Convergent Evolution During Local Adaptation to Patchy Landscapes

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

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

July 7, 2014

[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 as different alleles arise and spread in different patches, or through the spread of 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 how effectively selection acts 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 new, locally adapted allele to appear in a patch of habitat where it is favored through new 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 and how long does this signal persist for? To answer these questions we decompose the migration–selection equilibrium surrounding an already adapted patch into families of migrant alleles, 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.

1 Introduction

The convergent evolution of similar phenotypes in response to shared selection pressures is a testament to the power of selection to repeatedly 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 can also occur within species, if individuals within a species adapt to the same environment, in parallel, 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. larger mutational target 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 standing variation for an allele increases the probability that the selective

sweep will be *soft* (beginning from a base of multiple copies), leading 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 spatially 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) [**?**Turner et al., 2010, Macnair, 1991, Schat et al., 1996]; 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 is on a number of black lava flows has much darker pelage than on intervening, much lighter soil [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, 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 section 2.2 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 section 2.4, where we develop an understanding of the process by which alleles move from an existing patch to colonize a novel patch, culminating in equation (12) for the rate at which this happens. We combine these two results in section 2.5 to discuss to the probability of parallel adaptation, equation (14). To understand the genomic signal left by adaptations shared by migration in section 2.6 we approximate the time it will take for an allele to transit between patches, equation (19), and use this to approximate the length of haplotype that hitchhikes with it, equation (21). We then quantify how quickly this haplotype decays within patches equation (24) and use this to obtain properties about the equilibrium distribution of haplotype lengths, equation (25) in section 2.7. Finally, in section 2.8 we apply this work to understand the geographic scale of convergent evolution in *Chaetodipus intermedius*.

2 Methods

2.1 Model of a patchy landscape

Consider a population spread across a landscape, within which are patches of habitat to which individuals are (initially) poorly adapted, surrounded by large areas to which the population is well 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 $1 + s_p$ relative to the unmutated type (b) in these "poor" habitat patches. Outside of these patches we assume that the new allele has a relative fitness $1 - s_m$ when rare in the intervening areas, with s_p and s_m both positive. (Here we take "fitness" to be the intrinsic growth rate of the type when rare.) To simplify matters, we assume that the disadvantage s_m is sufficiently large that, on the relevant timescale, the allele is very unlikely to fix locally in the regions where it is at a disadvantage. (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 are mainly interested in whether the allele can escape initial loss by drift when rare. Therefore, we do not have to 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).

Past work on spatial models We will make use of several previous results from the literature on migration in spatially varying selection models. 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 $\sigma/\sqrt{s_m}$. 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 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].

2.2 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 arising at location x relative to 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 symmetric about the edge of the patch. We examine this more generally via exact calculation of $p(x)$ in Appendix A.

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 ($2s_p/\xi^2$). 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 $1/\lambda_{\text{mut}}$.

2.3 Simulation methods

First we briefly describe the simulations used below for illustration and validation. We simulated forward-time dynamics of the number of alleles of each type in a regular grid 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 migrated to demes no more than six steps away according to a symmetric migration kernel. 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 reproduction rate in simulations for type b alleles was 30%; this was then multiplied by $1 + s$ to get the probability of reproduction for type B. These s are the values 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 B alleles, then each lineage picked a label uniformly from $1 \dots n$, and those picking the same label coalesced. Since reproduction is Poisson, this correctly samples from the distribution of lineages given the population dynamics.

2.4 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 dissect the process by which an allele transits between the patches, along the way obtaining other useful information about the genealogy of B alleles.

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 location 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. These asymptotics fit simulations quite well, as shown in Figure 1. To be clear about the assumptions (especially regarding scaling) implicit here, we provide a derivation in Appendix B; as well as justification for the asymptotics in Appendix B.1. In one dimension, the equation can be solved to give $q(x)$ in terms of Jacobi elliptic functions; see Appendix B.2. 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).

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

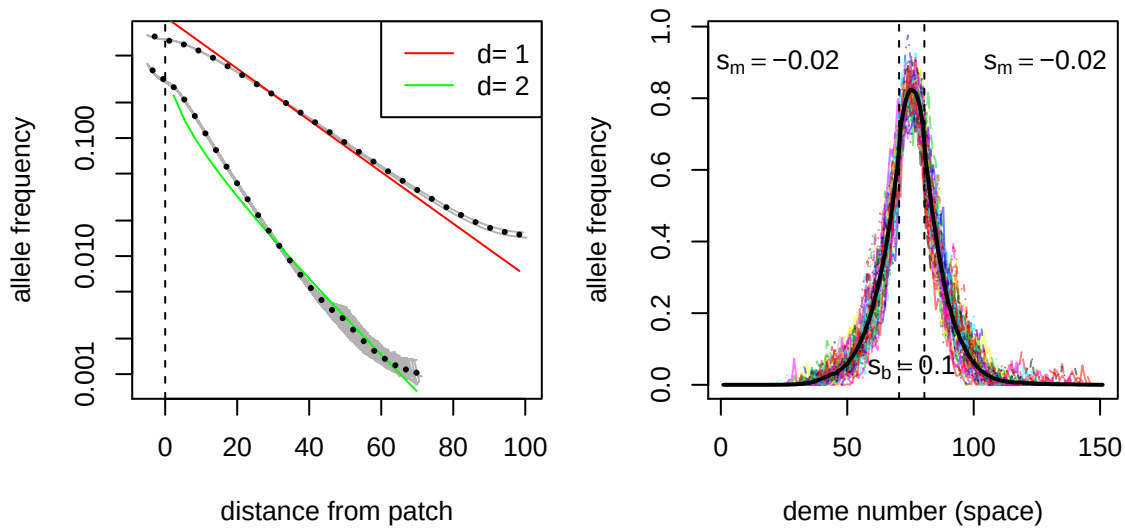

Figure 1: **(A)** Mean occupation frequencies of individual demes (grey dots) and averaged over demes at similar distances (black circles), on a one-dimensional grid of 201 demes, and a two-dimensional grid of 101×101 demes. Superimposed in color is the decay predicted by (3), with C chosen arbitrarily to give a reasonable fit. (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 a mean dispersal distance of $\sigma \approx 1.66$ times the deme spacing.

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 (12) for the rate of effective migration, an approximation that is valid if the distance R is large (at least five times $\sigma/\sqrt{2s_m}$, say).

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 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

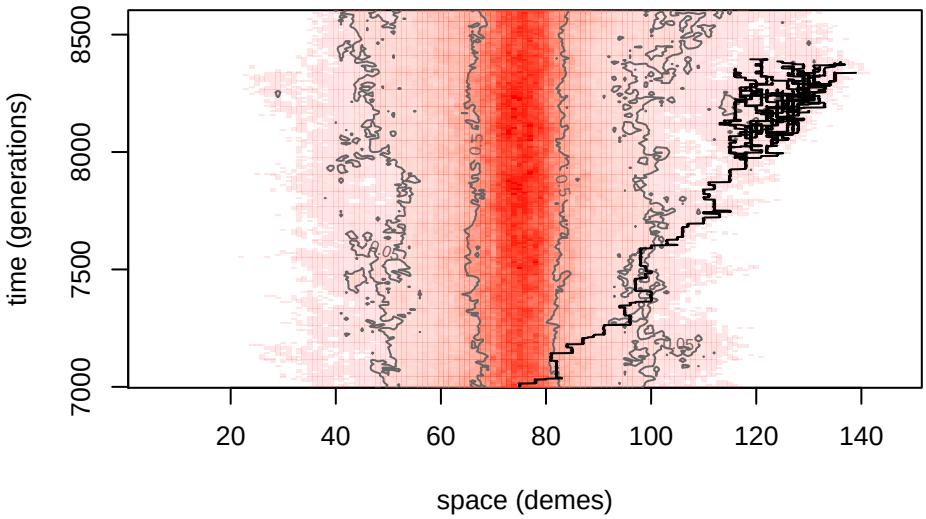

Figure 2: A segment of the temporal dynamics of the same simulation as Figure 1B. Space is again shown across the horizontal axis, with time on the vertical axis; demes are colored darker red the greater their allele frequency at the corresponding time. The 50% and 5% frequency contour lines are in grey,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

is then

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

and the quantity we wish to compute, the effective rate of 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. 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 $q(r)$ in (4). Note that λ_{mig} reflects the mean number of establishing migrant families per unit time, rather than individuals – a naive calculation using $q(x)$ that assumes migrants are independent would overestimate the effective migration rate.

2.4.1 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 patch where it is established. Consider one such family, that lives far from the original patch. As noted above, we will model this by assuming that each individual migrates, then reproduces, independently of all other family members. The mean squared migration distance is σ^2, takes the migrant in a random direction, and if migration takes the individual into a patch where the B allele is favored, we drop the individual from our 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, the mean number of offspring per individual is $\exp(-s_m) \approx 1 - s_m$, and the variance is ξ^2. 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 $e^{-s_m t}/(1 - k_e(t)) \to \mathbb{E}[K]$ as $t \to \infty$. In other words, the chance of survival to time t is 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 which sprouts independent "branches" that may or may not survive until t. Concretely, if we write Z_t for the number of offspring in the branching process alive at time t, then this gives the following decomposition:

$$(Z_t \mid Z_t > 0) \stackrel{d}{=} 1 + \sum_{s=0}^{t} \sum_{k=1}^{W_s - 1} Z_{t-s}^{(s,k)}, \tag{6}$$

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, conditioned on $W_s \geq 1$.

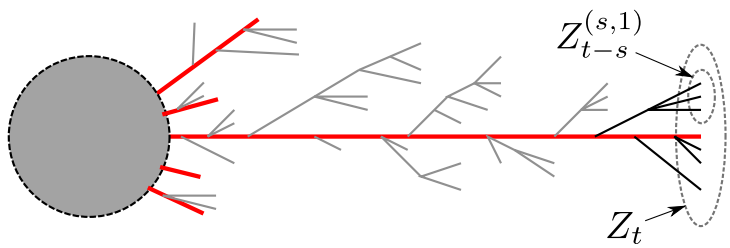

Figure 3: Cartoon of the decomposition described in the text. Migrants families that get further than r_0 from the already-established patch (grey circle) are each composed of a "trunk" lineage (in red), whose spatial motion we follow, embellished by transient branches. For the depicted long-lived family, 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).

This suggests that we add in the spatial motion by following only the motion of the "trunk" (i.e. the red line in figure 3), whose motion is is approximately Brownian, with variance σ^2 per generation. (Recall σ is the dispersal distance.) We can use these two ingredients, Brownian motion and a branching process, to compute the terms of equations (4) and (5).

The outflux of migrant alleles

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. Suppose that there is a constant outflux of migrant families from each point $x \in \partial B(r_0)$, such that the long-term proportion of individuals near a point x in $\partial B(r_0)$ that initiate new migrant families is $\gamma(x)$. To make expressions (4) and (5) more concrete, we need to replace "outflux of families" by a sum

over points at distance r_0 from the patch of the outflux γ at that point, multiplied by the relevant second quantity. By 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, 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 can be thought of as the expected total number of offspring of a family beginning at x that ever live at y. Concretely, it is the density of $\mathbb{E}[\int_0^t Z_t(\cdot)dt]$, i.e. the smooth function $u(x,y)$ that, integrated over a set S, gives the expected total number of individuals ever living in S:

$$\mathbb{E}\left[\int_0^\infty Z_t(S)dt\right] = \int_S u(x,y)dy. \tag{7}$$

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.

$$u(x,y) = \int_0^t e^{-s_m t} p_t(x,y)dt. \tag{8}$$

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

$$q(y) = \int_{\partial B(r_0)} \gamma(x)u(x,y)dx. \tag{9}$$

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, which we assume is circular, has relatively small area A and is centered at position y. It would be possible to analyze directly the probability that a given migrant family establishes in the patch, as in Barton [1987]; but we make a simpler approximation, approximating this probability 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, successfully establishing 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 $\gamma(x)$:

$$\lambda_{\text{mig}}(y) = \rho \int_{\partial B(r_0)} \gamma(x)h_A(x,y)f(A)dx. \tag{10}$$

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 that $(1-k_e(t)) \approx e^{-s_m t}/\mathbb{E}[K]$, it turns out (see Appendix C) that

$$h_A(x,y) \approx \frac{g(A)}{\mathbb{E}[K]} u(x,y), \tag{11}$$

where $g(A)$ is the total expected occupation time of the patch given the family arrives, and recall that $\mathbb{E}[K]$ is the expected family size after a large number of generations, given survival. Since the integrands in expressions 9 and 10 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 probability $f(A)$ that the family does not die out once it arrives at the new patch is approximately $\mathbb{E}[1-(1-p_e)^K] \approx 2s_p\mathbb{E}[K]/\xi^2$, if each member arriving has chance p_e of establishing, independently. Combining with expressions for $g(A)$ in Appendix C, we obtain that

$$\lambda_{\text{mig}}(y) \approx q(y) \times \frac{\rho s_p}{\xi^2 s_m} \tag{12}$$

$$\approx C\frac{\rho s_p}{\xi^2 s_m}\left(|x|\sqrt{2s_m}/\sigma\right)^{-(d-1)/2} \exp(-|x|\sqrt{2s_m}/\sigma). \tag{13}$$

This does 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. For more discussion, including an expression for the outflux of migrants $\gamma(x)$, see Appendix C.

2.5 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 (or both, or neither). 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) and successfully begins to establish, it becomes established in the poor habitat patch rapidly. 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 (12) 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 (see figure 1).

The first thing to notice is that effective migration and mutation rates will be on the same order if $A\mu \approx (1/s_m)\exp(-R\sqrt{2s_m}/\sigma)$, where recall that 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 $\rho(1/s_m)\exp(-R\sqrt{2s_m}/\sigma)$, which depends fairly strongly on the selective disadvantage s_m between patches. Equivalently, the rates are roughly equal if $R/\sigma = \log(1/(s_m A\mu))/\sqrt{2s_m}$, 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 1,000 generations ($\rho A\mu = .001$) and the population density to be 1 ($\rho = 1$, so $\log(A\mu) \approx 7$) then adaptation is mostly parallel for patches separated by gaps larger than $R/\sigma > (7 + \log(s_m))/\sqrt{s_m}$. If the selective pressure is fairly strong (say, $s_m = .05$), then for migration to be slower than mutation, the distance between patches must be at least 30 times the dispersal distance (i.e. $R/\sigma > 30$) for convergence to occur. If s_m is much smaller, say $s_m = .001$, then the patches must be separated by several hundred times the dispersal distance (i.e. $R/\sigma > 300$). The situation does not change much if the influx is quite a bit higher – even if $A\mu = 1$ (so $\log(A\mu) \approx 2.3$), the required separation between patches is only reduced to $R/\sigma > 10$ with $s_m = .05$ and $R/\sigma > 150$ for $s_m = .001$. On the other hand, if the separation is much smaller than this transitional value, then adaptations are likely to be shared between patches.

One way to visualize this is by a phase diagram. In figure 4, we give a contour plot for the time until adaptation by migration ($T_{\text{mig}} = 1/\lambda_{\text{mig}}$) and by new mutation ($T_{\text{mut}} = 1/\lambda_{\text{mut}}$) as a function of various parameters. This gives us an idea of not only the parameter space where parallel adaptation is likely ($T_{\text{mig}} \ll T_{\text{mut}}$), but also where adaptation by any means takes a very long time (T_{mig} and T_{mut} both very large), an important check in applications of these results.

Taking these approximations at face value, and assuming both rates are small, 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 ($p_e \approx 2s_b/\xi^2$), 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

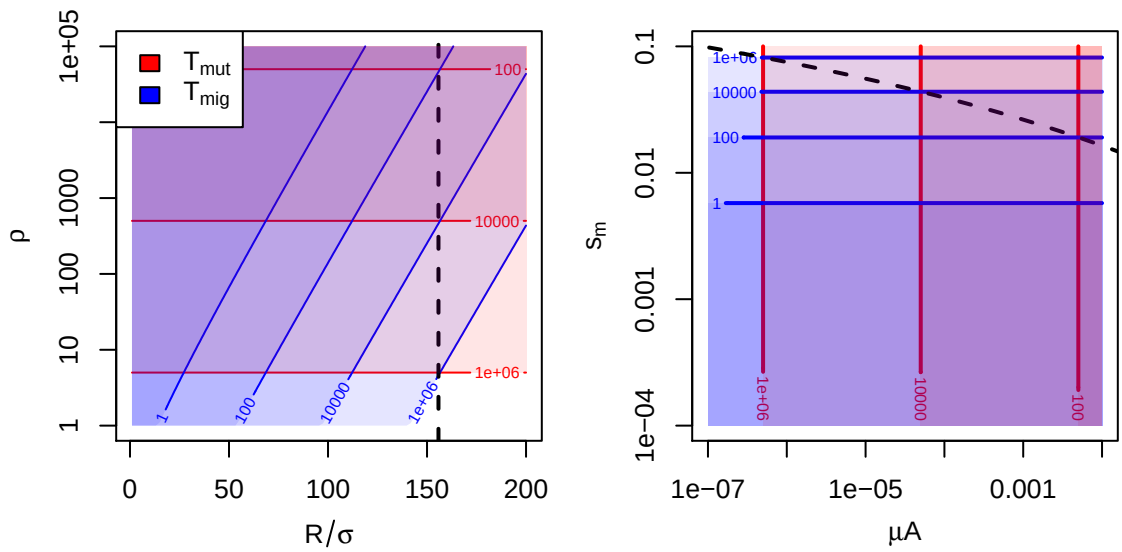

Figure 4: Shaded contour plots of mean time until adaptation in a new patch occurs by migration (T_{mig}, blue) or by mutation (T_{mut}, red), as a function of population density (ρ), distance to an already adapted patch in units of dispersal distance (R/σ), strength of deleterious selection (s_m), and mutation rate scaled by patch area (μA). The black, dotted line shows values where $T_{\text{mig}} = T_{\text{mut}}$. Default parameter values were $A = 100$, $\rho = 10$, $\mu = 10^{-8}$, $s_p = 0.1$, $s_m = 0.01$, $R/\sigma = 50$, and $\xi^2 = 1$. Note that population density ρ and mutation influx $A\mu$ are shown on a log scale.

the second adaptation is a new mutation, i.e. the probability of parallel adaptation, is

$$\frac{\lambda_{\text{mut}}}{\lambda_{\text{mut}} + \lambda_{\text{mig}}(R)} = \frac{A\mu s_m}{2A\mu s_m + \left(R\sqrt{2s_m}/\sigma\right)^{-(d-1)/2} \exp\left(-\sqrt{2s_m}R/\sigma\right)}, \tag{14}$$

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

2.5.1 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 mutational 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 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 (12), is

$$\lambda_{\text{mig}}(r_i) = C\frac{\rho s_p}{\xi^2 s_m} \sum_{i=1}^{k} \left(r_i\sqrt{2s_m}/\sigma\right)^{-(d-1)/2} \exp\left(-r_i\sqrt{2s_m}/\sigma\right), \tag{15}$$

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

$$\frac{(r_i)^{-(d-1)/2} \exp\left(-r_i\sqrt{2s_m}/\sigma\right)}{\sum_{j=1}^{k} (r_i)^{-(d-1)/2} \exp\left(-r_i\sqrt{2s_m}/\sigma\right)}. \tag{16}$$

These specify the transition rates of the process.

Since the compound parameter $s_p\rho/\xi^2$ is common to both 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 that 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 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 by distance is the same for 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 $A\mu s_m/\exp(-sqrt2s_m r/\sigma)$. 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, which 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.

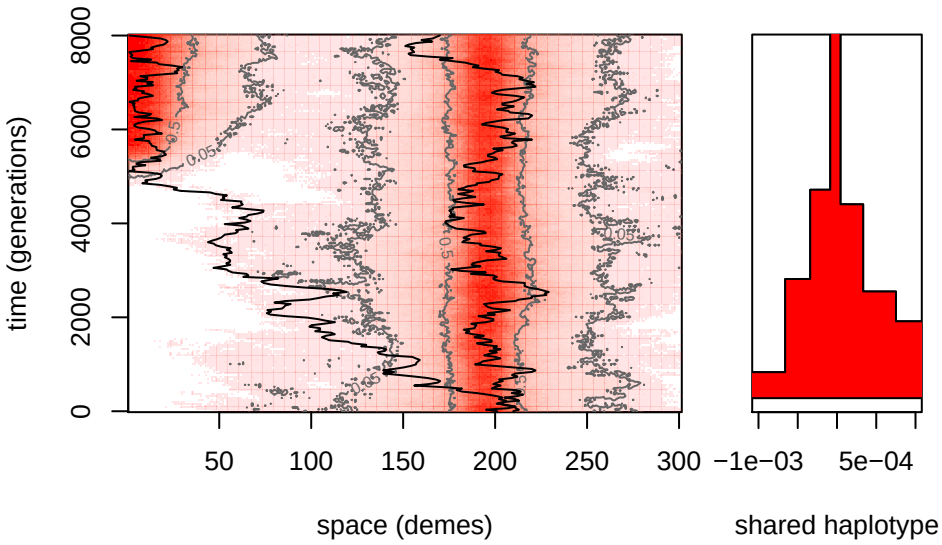

Figure 5: **Adaptation by migration:** An initially unadapted patch (10 demes at the left end of the range) is colonized around generation 5000; a representative lineage is shown in the new patch, which follows the original migrant "trunk" lineage between the patches and coalesces with a lineage in the 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.

2.6 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 individuals within a patch will 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.

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 each reproductive event in locations where the allele is at low frequency will likely recombine haplotypes of the b allele closer to the adaptive allele. In this case, the haplotype is literally hitchhiking across geography! Figure 5 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 5), 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, we assume this is rare), 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); but these factors should be minor 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. Since $\mathbb{E}[K] \approx 1/s_m$, it is sensible to make one further approximation, replacing this random time by an exponential with rate s_m. Once we do this, we can use standard results on hitting times of d-dimensional Brownian motion that is killed at rate s_m – Borodin and Salminen [see 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

$$\mathbb{E}[e^{-\ell\tau}] = \begin{cases} e^{-R\sqrt{2(s_m+\ell)}/\sigma} & d = 1 \\ \frac{K_0((R+w)\sqrt{2(s_m+\ell)}/\sigma)}{K_0(w\sqrt{2(s_m+\ell)}/\sigma)} & d = 2, \end{cases} \tag{17}$$

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

$$\mathbb{E}[e^{-\ell\tau}|\tau < \infty] = \exp\left\{-\frac{R}{\sigma}\left(\sqrt{2(\ell + s_m)} - \sqrt{2s_m}\right)\right\}. \tag{18}$$

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

$$\mathbb{E}[\tau] = (R\sqrt{2s_m}/\sigma) \times 1/(2s_m) \tag{19}$$

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

$$\mathbb{P}\{L > \ell\} = \mathbb{E}[e^{-\ell\tau}|\tau < \infty], \tag{20}$$

and so (one minus) equation (18) 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

$$\mathbb{E}[L] = \sigma\sqrt{2s_m}/R + \sigma^2/R^2 \tag{21}$$

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 (19) 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 (21) 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 $\sqrt{s_m}$, rather than multiplied; it appears here because $\sigma/\sqrt{s_m}$ is a length scale, and to convert it to a speed this is divided by $1/s_m$. This intuitive picture can be expanded by a central limit theorem: for large R, the deviation of τ about its mean is asymptotically Gaussian, and the haplotype halflength L is exponential.

2.7 Length of the shared haplotype within a patch

The other key issue in thinking about the population genetic signal of adaptation in spatial patches is the process by which haplotypes are whittled down within patches, which plays into a number of aspects of haplotype sharing. 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 21) 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. We first develop a simple approximation to the rate at which recombination breaks down haplotypes within patches.

Refer again to Figure 5. When a lineage is in a region with high frequency of B alleles (dark red), it is very unlikely to recombine off the B background. However, when it wanders out of the patch, into regions with lower frequency of B alleles, when it recombines it will usually be on to the background of the b allele.

A detailed treatment of this is well beyond the scope of this paper, but we can get a good idea of the genomic scale of these effects through heuristic arguments. To understand the probability of a recombination moving our lineage off of the B background we need to understand what fraction of the time our lineage spends in geographic regions where b is common. Consider the lineage of a single B allele, moving back through time, in a patch at migration-selection balance, i.e. with the local proportion of B alleles near x equal to $p(x)$, and suppose that the lineage stays close enough to the patch that fluctuations about $p(x)$ are small. (Note that this assumption differs from previous sections, but here our focus is on a typical lineage in the patch rather than the rare lineages that make it far from the patch). If we suppose that type B individuals at location y reproduce at rate $1 + s(y)$, and the chance their offspring move to x is $m(y, x)$ (with $\sum_x m(y, x) = 1$), then the total rate of influx of B alleles from y into x is $\rho p(y)(1 + s(y))m(y, x)$. As argued by Hudson and Kaplan [1988], the rate at which a B lineage moves from x to y is this influx divided by the local abundance of B alleles, or $p(y)(1 + s(y))m(y, x)/p(x)$. After the same large-population, continuous-space limit that underlies the rest of this paper (Appendix B), if the assumption that fluctuations in $p(x)$ are small, this motion should converge to a diffusion process [as assumed by e.g. Hallatschek and Nelson, 2008], notwithstanding some technical difficulties (e.g. discontinuity of $s(x)$). This diffusion, when at x, is pulled toward regions with more B alleles by a drift term $(1 + s(x))\frac{d}{dx}\log(p(x)(1 + s(x)))$, where $s(x) = s_p$ if x is in the patch and $s(x) = s_m$ otherwise, and moves at speed $\sqrt{1 + s(x)}$ (which is nearly constant, since $s(x)$ is small). In Itô differential notation, the location X_t of a B lineage t generations ago satisfies

$$dX_t = (1 + s(X_t))\frac{d}{dx}\log(p(X_t)(1 + s(X_t)))dt + \sigma\sqrt{1 + s(X_t)}dB_t, \tag{22}$$

where B_t is a standard Brownian motion. Furthermore, it is easy to verify that the stationary distribution

of this process has density

$$\pi(x) = \frac{p(x)^2(1+s(x))}{\int_y p(y)^2(1+s(y))dy}. \tag{23}$$

A lineage at a locus at genetic distance r that begins linked to a B allele moves as a B lineage up until the most recent time it recombined from a b background to a B background, which occurs when at x with rate $r(1-p(x))(1+s(x)/2)$. The Feynman-Kac formula could then be used to write down differential equations solved by various moments of the haplotype length as a function of the sampling location of the lineage; instead, we only use the above as a guide for rough approximation.

Based on this picture of the lineage bouncing around in the patch, we propose a simple heuristic to describe the probability of recombination. If recombination is on a slower time scale than mixing within the patch, then the time until the last recombination is the integral of $r(1-p(x))(1+s(x)/2)$ against the stationary distribution (equation 23). Within the patch, $p(x) \approx 1$, so recombination off the B background is unlikely to occur, and so the rate of recombination depends to a first approximation on the proportion of the patch in which $p(x)$ is intermediate, i.e. where both B and b are present. This region where $p(x)$ is intermediate has width of order $\sigma/\sqrt{s}$, where $s = \min(s_p, s_m)$. As we increase the size of a patch, the area on which $p(x)$ is intermediate scales with the perimeter of the patch. The per generation rate of recombination of a locus linked to the b background is thus reduced by a factor roughly equal to the proportion of the patch lying within distance $\sigma/\sqrt{s}$ of the boundary, so the effective recombination rate is

$$r_{\text{eff}} = \begin{cases} r\frac{\sigma/\sqrt{s}}{\sigma/\sqrt{s}+A} & d=1 \\ r\frac{\pi\sigma/\sqrt{s}\left(\sigma/\sqrt{s}+\sqrt{A/\pi}\right)}{A+\pi\sigma/\sqrt{s}\left(\sigma/\sqrt{s}+\sqrt{A/\pi}\right)} & d=2, \end{cases} \tag{24}$$

so that T generations after the founding of a patch we expect a haplotype of genetic length $1/(r_{\text{eff}}T)$ Morgans to remain linked to the B allele in a typical lineage. Note that the length of haplotype decays at a rate roughly proportional to $\sqrt{s}/\sigma$, as the narrower the cline the less opportunity for recombination to move our lineage off the B background.

Finally we can consider the expected length of haplotype shared within a patch. To do this, we compare the rate of recombination between B and b haplotypes (equation 24) to the rate of coalescence of B haplotypes within a patch. We will assume that mixing within the patch is fast compared to the rate of drift, so that the rate of coalescence of B alleles within a patch is $\xi^2/(A\rho)$, the "variance effective population size" of the patch (this bypasses the well-known difficulties to spatial coalescence [Felsenstein, 1975, Barton et al., 2002]). The rate of recombination and coalescence will be on the same order if $\xi^2/(A\rho) \approx r(\sigma/\sqrt{s})A^{-1/d}$, and hence the genetic length of the haplotype shared between pairs of individuals within the same patches at equilibrium is on the order of

$$r = \frac{\sqrt{s}\xi^2}{\sigma\rho A^{(d-1)/d}}. \tag{25}$$

This does not depend on A in one dimension, but decreases with $\sqrt{A}$ in two dimensions. In applications, this can be compared to equation (21) for the length of shared haplotype between patches.

2.8 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 assessed. 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). These numbers give us a sense of a plausible range of characteristic lengths. 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 $s_m = 1/\sqrt{3}$ and $1/\sqrt{30}$ are reasonable values.

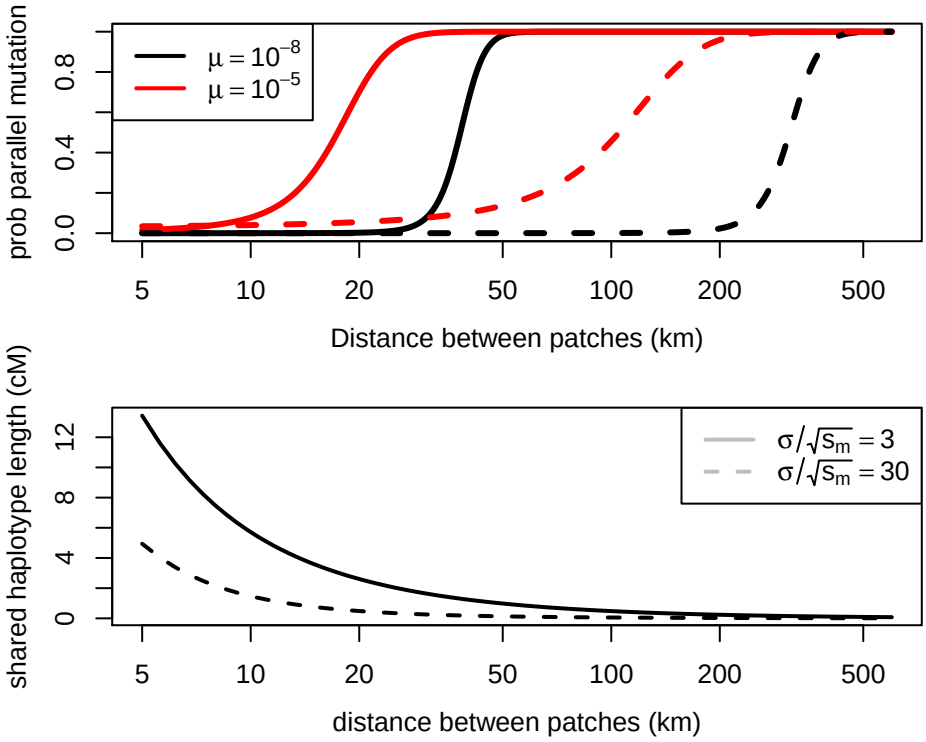

Figure 6: Top panel: The probability of parallel mutation as a function of the distance between environmental patches for two different cline widths and two different mutation rates (using equation (14)). The parameters were chosen to match the example of *Chaetodipus intermedius*; see text. Bottom panel: The initial genetic length of the haplotype shared between patches due to adaptation via migration as a function of the distance between environmental patches.

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 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 6 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 (21) to calculate how long the founding haplotype shared between populations is expected to be, as shown in Figure 6. 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.

Furthermore, given the relatively small radius of the patches of dark substrate (often $< 10\text{km}$), we should expect the haplotype to be rapidly whittled down within patches after establishment (equation 24). We estimate that the effective recombination rate is no less than 50% of the actual recombination rate. Indeed, if we take the population density to be roughly $\rho \approx 10$ per hectare (M. Nachman, personal communication), and the variance in offspring number to be $\xi^2 = 3$ (mostly to get round numbers), then the shared haplotype length within patches from equation (25) is only expected to be on the order of hundreds or thousands of bases (10^{-3}–10^{-4}cM). This suggests that unless the migration has happened fairly recently then it will be difficult to identify shared haplotypes (how recently depends on the background level of LD in the population).

3 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. This is in contrast with models of panmictic populations [Hermisson and Pennings, 2005, **?**, **?**] 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 **?**, 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 $\rho(1/s_m)\exp(-R\sqrt{2s_m}/\sigma)$. Equivalently, the critical gap size between patches past which adaptation is likely independent is $R = (\sigma/\sqrt{2s_m})\log(1/(s_m A\mu))$. 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 tended to focus 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 very low, any adaptation that does occur may have to make use of standing variation. The unstructured case was studied by Hermisson and Pennings [2005]; we study the continuous, spatial case in an upcoming paper. 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 note that the initial spatial distribution of standing variation may be).

Dominance We have mostly swept the issue of dominance under the rug 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. 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.

Long distance migration We have also ignored the possibility of very long distance migration, instead focusing on local dispersal (and using a Gaussian model). 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 sucked up in hurricanes and the like) 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 parallel 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 (21). 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 patch geometries Here we have studied circular patches of habitat at long distances from each other, while real habitat is can be much more complex – archipelagos of patches of varying sizes, or 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, and it is possible that our techniques could be applied in their much more general setting, as both are based, fundamentally, on branching process approximations.

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 convergently 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].

References

David 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`.

Dorald M. Allred and D. Elden 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, January 2008. doi: 10.1016/j.tree.2007.09.011.

NH 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, A.M. Etheridge, J. Kelleher, and A. Vber. 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.

Nick H. Barton, Frantz Depaulis, and Alison M. Etheridge. Neutral evolution in spatially continuous populations. *Theoretical Population Biology*, 61(1):31–48, February 2002. URL http://dx.doi.org/10.1006/tpbi.2001.1557.

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.

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

Robert Stephen Cantrell and Chris 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.

Robert Stephen Cantrell and Chris 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.

Brigitte Chauvin, Alain Rouault, and Anton 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, January 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.

Donald 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.

Lee 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.

Peter Donnelly and Paul 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?_ob=ArticleURL&_udi=B6V1B-45FKRNK-2F&_user=4420&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_version=1&_urlVersion=0&_userid=4420&md5=dff57a00ce4ce9c390959575a1ea8009.

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

Joseph Felsenstein. A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist*, 109(967):359–368, 1975. ISSN 00030147. URL http://www.jstor.org/stable/2459700.

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.

Norman R. French, Theodore Y. Tagami, and Page 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.

Jochen 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, With one CD-ROM (Windows, Macintosh and UNIX).

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

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.

O Hallatschek and D R Nelson. Gene surfing in expanding populations. *Theor Popul Biol*, 73(1):158–170, February 2008. doi: 10.1016/j.tpb.2007.08.008. URL http://www.ncbi.nlm.nih.gov/pubmed/17963807.

J Hermisson and P S Pennings. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352, April 2005. doi: 10.1534/genetics.104.036947. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=15716498.

H. E. Hoekstra. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity (Edinb)*, 97(3):222–234, Sep 2006.

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, November 2005. ISSN 0018067X. URL http://dx.doi.org/10.1038/sj.hdy.6800600.

R R Hudson and N L Kaplan. The coalescent process in models with selection and recombination. *Genetics*, 120(3):831–840, November 1988. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1203560/.

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

Donald 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.

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

Thomas 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/B6VJ1-45BCHTS-G/2/815bb21b2179c1f813f44cc1e5f3785a.

Simon A. Levin, Helene C. Muller-Landau, Ran Nathan, and Jérôme 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.

Yuan Lou and Eiji 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/getRecord?id=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.

Arnaud Martin and Virginie 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.

Takeo 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, February 1974. URL http://www.ncbi.nlm.nih.gov/pubmed/4407212.

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, April 2003. doi: 10.1073/pnas.0431157100. URL http://www.ncbi.nlm.nih.gov/pubmed/12704245.

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

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

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

Pleuni S. Pennings and Joachim 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.

M. Perman, J. Pitman, and M. Yor. Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1):21–39, 1992. URL http://www.springerlink.com/content/n08754882308h67n/.

Jim 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.

Edward 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.

Peter Ralph and Graham 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.

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

Erica Bree Rosenblum, Holger Römpler, Torsten Schöneberg, and Hopi 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.

Henk Schat, Riet Vooijs, and Eric 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.

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

Montgomery 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.

Cynthia C. Steiner, Holger Rompler, Linda M. Boettger, Torsten Schoneberg, and Hopi 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, November 2013. doi: 10.1038/nrg3483. URL http://www.ncbi.nlm.nih.gov/pubmed/24105273.

TL Turner, EC Bourne, EJ Von Wettberg, TT Hu, and SV Nuzhdin. Population resequencing reveals local adaptation of arabidopsis lyrata to serpentine soils. *Nat. Genet.*, 42(3):260–3, January 2010. URL http://www.ncbi.nlm.nih.gov/pubmed/20101244.

Ying Zhen, Matthew L. Aardema, Edgar M. Medina, Molly Schumer, and Peter 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 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 (a spatial branching process in this case). Furthermore, whether or not a new, beneficial mutation establishes or is lost to demographic stochasticity is determined by this 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 7.

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.

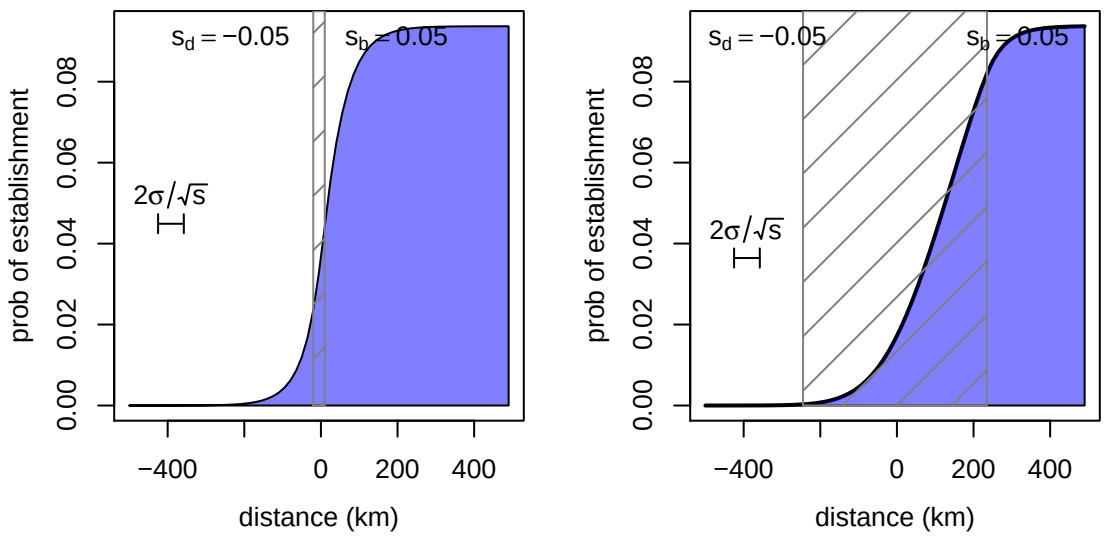

Figure 7: The probability of establishment of a new mutant allele as a function of distance from the edge of the region where it is beneficial, in an abrupt transition (left) and a gradual transition (right). The allele is deleterious to the left and beneficial to the right (with selection coefficients ∓ 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.

B 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], Pollak [1966] (or **?** or **?** or Haldane [1948]) for other arguments in equivalent models, or Etheridge [2000] and/or Dawson [1993] for general theory used below. Following is a description of a discrete 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 after a random lifetime, dying, and then leaving behind in the same deme a random number of offspring with distribution given by X; each offspring then migrates independently to a new location chosen randomly from the distribution given by $x + R$,

where she replaces a randomly chosen individual. If $x+R$ is outside of the range, then she perishes. 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,ϕ)); similarly, $\widetilde{X}_\phi$ is the number lost if the new offspring do not carry the allele (i.e. hypergeometric with parameters $(X,1-\phi)$). (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

$$\begin{aligned}\frac{\partial}{\partial t}\mathbb{E}\left[f(\Phi_t^N)\right] &= N\sum_x\Bigg\{\mathbb{E}\left[(1+s(x+R))\Phi_t^N(x+R)\left(f\left(\Phi_t^N+\frac{\bar{X}_{\Phi_t(x+R)}}{N}\delta_x\right)-f(\Phi_t^N)\right)\right] \\ &\qquad +\mathbb{E}\left[\left(1-\Phi_t^N(x+R)\right)\left(f\left(\Phi_t^N-\frac{\widetilde{X}_{\Phi_t(x+R)}}{N}\delta_x\right)-f(\Phi_t^N)\right)\right]\Bigg\} \qquad (26)\\ &= \mu\sum_x\mathbb{E}\left[\left(\partial_{\phi(x)}f(\Phi_t)\right)\left\{\Phi_t(x+R)-\Phi_t(x)+s(x+R)\Phi_t(x+R)(1-\Phi_t(x))\right\}\right]+O\left(\frac{1}{N}\right). \qquad (27)\end{aligned}$$

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 $s(x)\epsilon$, and suppose that $\Phi_t(x)$ is deterministic (and sufficiently differentiable) and let $\xi(t,x)=\Phi_{t/\epsilon}(x)$; then the previous equation with $f(\Phi)=\Phi(x)$ converges to the familiar form:

$$\partial_t\xi(t,x)=\mu\left(\frac{\sigma^2}{2}\sum_{k=1}^d\partial_{x_k}^2\xi(t,x)+s(x)\xi(t,x)(1-\xi(t,x))\right). \qquad (28)$$

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 (see section B.2); 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 (28) is

$$\partial_t\xi(t,r)=\mu\left(\frac{\sigma^2}{2}\partial_r^2\xi(t,r)+\sigma^2\frac{d-1}{2r}\partial_r\xi(t,r)+s(r)\xi(t,r)(1-\xi(t,r))\right). \qquad (29)$$

B.1 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 (29), a radially symmetric solution with $s(r) = -s < 0$ for all $r > r_0$ solves, for $r_0 < r < \infty$,

$$\partial_r^2 q(r) + \frac{d-1}{r}\partial_r q(r) - \frac{2s}{\sigma^2} q(r)(1-q(r)) = 0 \tag{30}$$

$$\lim_{r\to\infty} q(r) = \lim_{r\to\infty} \partial_r q(r) = 0 \qquad 0 < q(r) < 1 \tag{31}$$

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

$$\partial_r^2 u(r) + \frac{d-1}{r}\partial_r u(r) - \frac{2s}{\sigma^2} u(r) = 0 \tag{32}$$

$$\lim_{r\to\infty} u(r) = \lim_{r\to\infty} \partial_r u(r) = 0 \qquad . \tag{33}$$

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

$$u(r) = C'(r-r_0)^{(2-d)/2} K_{(2-d)/2}\left((r-r_0)\sqrt{2s}/\sigma\right), \tag{34}$$

where C' and r_0 are chosen to match boundary conditions. Asymptotics of Bessel functions [Gradshteyn and Ryzhik, 2007, 8.451.6] then imply that

$$q(r) \approx C r^{(1-d)/2} \exp\left(-r\sqrt{2s}/\sigma\right) + O(1/r), \tag{35}$$

where C is a different constant.

B.2 The exact solution to the equilibrium frequency

Here we describe how to solve (30) 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

$$\partial_x^2 \phi(x) = -2s(x)\phi(x)(1-\phi(x))/\sigma^2, \tag{36}$$

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

$$(\partial_x \phi(x))^2 = -\int^x 2s(x)\phi(x)(1-\phi(x))\partial_x\phi(x)dx \tag{37}$$

$$= -s_i\phi(x)^2\left(1 - \frac{2}{3}\phi(x)\right) - K_i \quad \text{for } x \in [x_i, x_{i+1}), \tag{38}$$

if we define

$$K_i = -s_i\phi(x_i)^2\left(1 - \frac{2}{3}\phi(x_i)\right) - (\partial_x\phi(x_i))^2. \tag{39}$$

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 $V(\phi) \le 0$. (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

$$F_i(\phi) = \int_{\phi_i^*}^{\phi} \frac{d\psi}{\sqrt{-V_i(\psi)}}, \tag{40}$$

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.

B.3 Doing the integrals

An important ingredient in the above method is the integral (40), 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 $k \to 1$, $\mathrm{sn}(x; k) \to \sinh(x)$.

C Hitting and occupation

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,

$$u(x,S) = \mathbb{E}\left[\int_0^\infty Z_t(S)dt\right] \tag{41}$$

$$= \int_0^\infty \mathbb{E}[Z_t]p_t(x,S)dt \tag{42}$$

$$= \int_0^\infty e^{-s_m t}p_t(x,S)dt \tag{43}$$

$$= \int_0^\infty \mathbb{P}^x\{\tau_\dagger > t \;\&\; B_t \in S\}dt \tag{44}$$

$$= \mathbb{E}^x\left[\int_0^{\tau_\dagger} \mathbf{1}_S(B_t)dt\right], \tag{45}$$

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 $B_{\min(t,\tau_\dagger)}$, by the strong Markov property this is equal to

$$u(x,S) = \mathbb{P}^x\{\tau_S < \tau_\dagger\}\mathbb{E}^{\mu_S(x)}\left[\int_0^{\tau_\dagger} \mathbf{1}_S(B_t)dt\right], \tag{46}$$

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 $h(x,S)$. (In the main body, 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

$$h(x,S) \approx \int_0^\infty (1-k_e(t))\mathbb{P}\{\tau_S \in dt\} \tag{47}$$

$$= \int_0^\infty e^{s_m t}(1-k_e(t))e^{-s_m t}\mathbb{P}\{\tau_S \in dt\} \tag{48}$$

$$= \int_0^\infty e^{s_m t}(1-k_e(t))\mathbb{P}\{\tau_S \in dt \,\&\, t < \tau_\dagger\} \tag{49}$$

$$\approx \frac{1}{\mathbb{E}[K]}\int_0^\infty \mathbb{P}\{\tau_S \in dt \,\&\, t < \tau_\dagger\} \tag{50}$$

$$= \frac{1}{\mathbb{E}[K]}\mathbb{P}^x[\tau_S < \tau_\dagger]. \tag{51}$$

Therefore, we have that

$$h(x,S)\mathbb{E}[K] \approx \mathbb{P}^x[\tau_S < \tau_\dagger] = u(x,S)/g(A). \tag{52}$$

If we let $q(S) = \int_S q(y)dy$ be the total equilibrium occupation of S, then by equations (9) and (10),

$$q(S) = \int_{\partial B(r_0)} \gamma(x)u(x,S)dx \tag{53}$$

$$\approx g(A)\mathbb{E}[K]\int_{\partial B(r_0)} \gamma(x)h(x,S)dx \tag{54}$$

$$= \frac{g(A)\mathbb{E}[K]}{\rho f(A)}\lambda_{\text{mig}}(y). \tag{55}$$

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 $d = 1$,

$$g(A) = \int_0^{2w} \frac{e^{-y\sqrt{2s_m}}}{\sqrt{2s_m}} dy \tag{56}$$

$$= (1 - e^{-2w\sqrt{2s_m}})/(2s_m) \tag{57}$$

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

$$g(A) = \int_0^{2w} 2yK_0(y\sqrt{2s_m})dy \tag{58}$$

$$= \frac{1}{s_m} \int_0^{2w\sqrt{2s_m}} yK_0(y)dy \tag{59}$$

$$= \frac{1}{2s_m} \left(1 - 2w\sqrt{2s_m}K_1(2w\sqrt{2s_m})\right). \tag{60}$$

In either case, $g(A) \approx 1/(2s_m)$, which is the approximation we use in the main text.

C.1 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 (9),

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

for any y. Take $y = (r, 0)$. Note that after scaling by $\sqrt{2s_m}/\sigma$,

$$\frac{u(x, y)}{q(y)} \simeq C' \frac{\|y - x\|^{-(d-1)/2} e^{-\|y-x\|}}{\|y\|^{-(d-1)/2} e^{-\|y\|}} \tag{61}$$

$$= C' \|\hat{x} - x/r\|^{-(d-1)/2} e^{-r(\|\hat{x}-x/r\|-1)}. \tag{62}$$

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

$$r(\|\hat{x} - x/r\| - 1) = r\left(\sqrt{\left(1 - \frac{r_0}{r}\cos\theta\right)^2 + \frac{r_0^2}{r^2}\sin^2\theta} - 1\right) \tag{63}$$

$$= -2r_0 \cos\theta + O(1/r). \tag{64}$$

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

$$\frac{u(x, y)}{q(y)} \longrightarrow_{r\to\infty} C' e^{-2r_0 \cos\theta}. \tag{65}$$

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

$$\int_{B(r_0)} \frac{u(x, y)}{q(y)} dx \to \gamma(r_0) \quad \text{as } y \to \infty \tag{66}$$

$$= C' \int_0^{2\pi} e^{-2r_0 \cos\theta} d\theta \tag{67}$$

$$= 2\pi C' I_0(-2r_0), \tag{68}$$

where I_0 is a modified Bessel function of order zero of the first kind. The effect of the scaling is to express r_0 in natural units, i.e. changing the result to $2\pi C' I_0(-2\sqrt{2s_m}r_0/\sigma)$.