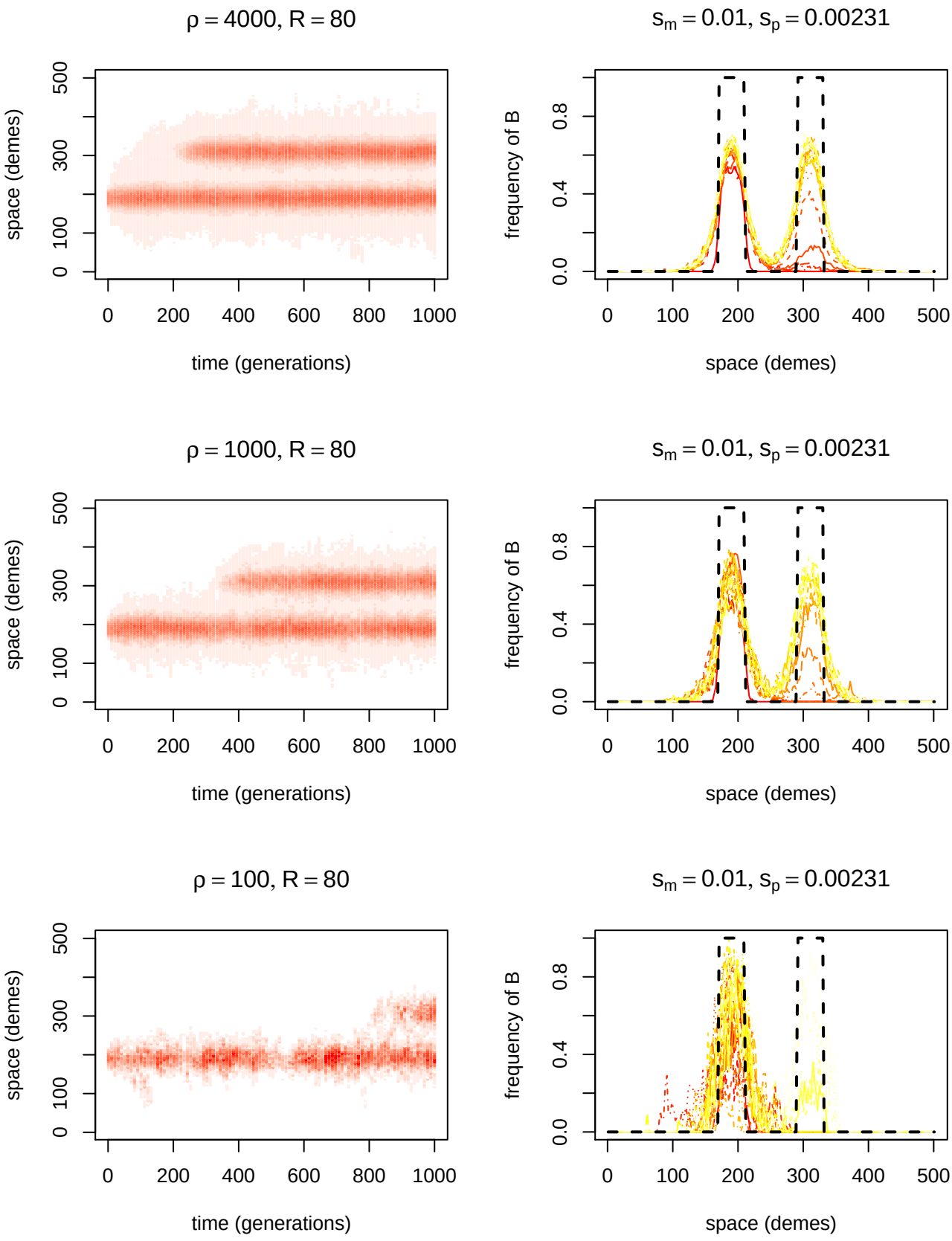

Figure S8. Randomly chosen simulations of adaptation by migration with $s_m = 0.01$, $R = 80$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

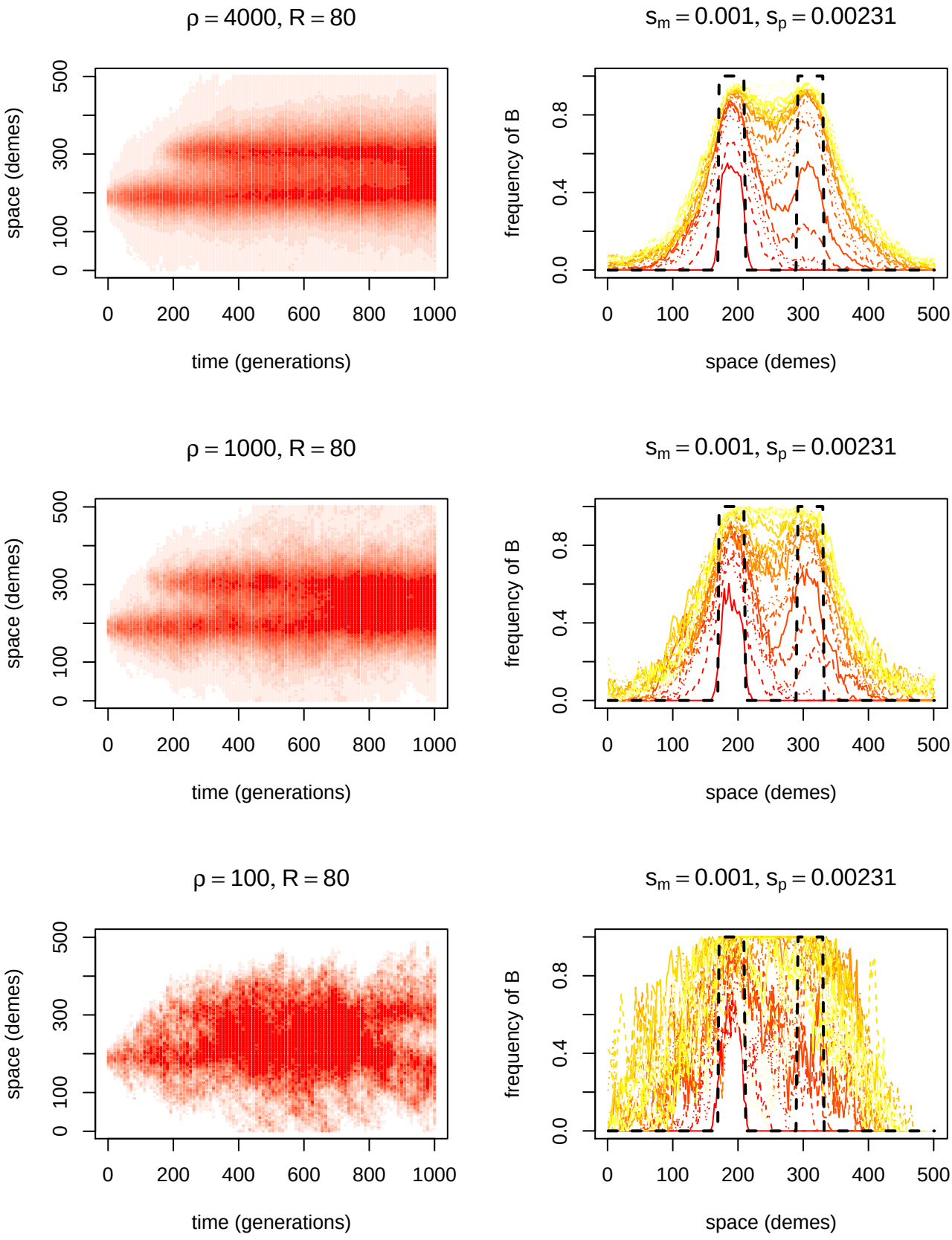

Figure S9. Randomly chosen simulations of adaptation by migration with $s_m = 0.001$, $R = 80$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is advantageous.

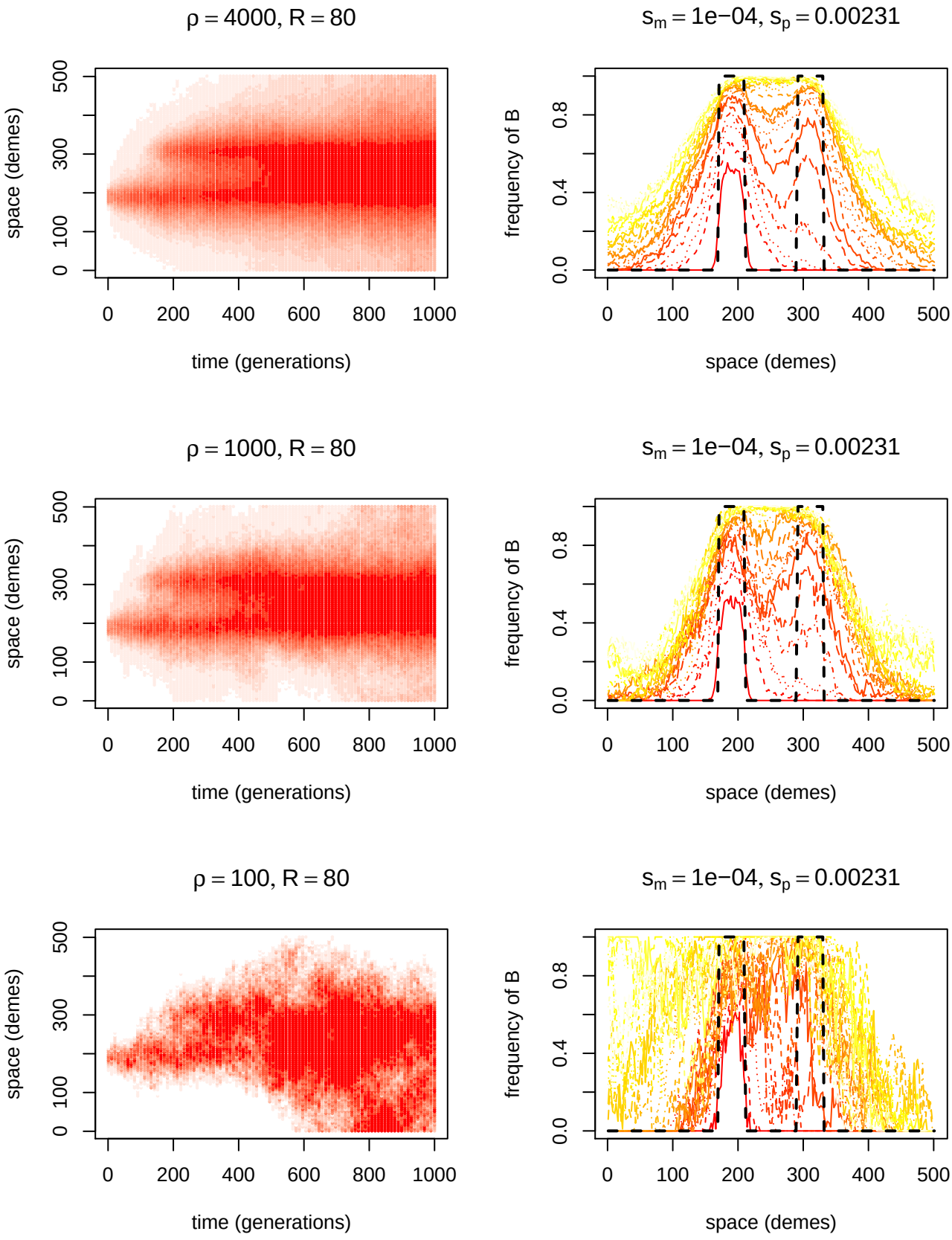

Figure S10. Randomly chosen simulations of adaptation by migration with $s_m = 0.0001$, $R = 80$, $\sigma \approx 1$, and ρ varying. On the left of each is a space-time heatmap of the local frequency of B alleles; and on the right are twenty-five curves showing the frequencies of B at evenly spaced time points (i.e. each line represents a vertical slice through the plot on the left); dotted black lines indicate the patches where B is