Determining selection across heterogeneous landscapes: a perturbation-based method and its application to modeling evolution in space

Jonas Wickman[1], Sebastian Diehl[2], Bernd Blasius[3], Christopher A. Klausmeier[4], Alexey B. Ryabov[3] and Åke Brännström[1,5]

[1]Integrated Science Lab, Department of Mathematics and Mathematical Statistics, Umeå University, SE-90187, Umeå, Sweden
[2]Integrated Science Lab, Department of Ecology and Environmental Science, Umeå University, SE-90187, Umeå, Sweden
[3]Institute for Chemistry and Biology of the Marine Environment, Carl-v-Ossietzky University Oldenburg, Carl-von-Ossietzky-Straße 9-11, D-26111 Oldenburg, Germany
[4]W. K. Kellogg Biological Station and Department of Plant Biology, Michigan State University, Hickory Corners, Michigan 49060, USA
[5]Evolution and Ecology Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, 2361 Laxenburg, Austria

Preprint. Accepted for pulication in The American Naturalist.

Abstract

Spatial structure can decisively influence the way evolutionary processes unfold. Several methods have thus far been used to study evolution in spatial systems, including population genetics, quantitative genetics, moment-closure approximations, and individual-based models. Here we extend the study of spatial evolutionary dynamics to eco-evolutionary models based on reaction-diffusion equations and adaptive dynamics. Specifically, we derive expressions for the strength of directional and stabilizing/disruptive selection that apply in both continuous space and to metacommunities with symmetrical dispersal between patches. For directional selection on a quantitative trait, this yields a way to integrate local directional selection across space and determine whether the trait value will increase or decrease. The robustness of this prediction is validated against quantitative genetics. For stabilizing/disruptive selection, we show that spatial heterogeneity always contributes to disruptive selection and hence always promotes evolutionary branching. The expression for directional selection is numerically very efficient, and hence lends itself to simulation studies of evolutionary community assembly. We illustrate the application and utility of the expressions for this purpose with two examples of the evolution of resource utilization. Finally, we outline the domain of applicability of reaction-diffusion equations as a modeling framework and discuss their limitations.

Introduction

Heterogeneous landscapes provide both spatial variation in selective regimes and opportunities for geographic isolation. The notion that spatial heterogeneity should promote evolutionary diversification is therefore intuitively appealing. Although this intuition is sometimes foiled (Day 2001; Ajar 2003), several individual-based simulation and population genetics studies do indeed suggest that spatial structure can contribute to evolutionary branching and speciation (Felsenstein 1976; Doebeli and Dieckmann 2003; Servedio and Gavrilets 2004; Gavrilets and Vose 2005; Haller et al. 2013). However, understanding the interactions between spatial structure, population dynamics, and evolutionary dynamics has proved challenging. It is well recognized that the direction and magnitude of selection experienced by a population in a heterogeneous landscape depends both on the spatial pattern of local selection pressures and on the movement rates of individuals and genes across the landscape (Thompson 1999; Blondel et al. 2006). It is far from obvious under which circumstances this interplay of local selection, resulting from local ecological dynamics, and homogenizing dispersal results in directional selection, evolutionary stasis, or evolutionary diversification. It has, for example, been conjectured that mild geographical population structure may – paradoxically – be critical to the maintenance of evolutionary stasis at the species level over longer periods of time (Eldredge et al. 2005), in spite of widespread directional local selection (Kingsolver and Diamond 2011). Clearly, there is a need for a theory of evolution in space that can describe how spatially integrated selection in heterogeneous environments is driven by the interplay of local ecological dynamics and dispersal.

The development of such a theory has proceeded along several venues. The perhaps earliest attempts trace back to diallelic one-locus models exploring the invasion of beneficial alleles and the evolution of polymorphisms in a continuous one-dimensional habitat (Fisher 1937; Haldane 1948; Fisher 1950; Slatkin 1973). Several extensions of these models have been made to include more than one locus (Slatkin 1975; Barton 1983), continuous polygenic characters (Slatkin 1978; Barton 1999), or density dependence and multidimensional space (Nagylaki 1975) among other examples. However, in these studies, the ecological dynamics are typically very simple and selection pressures are prescribed rather than derived from density- and frequency-dependent interactions among phenotypes.

Conversely, models encompassing more realistic ecological interactions have typically neglected the underlying genetics altogether in favor of studying the evolution of phenotypes directly. A frequently employed approach has been to use spatially explicit, individual-based models that compute trait-dependent reproduction and inheritance directly based on various rules (Doebeli and Dieckmann 2003; Gavrilets and Vose 2005; Mágori et al. 2005; Birand et al. 2012; Haller et al. 2013; Kubisch et al. 2014). These models exhibit potentially great ecological realism and have broad applicability but suffer from two limitations: they are computationally very demanding and the results

derived from these models are often difficult to analyze and check for robustness (but see the discussion section on moment-closure approximations for further description and references).

Another approach to the inclusion of more realistic ecological interactions is to extend ecological models based on reaction-diffusion equations (a specific class of partial differential equations; see e.g., Britton et al. 1986; Holmes et al. 1994; Cantrell and Cosner 2004) to an evolutionary setting. Eco-evolutionary processes can then be studied using either a quantitative genetics (Lande 1979; Lande and Arnold 1983) or an adaptive-dynamics (Metz et al. 1992; Dieckmann and Law 1996; Geritz et al. 1998) framework. The former accounts for standing variation in trait values and can be used to answer questions about the rate of evolutionary change and the degree of spatial variation of trait values within species caused by limited gene flow. These questions have already been touched upon by Kirkpatrick and Barton (1997), Case and Taper (2000) and Norberg et al. (2012) among other studies. Adaptive dynamics, in turn, is concerned with mutation-driven evolution and is typically used to determine population-level selection gradients and phenomena such as evolutionary branching. These questions have received very little attention in the setting of spatial adaptive dynamics (but see e.g., Mizera and Meszéna 2003; Troost et al. 2005). In particular, general mathematical expressions describing directional as well as stabilizing or disruptive selection have not yet been developed for reaction-diffusion models in continuous space. This is unfortunate, as some questions, such as studying the direction of evolution of a trait in a spatially heterogeneous environment or how this spatial heterogeneity affects the potential for disruptive selection and subsequent diversification can be easier to address in the reaction-diffusion adaptive-dynamics framework than with the aforementioned approaches.

In this paper we develop central parts of an adaptive-dynamics theory of evolution in space for populations whose ecological dynamics are well described by reaction-diffusion equations, or certain metapopulation models. Specifically, we derive expressions for directional and stabilizing/disruptive selection acting on a quantitative phenotypic trait in organisms inhabiting a spatially heterogeneous landscape. These expressions are exact in the sense that they introduce no further approximations beyond the conceptual ones that underly the use of adaptive dynamics and reaction-diffusion equations. Our effort serves two primary purposes: First, it enables us to draw general conclusions about the evolution of traits in such systems. Specifically, we determine how local selection pressures should be used in a weighted average across space to ascertain the population-level direction of selection of quantitative traits and show how this result can be understood in terms of reproductive value. We test that this result is consistent with a quantitative genetics model in the limit of small standing genetic variation. We also show that spatial structure always contributes to disruptive selection in the absence of directional selection. Second, the derived expressions enable efficient numerical investigation of spatial evolutionary dynamics in eco-evolutionary systems described by reaction-diffusion equations. We illustrate the application and utility of the developed

techniques in two examples of the evolution of resource utilization in a heterogeneous environment. Finally, we discuss the limitations of reaction-diffusion equations as a modeling framework and delineate the range of eco-evolutionary systems over which they are applicable.

Evolutionary dynamics in spatially structured populations

Imagine a population inhabiting a heterogeneous landscape over which selection pressures for a quantitative trait, such as the size of a body part, varies. In certain regions, selection favors larger trait values while in others it favors smaller trait values than the current population average. Will selection result in an increase or a decrease of the mean trait value across the population? The intuitive approach of simply averaging selection across all members of the populations gives the wrong result. Below we show that one needs to take a particular weighted average with a disproportionate contribution from highly populated areas. We demonstrate that the expression for directional selection is valid in the frameworks of both adaptive dynamics and quantitative genetics, provided that standing genetic variation is small.

Invasion fitness in spatially structured populations

Before we start our analysis of selection in reaction-diffusion equations, we present a brief summary of adaptive-dynamics theory (Metz et al. 1992; Dieckmann and Law 1996; Geritz et al. 1998), which is a framework for studying evolutionary processes by calculating the so-called 'invasion fitness' of a rare mutant phenotype in a community of one or more resident phenotype(s) at equilibrium. This invasion fitness is the long term per capita growth rate of the mutant while it is still rare compared to the resident(s) and is thus a measure of the mutant's ability to cope with the environment set by the resident(s) while its own influence on the environment is still negligible. Write A_r and A_m for the densities of a resident and a mutant phenotype respectively, and let χ_r and χ_m be traits that uniquely characterize each of the two phenotypes (we shall throughout the text use subscripts r and m for quantities pertaining to residents and mutants, respectively). In a spatially unstructured system the growth of the resident population is governed by:

$$\frac{\mathrm{d}A_\mathrm{r}}{\mathrm{d}t} = G(E_\mathrm{r}, \chi_\mathrm{r})A_\mathrm{r}, \tag{1}$$

where the net per capita growth rate G depends on the environment set by the resident, $E_\mathrm{r} = E(\chi_\mathrm{r})$, and on the resident's trait value. When the resident is at equilibrium, with density A_r^* setting an equilibrium environment E_r^*, and the mutant is still rare enough to not affect the environment, the mutant's growth rate can be expressed as:

$$\frac{\mathrm{d}A_\mathrm{m}}{\mathrm{d}t} = G(E_\mathrm{r}^*, \chi_\mathrm{m})A_\mathrm{m}. \tag{2}$$

The above equation is a linear, first order, ordinary differential equation. The solution is

$$A_{\mathrm{m}}(t) = c\, \exp(\, G(E_{\mathrm{r}}^*, \chi_{\mathrm{m}})\, t\,), \tag{3}$$

where c is a constant representing the initial density of the mutant. This solution describes the invasion well as long as the mutant remains rare relative to the resident. From this, we see that the mutant density will increase if G is positive, and decline if G is negative. Thus $G(E_{\mathrm{r}}^*, \chi_{\mathrm{m}})$, the exponential growth rate of the mutant while still rare, is considered the invasion fitness of the mutant. For a given resident trait, we can think of G as being a function of the trait value χ and calculate the directional selection acting on a resident phenotype by computing how invasion fitness changes with changes in the trait value. In adaptive dynamics the selection gradient, defined as

$$D(\chi_{\mathrm{r}}) = \left.\frac{\partial G(E_{\mathrm{r}}^*, \chi)}{\partial \chi}\right|_{\chi=\chi_{\mathrm{r}}}, \tag{4}$$

is the measure of the strength and direction of this directional selection.

In a spatially heterogeneous system, the density of a spatially structured population depends on both time and space, and can be denoted by $A(t, \mathbf{x})$, where $\mathbf{x} = (x_1, ..., x_n)$ describes the location in $n = 1$, 2, or 3 spatial dimensions. We assume that the ecological dynamics can be described by a partial differential equation

$$\frac{\partial A(t, \mathbf{x})}{\partial t} = F(E(\chi), \mathbf{x}, \chi)A. \tag{5}$$

Here, the local rate of change is described by a differential operator F, which typically depends not only on the local population density $A(\mathbf{x})$, the spatial coordinate $\mathbf{x}$, and the trait under selection, χ, but also contain spatial derivatives describing passive transport or active movement in space. One important class of such systems are reaction-diffusion systems, where the differential operator takes the form:

$$F(E(\chi), \mathbf{x}, \chi)A = G(E(\chi), \mathbf{x}, \chi)A + d(\chi)\Delta A. \tag{6}$$

Here the function G describes the density-dependent net growth of the phenotypes at different points in space (mediated through their effect on the environment E), d is the diffusivity, and $\Delta A = \frac{\partial^2 A}{\partial x_1^2} + ... + \frac{\partial^2 A}{\partial x_n^2}$ is the Laplacian of A, which describes the random movement or transport of individuals from more to less densely populated locations in all n directions. When we derive in the next section an expression for the selection gradient in heterogeneous space we initially consider only such random (diffusive) movement. Further down, in Example 2, we extend the approach to include constant directional movement (advection).

Much like the non-spatial case, considering the case of a mutant that is initially very rare results in a linear partial differential equation describing the mutant's population growth:

$$\frac{\partial A_{\mathrm{m}}}{\partial t} = F(E_{\mathrm{r}}^*, \mathbf{x}, \chi_{\mathrm{m}})A_{\mathrm{m}}. \tag{7}$$

This equation cannot be solved in the same way as the ordinary differential equation of the non-spatial case, since F contains partial spatial derivatives. There is, however, by the theory of linear differential equations, an expression for the solutions,

$$A_{\mathrm{m}}(\mathbf{x}, t) = \sum_{i=0}^{\infty} c_i e^{\lambda_i t} A_i(\mathbf{x}). \tag{8}$$

Each term in the above sum is a different possible solution to Eq. 7, where each solution assumes a specific (temporally constant) shape of the spatial density distribution of individuals $A_i(\mathbf{x})$, while the total density of individuals is initially determined by c_i and subsequently increases or decreases over time at an exponential rate determined by the exponent λ_i. The functions $A_i(\mathbf{x})$ and numbers λ_i are the so called eigenfunctions and eigenvalues of the operator F, and solve the equation $FA_i(\mathbf{x}) = \lambda_i A_i(\mathbf{x})$. The general solution, Eq. 8, emerges from the combined contributions of these specific solutions. Eventually, the term with the largest λ_i will dominate and determine the long-term exponential growth rate of the mutant population. This quantity, known as the dominant eigenvalue, is written λ_{d} and is the natural generalization of invasion fitness for a mutant in a spatially structured system.

A perturbation-based method for calculating selection across heterogeneous space

Having introduced the concept of invasion fitness for spatially structured populations, we next derive expressions for disruptive/stabilizing selection in populations that follow reaction-diffusion dynamics. The approach rests on the assumption in adaptive dynamics that single mutations have small phenotypic effects. Hence, the selection gradient is essentially describing first order, or weak, selection. This is not in itself a new approach (especially for matrix models, see e.g., Van Baalen and Rand 1998; Caswell 2001; Rousset 2004), but by using perturbation theory for operators describing reaction-diffusion dynamics in systems with absorptive, reflective, or periodic boundaries, we are able to compute an expression for the selection gradient that does not involve any explicit computations of eigenvalues or vectors. This has two major benefits. First, it allows us to evaluate selection based only on the ecological equilibrium dynamics, which means that some immediate, general conclusions can be drawn regarding the direction in which a trait will evolve by averaging local selection gradients. Second, it provides great savings in terms of numerical computations of selection gradients when simulating discretized versions of evolutionary reaction-diffusion systems. This approach likewise yields some new insight into how environmental heterogeneity contributes to disruptive selection, and how diversification is affected by this additional disruptive selection.

Perturbation theory can be used to determine directional selection. If an unknown differential operator can be expressed as the sum of a known operator plus some small disturbance, it is often possible to calculate approximations of the unknown operator.

For instance, we know that the eigenvalue of the operator F of Eq. 7 must be zero when it is operating on the resident phenotype at ecological equilibrium, since the equilibrium density by definition will neither grow nor decline. As such Eq. 7 will be

$$0 = \frac{\partial A_{\mathrm{r}}^*}{\partial t} = F A_{\mathrm{r}}^*. \tag{9}$$

In a reaction-diffusion system like Eq. 6, knowing that the operator acting on the equilibrium density of the resident has a dominant eigenvalue of zero, we can perform perturbation calculations to derive an expression for the selection gradient $D(\chi_{\mathrm{r}})$ for a resident phenotype in ecological equilibrium (see Appendix A for details):

$$D(\chi_{\mathrm{r}}) = \frac{1}{\int A_{\mathrm{r}}^{*2} \mathrm{d}\mathbf{x}} \left(\int A_{\mathrm{r}}^{*2} \left. \frac{\partial G(E_{\mathrm{r}}^*, \mathbf{x}, \chi)}{\partial \chi} \right|_{\chi_{\mathrm{r}}} \mathrm{d}\mathbf{x} - \left. \frac{\partial d(\chi)}{\partial \chi} \right|_{\chi_{\mathrm{r}}} \int |\nabla A_{\mathrm{r}}^*|^2 \mathrm{d}\mathbf{x} \right). \tag{10}$$

Eq. 10 describes the selection gradient for a resident phenotype with equilibrium density A_{r}^*, in the equilibrium environment E_{r}^* set by the resident. $\partial G(E_{\mathrm{r}}^*, \mathbf{x}, \chi)/\partial \chi|_{\chi_{\mathrm{r}}}$ describes how per capita growth changes, and $\partial d(\chi)/\partial \chi|_{\chi_{\mathrm{r}}}$ how diffusivity changes with changing trait. The integral is over the entire available space. The gradient ∇A_{r}^* is a vector describing the slope of the resident's density distribution in each spatial direction, and $|\nabla A_{\mathrm{r}}^*|$, the Euclidian vector norm of ∇A_{r}^*, is the maximum rate of change, which is always in the direction of the gradient.

From Eq. 10 we can draw general conclusions about directional selection in spatial systems. First, if diffusivity is trait independent i.e., if $\partial d/\partial \chi = 0$, then Eq. 10 simplifies to

$$D(\chi_{\mathrm{r}}) = \frac{1}{\int A_{\mathrm{r}}^{*2} \mathrm{d}\mathbf{x}} \int A_{\mathrm{r}}^{*2} \left. \frac{\partial G(E_{\mathrm{r}}^*, \mathbf{x}, \chi)}{\partial \chi} \right|_{\chi_{\mathrm{r}}} \mathrm{d}\mathbf{x}. \tag{11}$$

In the complete absence of diffusion $\partial G(E_{\mathrm{r}}^*, \mathbf{x}, \chi)/\partial \chi|_{\chi_{\mathrm{r}}}$ is the local selection gradient at each point in space. The selection gradient acting on the population of the resident phenotype as a whole (described by Eq. 11) can thus be interpreted as a weighted average of local selection gradients. The weights are A_{r}^{*2}, meaning that the contributions to the selection gradient at points in space where the resident is abundant are disproportionately stronger than where it is rare. This squared term can be understood by analogy to a classical result for sensitivity analysis of matrix models, where the sensitivity of growth rate depends on both the right and left eigenvectors, corresponding respectively to the stable population distribution and the individual reproductive value (Caswell 2001). In the reaction-diffusion equations we investigate, these right and left eigenvectors are identical due to the spatial symmetry of dispersal. Each is given by $A^*(\mathbf{x})$, resulting in the squared weighting term in Eq. 11 (Appendix A).

Second, if instead the trait does not affect local fitness but only the dispersal rate, i.e., if $\partial G/\partial \chi = 0$, Eq. 10 simplifies to:

$$D(\chi_{\mathrm{r}}) = -\frac{1}{\int A_{\mathrm{r}}^{*2} \mathrm{d}\mathbf{x}} \left. \frac{\partial d(\chi)}{\partial \chi} \right|_{\chi_{\mathrm{r}}} \int |\nabla A_{\mathrm{r}}^*|^2 \mathrm{d}\mathbf{x}. \tag{12}$$

Two conclusions directly follow. First, since $|\nabla A^*_{\mathrm{r}}|$ is the local slope of the resident's density distribution, selection on diffusivity depends on the degree to which the resident is heterogeneously distributed in the landscape. Hence, if the distribution of the resident is completely homogeneous, then the selection gradient will be zero. Second, since both $\int A^{*2}_{\mathrm{r}} \mathrm{d}\mathbf{x}$ and $\int |\nabla A^*_{\mathrm{r}}|^2 \mathrm{d}\mathbf{x}$ are always positive, selection will always be for lower diffusivity, which is in line with earlier investigations by Hastings (1983) and Dockery et al. (1998).

Eq. 10 is valid for reaction-diffusion equations with periodic, absorptive, and reflective boundary conditions, or some combination thereof. In Appendices A and E, this is extended to a larger class of systems.

Sympatric and parapatric sources of disruptive selection can be distinguished. When directional selection ceases and the selection gradient is zero, the resident is at what is known as an evolutionarily singular point. Here, the second derivative of invasion fitness with respect to a mutant's trait value must be determined to evaluate whether the evolutionarily singular point is a fitness maximum (evolutionarily stable strategy), at which no more evolution will occur, or a fitness minimum, at which a phenotype will split into two. A full account of selection at an evolutionarily singular strategy is given in Appendix B. Here, we present the three most important conclusions that arise from this analysis:

First, all other things equal, spatial heterogeneity in local selection regimes always promotes evolutionary branching and diversification. The reason is that the possibility for the mutant to assume other spatial distributions than the resident's always contributes positively to the second derivative of its invasion fitness (Eq. B.5).

Second, by splitting the equation for disruptive/stabilizing selection (Eq. B.5) into two terms, two cases of evolutionary branching can be distinguished which we with a slight abuse of terminology call sympatric and parapatric diversification. In the first case, when the primary source of disruptive selection is the spatially averaged ecological dynamics, a phenotype can branch into two phenotypes with exactly the same spatial distribution as their progenitor. In the second case, when the only source of disruptive selection is the environmental heterogeneity, i.e., spatial variation in directional selection, at least one of the new phenotypes must have a different spatial distribution than the progenitor. The second type of branching may occur even if the spatially averaged dynamics contribute stabilizing selection, such as in Example 1 below.

Third, if the diffusion rate is not under selection we can in many situations (see Appendix B for details) estimate an upper limit to the amount of disruptive selection (i.e., the second derivative of the invasion fitness with respect to the mutant trait) that can be contributed by environmental heterogeneity as:

$$2K \frac{\dfrac{1}{\int A^{*2} \mathrm{d}\mathbf{x}} \int A^{*2} \left(\left. \dfrac{\partial G}{\partial \chi} \right|_{\chi=\chi_{\mathrm{r}}} \right)^2 \mathrm{d}\mathbf{x}}{d}. \tag{13}$$

Here, K is a constant that depends only on the shape and size of the landscape. The numerator can be interpreted as the weighted variance of local selection gradients, in the same spirit as in Eq. 11. This means that the maximal amount of disruptive selection coming from environmental heterogeneity for any given landscape depends on the ratio between the overall variability in local directional selection regimes and diffusion rate d.

The results concerning directional selection are also valid in a spatial quantitative genetics setting in the limit of low standing variation. In a non-spatial setting, it has been noted that there are some strong similarities between adaptive dynamics and quantitative genetics (Waxman and Gavrilets 2005). We investigate to what extent our conclusions about directional selection derived in the setting of adaptive dynamics hold true also in a spatial quantitative genetics setting by using a model developed by Kirkpatrick and Barton (1997), which describes trait evolution in a population with normally distributed trait values at each point in space. We conclude that, in the limit of small standing genetic variation, the quantitative genetics model reduces to a reaction-diffusion equation for the ecological dynamics, and that trait evolution becomes completely determined by the same expression for directional selection that we derived in the adaptive-dynamics context. The details can be found in Appendix C.

Application of the new method to evolutionary community assembly in continuous space and in metacommunities – a general recipe and two examples

An important application of the new method for calculating selection across heterogeneous space is to the study of evolutionary community assembly. Since Eq. 10 makes it possible to efficiently calculate the direction and magnitude of selection acting on a set of resident populations, methods such as setting the rate of change of a trait to be proportional to the selection gradient, can be used to investigate the coevolutionary dynamics of ecological communities. Many methods for studying evolutionary community assembly and for numerically implementing the equations of the previous section exist. Below, we describe one such combination of methods, which is applicable to both continuous space and to landscapes of discrete habitat patches occupied by metacommunities.

Evolutionary assembly in continuous space. For numerical implementation we discretize (continuous) space into a lattice. This reduces the partial differential equation (6) to a large set of coupled ordinary differential equations, the density distribution $A(\mathbf{x})$ to a vector of densities $\mathbf{A}$, and the differential operator F to a square matrix. The discretized equation system is then integrated to ecological equilibrium. The invasion fitness of a mutant in the system at a given ecological equilibrium is then found by computing the dominant eigenvalue of the matrix F, and an expression analogous to

Eq. 10 can be derived (see Appendix D).

The evolutionary assembly process is computed in three principal steps. First, during evolutionary community assembly selection will be directional for most of the time, and several methods can be used to compute the evolutionary response to directional selection. For instance, to produce Fig. 2A in Example 1 below we set the rate of change of a trait to be proportional to the selection gradient, i.e., $\mathrm{d}\chi_\mathrm{r}/\mathrm{d}t = \epsilon D(\chi_\mathrm{r})$, where ϵ is a small number separating ecological and evolutionary time scales. ϵ was chosen to be sufficiently small for the ecological dynamics to be very close to equilibrium at all times. This type of gradient dynamics is commonly used in computing the effects of directional selection, and has the same general form as e.g., the canonical equation of adaptive dynamics (Dieckmann and Law 1996; Champagnat 2003) or the Lande equation (Lande 1979). For Fig. 3 in Example 1 we were only interested in finding the eco-evolutionary equilibria, and hence we used the secant method (see e.g., Press et al. 2007) with a bound on step-size to find where the selection gradient for each resident was zero.

Second, when directional selection ceases and an evolutionarily singular point is reached, direct numerical computations of the dominant eigenvalue of the matrix F for trait values close to each resident are used to determine whether each resident phenotype is at an evolutionary branching point or an evolutionarily stable strategy (ESS). If any of the residents is at an evolutionary branching point, a portion of its density is split off to form a new phenotype with a trait value close to the one of the branching resident, and further co-evolution of the new ensemble is calculated as previously. Third, once all phenotypes are at an ESS, further trait values with positive invasion fitness may still exist. In such cases, a mutant with a trait value at the global maximum in the fitness landscape is allowed to invade the ensemble. As this requires large mutational steps, it is done by direct numerical calculation of the invasion fitness, i.e., the dominant eigenvalue of F. The described three steps of the evolutionary assembly process are then repeated until no more positive invasion fitness is available. An example of the entire process (Fig. 2A) and the resulting fitness landscape (Fig. 2B) is illustrated in Example 1 below.

Evolutionary assembly in metacommunities. Our new method for the calculation of selection across a heterogeneous, but continuous, landscape is easily adapted to the setting of an evolving metacommunity on a set of discrete patches, where the dynamics are governed by coupled ordinary differential equations. Since the just described numerical implementation of continuous space scenarios is technically equivalent to the metacommunity formalism, we do not treat the latter any further in the main text but refer instead to Appendix D. There we develop in detail a formalism for the study of community assembly in an evolving metacommunity, where both local growth rates in different patches and dispersal rates between patches may depend on a selected trait of the evolving phenotypes. Additionally, in Appendix C, the results concerning the spatial quantitative genetics model of the previous section are restated for a metapop-

ulation model on discrete patches.

In the following sections we illustrate the application of our perturbation-based method to the study of evolution in heterogeneous space with two examples of systems where consumers compete for two limiting resources along spatial gradients and experience an evolutionary trade-off in the efficiency with which they can use the two resources. Example 1 is crafted as an easy to understand, heuristic case study assuming a highly regular, 2-dimensional geometry of resource supply and exclusively random (diffusive) dispersal of consumers. Example 2 is a much more realistic case study of phytoplankton competing for nutrients and light along opposite vertical gradients and includes the additional complication of directed movement (which requires a variable transformation described in Appendix E). It also shows the relative ease with which an existing ecological study can be extended to include evolutionary dynamics. In both examples we focus on the most relevant and challenging issue in a spatial context, i.e., the influence of the speed of (diffusive or directed) movement across heterogeneous space on evolutionary outcomes. In addition, we evaluate the computational efficiency of the perturbation techniques. The reader interested in learning to apply the method may also consult Appendix F, where a discrete example simple enough to solve analytically is treated.

Example 1: Randomly dispersing consumers competing for resources with spatially variable supply rates

In our first example randomly dispersing consumers compete for two heterogeneously distributed resources on a square landscape (with spatial coordinates x and y). Consumers take up resources at the coordinate at which they reside and move randomly through the landscape as described by a diffusion process. The two resources are substitutable, renew locally at each coordinate, and do not disperse. As an example we envision plants consuming two substitutable nutrients (e.g., nitrate and ammonia) and dispersing seeds randomly, but the scenario is applicable to other systems with primarily locally renewed resources (e.g., randomly dispersing herbivores consuming substitutable plants). The dynamics at each point in space of the density $A_k(t,x,y)$ of consumer k, and of the two resource densities $R_1(t,x,y)$ and $R_2(t,x,y)$ are described by the equations:

$$\frac{\partial A_k}{\partial t} = G(R_1,R_2)A_k - mA_k + d\Delta A_k \tag{14a}$$

$$\frac{\mathrm{d}R_1}{\mathrm{d}t} = r_1(K_1 - R_1) - \sum_k G_{\max}\frac{a_{1k}R_1}{1 + a_{1k}R_1 + a_{2k}R_2}A_k \tag{14b}$$

$$\frac{\mathrm{d}R_2}{\mathrm{d}t} = r_2(K_2 - R_2) - \sum_k G_{\max}\frac{a_{2k}R_2}{1 + a_{1k}R_1 + a_{2k}R_2}A_k. \tag{14c}$$

Consumers take up resources 1 and 2 and convert them into growth according to a function derived from the concept of the 'synthesizing unit' (Kooijman 1998) at a maximal rate $G_{\max}$. For the case of two substitutable resources this yields

$$G(R_1, R_2) = G_{\max} \frac{a_{1k} R_1 + a_{2k} R_2}{1 + a_{1k} R_1 + a_{2k} R_2}, \tag{15}$$

where a_{1k} and a_{2k} are consumer k's affinities for resources 1 and 2. In the limiting case of only one resource being present this function reduces to the familiar Monod function (Monod 1949) with $1/a_{1k}$ or $1/a_{2k}$, respectively, being half-saturation constants.

In the absence of consumers, resources 1 and 2 follow semi-chemostat dynamics with turnover rates r_1 and r_2 and maximum resource densities at each point in space $K_1(x, y)$ and $K_2(x, y)$. Semi-chemostat dynamics are commonly used to describe the renewal of abiotic resources such as mineral nutrients (Tilman 1982) and ensure, for the non-spatial case of two resources, that the system settles to a locally stable ecological equilibrium. Though not guaranteed for the spatial case, we observe only stable equilibrium dynamics for the range of parameters explored in this example.

For ease of interpretation, we assume a highly regular geometry of the resource supply landscape. Specifically, total resource supply is constant ($K_1 + K_2 = 2.05$ at each point in space), while the ratio K_1/K_2 varies in space in the form of two crossed saddle-shapes as in Fig. 1. We assume that all consumers have identical (and fixed) maximum growth rates $G_{\max}$, mortality rates m, and diffusive dispersal rates d, but that their affinities to resources 1 and 2 can evolve within the constraints described below.

Consumers are characterized by a trait χ_k which describes a trade-off between affinities for the two resources as

$$a_{1k} = \frac{1}{1 + e^{-(a_0 + \chi_k)}}, \quad a_{2k} = \frac{1}{1 + e^{-(a_0 - \chi_k)}}, \tag{16}$$

pictured in Fig. 1B. Here a_0 is a trade-off control parameter that makes the trade-off strong for $a_0 < 0$, linear for $a_0 = 0$, and weak for $a_0 > 0$. An example of a weak trade-off is shown in Fig. 1B ($a_0 = 2.5$), which is the one we used in the numerical examples below (Figs. 2 and 3). The trade-off for the resource affinities is constructed in such a way that there exists a unique value of trait χ that maximizes consumer growth for each resource ratio when the trade-off is weak. Consumers with $\chi > 0$ are better at acquiring resource 1 and consumers with $\chi < 0$ are better at acquiring resource 2. Consumers with $\chi = 0$ are perfect generalists. The landscape has reflective boundary conditions

$$\left.\frac{\partial A_k}{\partial x}\right|_{x=-1,x=1} = 0, \quad \left.\frac{\partial A_k}{\partial y}\right|_{y=-1,y=1} = 0, \tag{17}$$

i.e., consumers cannot disperse into or out of the landscape. The parameter values used in simulations are given in Table 1.

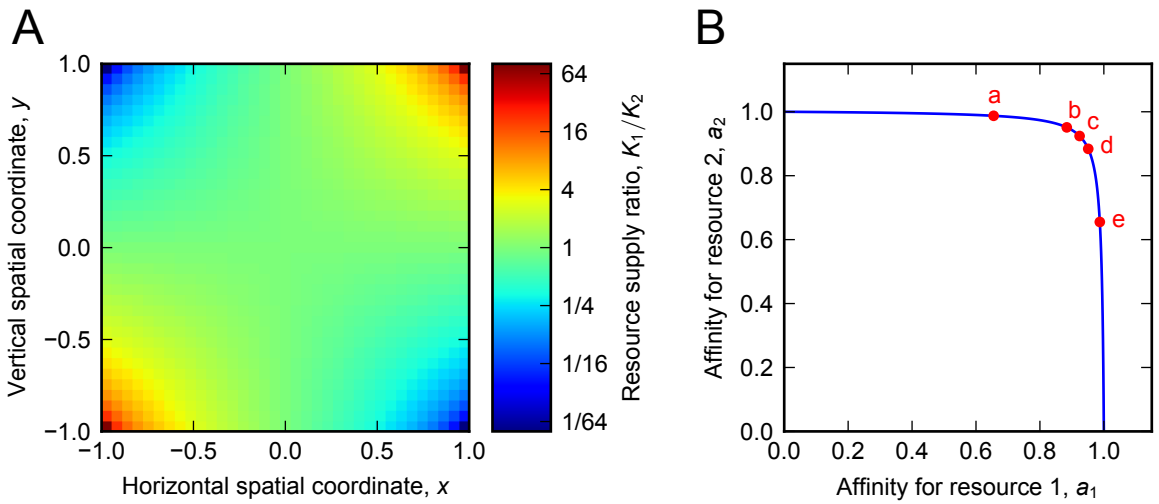

Figure 1: Example 1. (A) Resource supply landscape. The ratio K_1/K_2 of the maximum resource densities of the two resources is distributed as two crossed saddle-shapes in space. The sum of the maximum resource densities of the resources is constant so that $K_1 + K_2 = 2.05$. (B) The trade-off curve describing the relationship between consumers' resource affinities for the two resources. The concave down shape of the curve implies that the trade-off is weak and imposes a cost to specialization. Evolution of the trait χ is constrained to yield resource affinity combinations on the trade-off curve. The red dots indexed by a-e indicate the resource affinity of consumers a-e in Fig. 2.

Table 1: Parameters and state variables in Example 1.

Quantity	Definition	Value/range	unit
A_k	Density of consumer k		mass area^{-1}
$R_{1,2}$	Densities of resources 1 and 2		mass area^{-1}
m	Mortality rate	0.1	time^{-1}
d	Diffusion rate	$1.75 \cdot 10^{-6}$	area time^{-1}
$K_{1,2}$	Maximum resource densities of resources 1 and 2	See Fig. 1a	mass area^{-1}
$r_{1,2}$	Renewal rates of resources 1 and 2	1	time^{-1}
$a_{1k,2k}$	Resource affinity of consumer k for resources 1 and 2	(0,1)	area mass^{-1} time^{-1}
a_0	Trade-off strength control parameter	2.5	-
$G_{\max}$	Maximal consumer growth rate	1	time^{-1}
χ_k	Trait value for consumer k	$(-\infty, \infty)$	-
x	Horizontal spatial coordinate	(-1,1)	length
y	Vertical spatial coordinate	(-1,1)	length

To solve the ecological and evolutionary dynamics, we discretized the landscape to a 35×35 grid, and used the methods described above (*Evolutionary and assembly in continuous space*). Figure 2A shows an example of an evolutionary process for the resource landscape depicted in Fig. 1A, where we seeded the landscape with a single phenotype and ended up with a community of 11 different consumers. Once an eco-evolutionary equilibrium has been reached, different consumer phenotypes will have settled onto spatial distributions reflective of their degree of specialization on either resource (Fig. 2C), mirroring the distribution of the resource supply ratio (Fig. 1A).

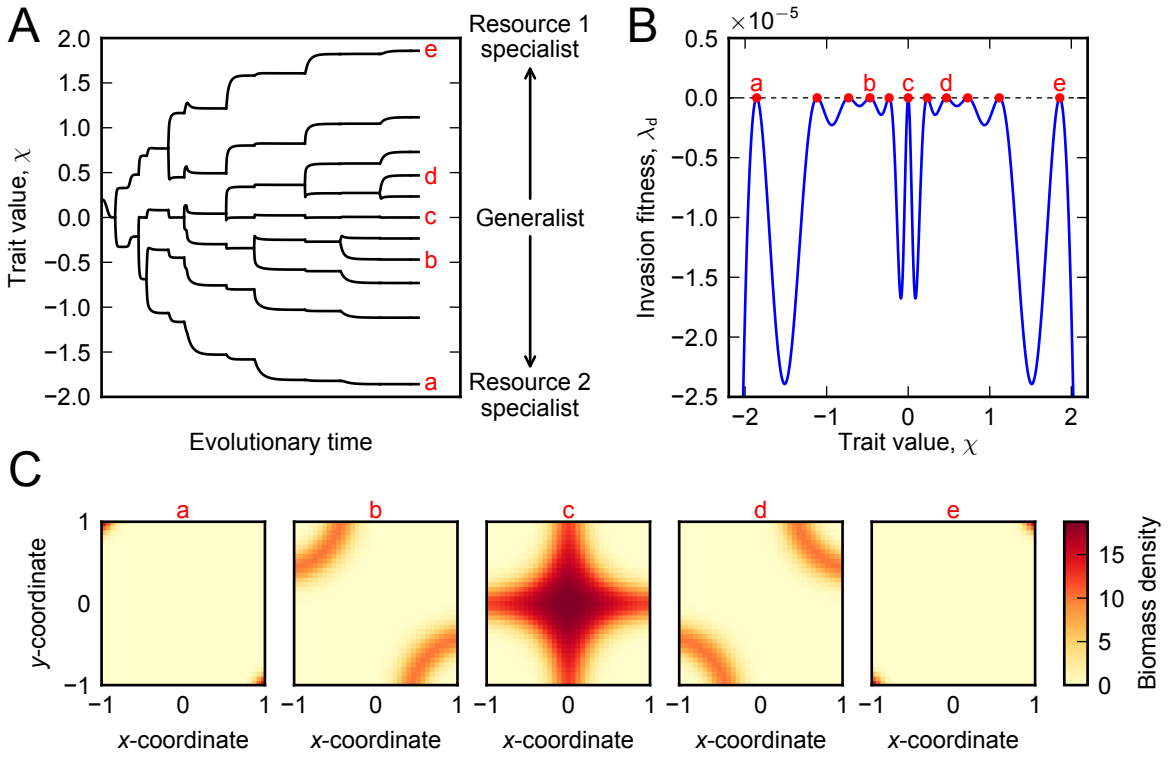

Figure 2: (A) An example of trait evolution on the resource supply rate landscape in Fig. 1, with parameters as in Table 1. Consumers continually evolve until an eco-evolutionary equilibrium is reached at the final time point. Letters a-e indicate which branch corresponds to which consumer density distribution in C. (B) Fitness landscape at the final point in time of A, showing that all consumers reside on local fitness maxima, with no further invasions possible. Red dots indicate resident consumers' trait values. (C) Equilibrium distribution in space of the five consumers indicated by letters a-e in panels A and B. Consumers with positive trait values are resource 1 specialists, consumers with negative trait values are specialized on resource 2, and consumers with trait value 0 (c) are generalists and have equal affinity for both resources. Darker shading indicates higher consumer density. Though pictured separately for clarity, there is significant spatial overlap among consumers, and several consumers can co-occur at the same spatial coordinate.

While the above findings are qualitatively intuitive, our approach greatly facilitates quantitative prediction of the exact number of evolving phenotypes, their trait values,

population sizes and spatial distributions. Note that quantitative prediction of these features is as easily achieved also when spatial variation in local selection is highly irregular, and thus precludes any intuition about even the qualitative nature of evolutionary outcomes. Quantitative prediction of evolutionary outcomes is only a simple task when the dispersal rate of consumers is either zero (allowing perfect local adaptation) or very high (preventing local adaptation altogether, see below). In the following we therefore explore in greater detail how these quantitative predictions depend on the consumers' rate of diffusion.

In a system that lacks spatial structure in resource supply, only one or two consumers can coexist when competing for two resources (sensu Tilman 1980). We observe the same phenomenon in the presence of spatial structure in resource supply when the diffusion rate of consumers is sufficiently high, because the rate of diffusion controls to what extent consumers experience environmental heterogeneity. Fig. 3 shows how the equilibrium distribution of trait values is affected by the rate of diffusion. When diffusion rates are sufficiently high all consumers more or less see only the average resource supply ratio, which is 1. This means that only a single generalist consumer will exist, due to the weak trade-off between resource affinities. We made a cut-off at a rate of diffusion yielding 11 consumers, since the computational complexity increases quite steeply with the number of consumers. In the limit of $d = 0$, one would expect to have one consumer for each ratio K_1/K_2, since there exists a unique χ optimizing growth for each ratio, and with no diffusion to propagate the consumers, each extant consumer will adapt to its local conditions.

To get a better understanding of the conditions that favor transition from a monomorphic to a polymorphic community we employ the methods detailed in Appendix B to determine the switchpoint between stabilizing and disruptive selection on a monomorphic consumer population for different diffusion rates. Due to the spatial symmetry of the supply ratios of resources R_1 and R_2 across the landscape, there is a spatially constant solution A_c for the density of the monomorphic consumer at the evolutionarily singular point $\chi_\mathrm{r} = 0$. We use this to analytically calculate an upper limit C for the amount of disruptive ($C > 0$) or stabilizing ($C < 0$) selection (i.e., the second derivative of the invasion fitness with respect to the mutant trait):

$$C \leq \frac{1}{\int A_c^2 \mathrm{d}\mathbf{x}} \int A_c^2 \left.\frac{\partial^2 G}{\partial \chi^2}\right|_{\chi=0} \mathrm{d}\mathbf{x} + \frac{1}{\int A_c^2 \mathrm{d}\mathbf{x}} \int A_c^2 \left(\left.\frac{\partial G}{\partial \chi}\right|_{\chi=0} \right)^2 \mathrm{d}\mathbf{x} \frac{L^2}{\pi^2 d}. \tag{18}$$

The first term is a weighted average of the curvatures of local selection, which here are downward concave (i.e., $\left.\frac{\partial^2 G}{\partial \chi^2}\right|_{\chi=0} < 0$) everywhere owing to the weak trade-off in resource-utilization. Due to the spatial symmetry in resource supply ratios this results in stabilizing selection across the entire landscape: average fitness is highest for $\chi_\mathrm{r} = 0$ because the local fitness gains of a phenotype with $\chi_\mathrm{r} \neq 0$ is outweighed by the inevitably higher local fitness losses in locations with the opposite resource supply ratios. The integral ratio in the second terms describes a weighted variance of local

directional selection. This ratio is multiplied by a measure of the size and shape of the landscape (here $L = 2$ is the length of each side of the square landscape) and is divided by the diffusion rate, d. As the average local selection is stabilizing (i.e. the first term in Eq. 18 is negative), evolutionary branching can only arise from a sufficiently large ratio between the variance in local selection regimes and the diffusion rate. In this specific system the solution for A_c and G does not depend on d, and we may thus set the above expression to zero to calculate a diffusion rate above which we know that selection will be stabilizing. This calculation yields $d = 3.86 \cdot 10^{-4}$, which agrees well with the value of d at which the system transitions from monomorphic to polymorhpic (Fig. 3).

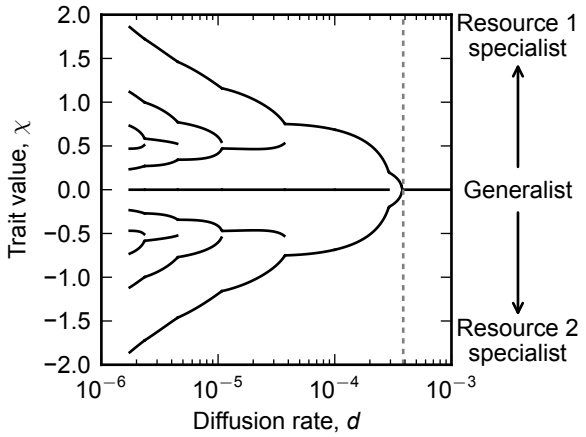

Figure 3: Example 1. Distribution of consumer trait values at evolutionary equilibrium for different rates of diffusion. When the rate of diffusion is high, a generalist can monopolize the entire space, whereas low rates leave room for many co-existing consumers. The vertical broken line indicates the upper limit for disruptive selection ($d = 3.86 \cdot 10^{-4}$), as calculated by Eq. 18 (see text). Parameters other than d are as in Table 1. See Appendix G for an account of the nature of the bifurcations, and why the diagram has gaps.

Example 2: Sinking algae in a water column

The previous example illustrated how phenotypic selection experienced by populations of randomly dispersing organisms depends both on the spatial pattern of local selection regimes and on the rate of dispersal across the landscape. In some systems additional aspects of movement have to be taken into account to understand evolutionary responses to spatially varying selection. For example, many organisms disperse directionally in fluid media (e.g., streams, coastal longshore currents, wind) or by gravity. Such systems can be modeled with reaction-advection-diffusion equations where, in addition to the diffusion term, an advection term describing directional movement is introduced (Speirs and Gurney 2001; Huisman et al. 2002; Anderson et al. 2005).

This directional movement presents a new type of problem. The necessary technical

condition allowing us to compute the expression for the selection gradient (Eq. 10) is that the differential operator describing the dynamics is what is known as "self-adjoint" (See Appendix A for details). Intuitively speaking, this condition, in the spatial case, means that the dispersal or transport rate from a coordinate x to a coordinate y is equal to that from y to x. In the discrete spatial case, this condition is equivalent to having a symmetric dispersal matrix. In the continuous case, this imposes conditions both on the form of the operator itself, as well as on the boundary conditions of the system. Advection breaks this symmetry, but for the case of constant advection this issue can be resolved by means of a variable-transform (Appendix E), which is what we have applied in the example below.

An example of a widespread ecological system described by reaction-advection-diffusion is sinking plankton algae growing in a pelagic environment. The spatial dynamics of this system are considerably more complex than those of Example 1 because, besides the added complication of directional algal movement, resources are also transported through space by either turbulent diffusion (dissolved nutrients) or directional flux (light). Such systems are typically characterized by opposing vertical resource gradients (with light availability decreasing and nutrient availability increasing with depth), setting up a smooth, continuous gradient of spatially varying selection for light vs. nutrient use capacities. As an illustration we use the so called "fixed stoichiometry" version of a phytoplankton model by Jäger et al. (2010), where nutrients from algae that have settled out of the water column are recycled in the bottom sediment. The equations that describe the ecological dynamics are reproduced in Appendix E together with the details of how the evolutionary dynamics were implemented.

The growth rate of the algae is proportional to the product $R/(M_j+R)\cdot I/(H_j+I)$ of two Monod functions describing the dependence of algal growth on available nutrients (R) and light (I). We introduce evolutionary dynamics to the system by letting the half-saturation constants of phenotype j for nutrient and light uptake M_j and H_j depend on a trait χ_j such that $M_j = M_0/\chi_j$ and $H_j = H_0\chi_j$. This implies a weak trade-off between the half-saturation constants, where phenotypes with a high value of χ are good nutrient competitors, and phenotypes with low χ are good light competitors (Fig. 4A).

To study evolutionary consequences of directional dispersal – in this case sinking – we considered a 50 m deep water column and let the system run to eco-evolutionary equilibrium (in the same way as in Example 1) for a range of algal sinking speeds, v. For the chosen parameter values, sinking speed not only affects the equilibrium distribution of phenotype trait values but also the number of co-existing phenotypes in a non-monotonic manner (Fig. 4B). With increasing sinking speed the phenotypes' depth distributions become centered at increasingly greater depths (Fig. 4C, D), and the overall biomass in the system generally decreases (Fig. 4E). For the investigated parameter range a strong light competitor, pictured in green in Fig. 4, tends to dominate the system.

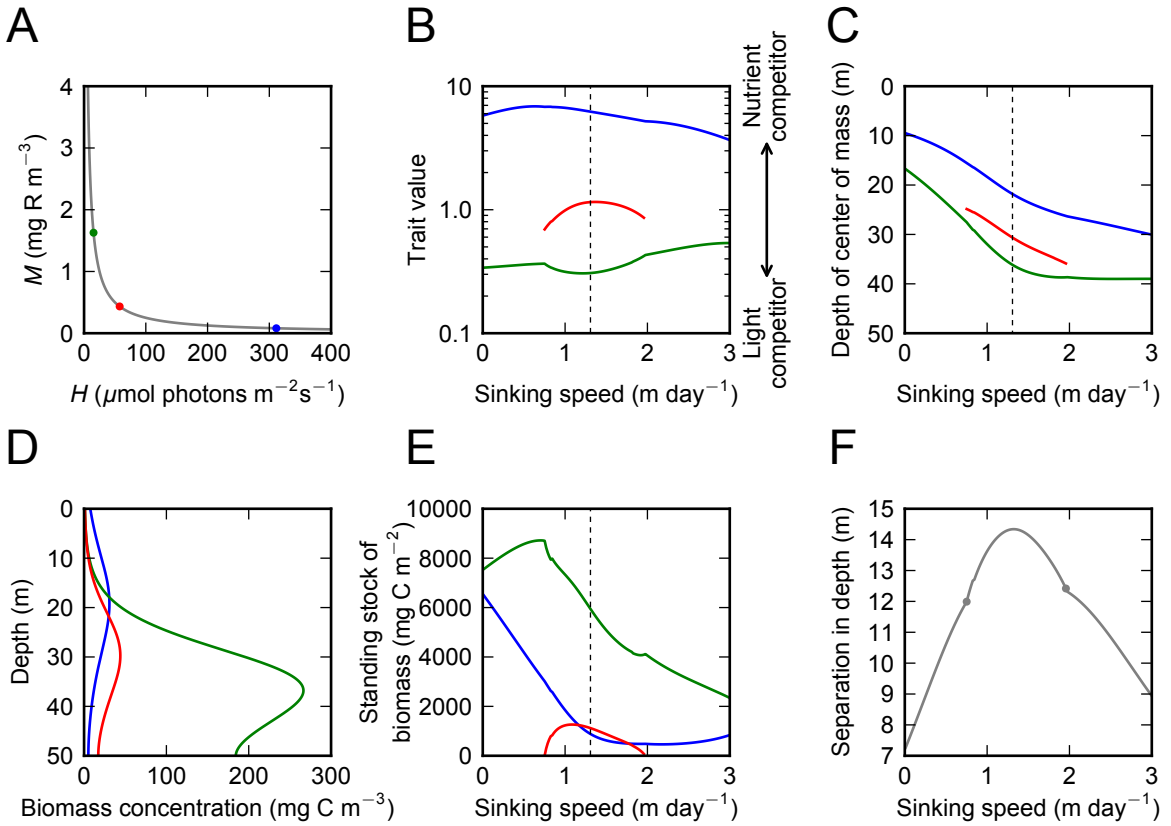

Figure 4: Example 2. (A) Trade-off between the half-saturation constants for light (H) and nutrient (M) uptake. The three dots show the combination of half-saturation constants for the three phenotypes at the sinking speed indicated by the broken line in B, C, and E. (B) Eco-evolutionary equilibrium distribution of trait values for different sinking speeds. Phenotypes with low trait values are good light competitors, and phenotypes with high trait values are good nutrient competitors. (C) The depth of the center of mass of the depth distributions of biomass for different sinking speeds. (D) Depth distribution of biomass concentration of the three phenotypes at the sinking speed indicated by the broken line in B, C, and E. (E) The standing stock of biomass of the algal phenotypes for different sinking speeds. The standing stock of a phenotype is its biomass concentration integrated over the entire water column. (F) Depth separation between the center of mass of the top and bottom phenotypes (blue and green in A-E). The dots delimit the range of intermediate sinking speeds over which the third (red) phenotype is part of the ensemble. In all panels, different colors represent different phenotypes with the color codes being consistent among panels.

It is far from obvious why intermediate sinking speeds generate the most diverse communities. One possible explanation could be that only at intermediate rates of sinking does the depth separation between the top and bottom phenotypes become large enough to admit an intermediate phenotype (see Fig. 4F). We limited our investigations to a realistic range of sinking speeds v for which we observed equilibrium dynamics, and for which the numerical approximation of the transformation of Eq. E.9 (Appendix E) was accurate at the implemented spatial resolution. When sinking speeds are increased further the system first moves into limit cycle dynamics until, at around $v = 4$ m day^{-1}, sinking losses become so large that no population is viable.

Computational benefits

Numerically, the main advantage of using the perturbation expression for calculating selection gradients, Eq. 10, is that it does not rely on any explicit computations of eigenvalues or eigenfunctions. If the expression weren't available, the selection gradient would have to be estimated numerically by calculating, for example, $(\lambda_{\mathrm{d}}(\chi_{\mathrm{r}}+\epsilon)-\lambda_{\mathrm{d}}(\chi_{\mathrm{r}}-\epsilon))/(2\epsilon)$, for some small number ϵ which would require the numerical calculation of the dominant eigenvalue of $F(\chi_{\mathrm{r}}+\epsilon)$ and $F(\chi_{\mathrm{r}}-\epsilon)$. Numerical calculation of eigenvalues of large matrices is however time-consuming, and hence using Eq. 10 can yield very significant computational time-savings. This is particularly true when setting the rate of change of a trait in time to be proportional to the selection gradient, in which case many evaluations of the selection gradient have to be made. As an example, we timed the calculations of the selection gradients for all residents in Fig. 3. The average time for numerical computation of the selection gradient using Matlab's (R2014a) 'eigs' function was 0.102 seconds, compared to $6.03 \cdot 10^{-5}$ for the perturbation calculation. This makes the perturbation calculation more than 1500 times faster. Using the above estimates of average computation times and (conservatively) assuming that the selection gradient has to be calculated only once per resident and time step, calculation of all selection gradients needed to produce Fig. 2A would have taken over 10 hours without the perturbation methods, but only roughly 21 seconds with them.

Discussion

Contributions to understanding selection in space

Apart from their use as efficient tools for modeling specific evolutionary scenarios, the expressions derived for directional and stabilizing/disruptive selection yield several novel, general insights. For example, when the trait under selection does not affect movement, Eq. 11 shows that the selection gradient in the spatial system can be understood as a weighted average of local selection gradients with weights proportional to the square of the resident population's equilibrium distribution. Eq. 11 can thus in

principle be used to predict the direction of evolution of a trait under selection in a heterogeneous environment. This would require accurate measurements of local population densities and of how per-capita growth rates change with trait values (i.e. of $\partial G/\partial\chi$). The latter is a challenging and labor intensive task in most natural systems. An alternative would be to investigate to what extent selection gradients measured using the approach of Lande and Arnold (1983) could serve as useful proxies for $\partial G/\partial\chi$, as there is already a wealth of such measurements (see e.g., Siepielski et al. 2013, and the references therein).

Eq. 11 makes furthermore a contribution to the unresolved issue of widespread local directional selection combined with evolutionary stasis (Merilä et al. 2001). Specifically, the expression shows that even if the population-level selection gradient is zero, as in evolutionary stasis, local measurements of the selection gradient will very likely be non-zero and point in different directions at different points in a heterogeneous environment. While this possibility has been conjectured previously, Eq. 11 lends mathematical support to it and could be used to ascertain to what degree space is responsible for the directional selection/stasis dichotomy in real populations.

Finally when it comes to disruptive selection in heterogeneous landscapes, the methods described herein (see Appendix B for details) can be used to determine whether the source of disruptive selection is a trade-off across the phenotype range in e.g., physiology or behavior, or the environmental heterogeneity prompting local adaptations. This was shown in Example 1, where the spatially averaged stabilizing trade-off in resource utilization could be overcome by disruptive selection contributed by variability in selection regimes provided that the rate of dispersal was sufficiently low. Depending on which type of disruptive selection is the primary contributor, one may forecast on a rough level the degree to which the branched phenotypes can be expected to have similar spatial extensions as their progenitor. Furthermore, if measurements of $\partial^2 G/\partial\chi^2$ could be obtained in addition to the local selection gradients, Eq. B.7 could be used to rule out disruptive selection, in the same way we analyzed Example 1.

Limitations of reaction-diffusion approaches

Reaction-diffusion models are not always appropriate for describing ecological dynamics. In particular, the discrete nature of individuals is not modeled, and the diffusion operator instantaneously propagates infinitesimally low densities across any region. This may lead to some pathological behavior, such as the somewhat infamous atto-fox (Mollison 1991), where 10^{-18} of an infected fox causes a re-invasion of rabies. Reaction-diffusion equations furthermore do not always replicate the limiting behavior of an underlying stochastic spatial process, as pointed out by Durrett and Levin (1994), although the authors remark that this issue can sometimes be alleviated by correcting the reaction terms by deriving them directly from the stochastic process. The shortcomings of reaction-diffusion equations thus practically invalidate their use for studying evolution

in settings where mutant-mutant interactions or limited movement are critical to the outcome. Examples include the evolution of cooperation, where spatial structure has been shown to allow local clusters of cooperators to invade a community of defectors (see e.g., Ferrière and Le Galliard 2001; Doebeli and Dieckmann 2003; Mágori et al. 2005; Lion and van Baalen 2008, for perspectives on these shorcomings). Before applying the framework developed in this paper, one should therefore carefully assess that neither the ecological nor the evolutionary processes of interest are critically dependent on the discrete nature of individuals or on limited dispersal rates of mutants.

Although there is no straightforward way of taking local mutant-mutant interactions into account in reaction-diffusion models, as in other frameworks (see next section), a possible way of modeling limited dispersal in reaction-diffusion systems is with non-linear diffusion terms. Using so-called slow diffusion, the spread of infinitesimal densities is no longer instantaneous. Such approaches have occasionally been used to model the ecology of populations, for example of microorganisms (Eberl et al. 2001; Tao and Winkler 2013) and insects (Turchin 1989). Using nonlinear diffusion to model spread implies that parts of a landscape may become inaccessible to some of the phenotypes. This in turn suggests that the location of invasion of a mutant would be critical to its success, as opposed to the linear diffusion models where the mutant spreads globally while still rare. An adaptive-dynamics framework for these types of non-linear diffusion equations does, however, not yet exist, and the formulation and analysis of such a framework would be a natural continuation of the material presented here.

Other analytical approaches to modeling spatial evolution

As the eco-evolutionary dynamics of spatially structured systems can be very complicated, some realism must inevitably be sacrificed to successfully derive analytical insights. Depending on the focus of interest, different approaches have been developed to handle specific phenomena, while neglecting other parts of ecological, evolutionary, or spatial dynamics for analytical feasibility.

For example, starting with Wright (1943) there has been a large body of population-genetics theory aiming at understanding genetic structure in finite, or locally finite, populations (see Rousset 2004, for a review). The framework considers dispersing individuals on a possibly infinite set of discrete patches, where the expected allele frequencies are tracked. These models have the advantage that they enable the study of both genetically explicit selection through a weak selection approximation, as well as the effects of genetic drift. Incorporating demographic dynamics into these models is, however, very complicated compared to phenotypic approaches such as adaptive dynamics or quantitative genetics, and most models simply consider population sizes to be constant (but see e.g., Rousset and Ronce 2004). Models of this type may also be used to investigate the evolution of helping behaviors (Rousset 2004; Lehmann et al. 2006), as they can take local mutant-mutant interactions into account.

Another approach to studying phenotypic evolution in spatially structured systems is to use adaptive-dynamics techniques coupled with moment-based approximations (Van Baalen and Rand 1998; Lion and van Baalen 2008; Lion 2015). Typically, the demographic dynamics of individual-based lattice models are approximated by deriving equations for the temporal change in the density of individuals (first moment) and of pairs (second moment) after which the moment hierarchy is 'closed' by replacing higher-order moments such as the density of triplets with expressions based on the lower-order moments. These methods have worked well for studying evolutionary processes in space where local mutant-mutant interactions are critical, such as the evolution of helping behaviors (Van Baalen and Rand 1998; Le Galliard et al. 2003; Ohtsuki et al. 2006; Lion and van Baalen 2009). A corresponding approximation method for populations in continuous space has been developed (Bolker and Pacala 1997; Law and Dieckmann 2000) and should be applicable to evolutionary studies (a somewhat different moment-based method has already been applied by North et al. 2011). Underlying these techniques is the implicit assumption that space, at least in a statistical sense, looks the same from the perspective of any focal individual. They therefore work best when the underlying environment is homogeneous, and are thus hard to apply to cases such as our Example 2.

Conclusions and outlook

While there will always be exceptions, we believe the following advice can be given for modeling phenotypic evolution of spatially structured populations: If the primary goal of the study is to understand how local individual interactions evolve or how evolution acts near to a mean-field limit, moment-based methods are appropriate. If the goal is to understand intraspecific variation in a trait throughout heterogeneous space, then reaction-diffusion quantitative-genetics models are best suited. Finally, if the goal is to study the evolution of a spatially constant trait in a heterogeneous environment, especially for finding evolutionarily stable communities or studying disruptive selection, then the methods detailed in this paper constitute the best alternative.

Selection in space continues to be a challenging problem, with no single theoretical framework striking the balance between tractability and insight on the one hand, and scope and realism on the other. We believe that this paper will provide the groundwork for using reaction-diffusion equations coupled with adaptive dynamics to answer questions about selection in space that were previously difficult to address.

Online Appendix A: Derivation of invasion criteria and selection gradient in spatially structured systems

In this appendix we derive expressions for the invasion fitness of a rare mutant, and the selection gradient of a resident for spatially structured systems described by reaction-diffusion equations. We extend the main result to all partial differential operators that are self-adjoint, and show the type of boundary conditions yielding self-adjoint operators for reaction-diffusion systems.

Self-adjoint operators and reaction-diffusion systems

Although the main text of this paper has been written specifically for reaction-diffusion equations with certain boundary conditions, the perturbation theoretic formulation of the selection gradient is valid for any self-adjoint operator. Formally, for a linear differential operator F to be self-adjoint it has to fulfill

$$\int A(\mathbf{x})(FB(\mathbf{x}))\mathrm{d}\mathbf{x} = \int (FA(\mathbf{x}))B(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{A.1}$$

for all smooth A and B with the same boundary conditions at the boundary. Whether the operator is self-adjoint may depend on both the operator itself and the boundary conditions of the space on which it operates. The operator of reaction-diffusion systems fulfill this criterion under boundary conditions of the form

$$a(\mathbf{s})A(\mathbf{s}) + b(\mathbf{s}) \left.\frac{\partial A(\mathbf{x})}{\partial \hat{n}}\right|_{\mathbf{x}=\mathbf{s}} = 0, \tag{A.2}$$

where $\mathbf{s}$ is a point on the boundary of the domain, and $\left.\frac{\partial A(\mathbf{x})}{\partial \hat{n}}\right|_{\mathbf{x}=\mathbf{s}}$ is the derivative on the boundary in the direction $\hat{n}$ pointing outwards from the boundary. $a(\mathbf{s})$ and $b(\mathbf{s})$ are some arbitrary smooth functions on the boundary that are not both zero at the same coordinate. This condition includes absorptive and reflective boundary conditions, as well as some more complicated ones like Eq. E.11b in Example 2.

Invasion criterion

In the main text we explained how the invasion fitness in a spatial system described by a partial differential equation could be computed by calculating the dominant eigenvalue of the differential operator describing the system dynamics. Here, we show how application of perturbation theory can be used in conjunction with this fact to produce expressions for the invasion fitness of rare a mutant with a trait value close to a resident.

Consider a linear differential operator F that is self-adjoint on a spatial domain Ω. Let this operator act on functions ψ and consider the eigenvalue problem $F\psi =$

$\lambda\psi$, where ψ is known as an eigenfunction of F if operating with F on ψ yields the same result as multiplication with a constant λ, known as an eigenvalue. Suppose this problem can be decomposed in such a way that

$$F = F_0 + F', \quad \psi = \psi_0 + \psi', \quad \lambda = \lambda_0 + \lambda', \tag{A.3}$$

where the eigenfunctions, ψ_0, and eigenvalues, λ_0, of the operator F_0 are known, and that F' is a small perturbation of F_0. Then, by perturbation theory, the first order approximation to λ, $\lambda^{(1)}$, is (see any quantum mechanics textbook e.g., Sakurai and Tuan 1985)

$$\lambda^{(1)} = \lambda_0 + \frac{\int_\Omega \psi_0 F' \psi_0 \mathrm{d}\Omega}{\int_\Omega \psi_0^2 \mathrm{d}\Omega} = \lambda_0 + \frac{\int_\Omega \psi_0 (F - F_0) \psi_0 \mathrm{d}\Omega}{\int_\Omega \psi_0^2 \mathrm{d}\Omega}. \tag{A.4}$$

In a reaction-diffusion system, let the time-evolution of the resident phenotype density $A_\mathrm{r}(\mathbf{x}, t)$ at coordinate $\mathbf{x}$ in a spatial domain Ω at time t be

$$\frac{\partial A_\mathrm{r}}{\partial t} = G_\mathrm{r}(E_\mathrm{r}, \mathbf{x}, \chi_\mathrm{r}) A_\mathrm{r} + d_\mathrm{r}(\chi_\mathrm{r}) \Delta A_\mathrm{r}, \tag{A.5}$$

with $G_\mathrm{r}(E_\mathrm{r}, \mathbf{x}, \chi_\mathrm{r})$ being the net per-capita growth rate of the resident phenotype for an environment E_r set by the resident, and $d(\chi_\mathrm{r})$ being a coefficient determining diffusion rate, for a phenotype with trait χ_r. For a resident at a stable equilibrium, with equilibrium density A_r^*, setting an equilibrium environment E_r^* the operator

$$F_\mathrm{r} \,:\, F_\mathrm{r} A_\mathrm{r}^* = (G_\mathrm{r}(E_\mathrm{r}^*, \mathbf{x}, \chi_\mathrm{r}) + d_\mathrm{r}(\chi_\mathrm{r}) \Delta) A_\mathrm{r}^* \tag{A.6}$$

will have a an eigenvalue $\lambda_{\mathrm{d,r}} = 0$, with associated eigenfunction A_r^*, since per definition a population in equilibrium neither grows nor declines. Furthermore, since the equilibrium is stable, there cannot be any eigenvalues larger than 0, since by Eq. 8 this would imply that a small deviation from the equilibrium would lead to the solution growing away from it. This means that the dominant eigenvalue of F_r has to be $\lambda_{\mathrm{d,r}}$.

A rare mutant in the environment set by the resident population will follow the dynamics

$$\frac{\partial A_\mathrm{m}}{\partial t} = G_\mathrm{m}(E_\mathrm{r}^*, \chi_\mathrm{m}) A_\mathrm{m} + d_\mathrm{m}(\chi_\mathrm{m}) \Delta A_\mathrm{m}, \tag{A.7}$$

where G_m is assumed not to depend on A_m, with

$$F_\mathrm{m} \,:\, F_\mathrm{m} A_\mathrm{m} = (G_\mathrm{m}(E_\mathrm{r}, \chi_\mathrm{m}) + d_\mathrm{m}(\chi_\mathrm{m}) \Delta) A_\mathrm{m}. \tag{A.8}$$

The difference of the operators is

$$F_\mathrm{m} - F_\mathrm{r} = (G_\mathrm{m} - G_\mathrm{r}) + (d_\mathrm{m} - d_\mathrm{r}) \Delta, \tag{A.9}$$

which we can identify as F' and insert into Eq. A.4 yielding a first-order perturbation expression for the dominant eigenvalue, which is the invasion fitness of a rare mutant,

$$\lambda_{\mathrm{d,m}}^{(1)} = \lambda_{\mathrm{d,r}} + \frac{\int_\Omega A_\mathrm{r}^* ((G_\mathrm{m} - G_\mathrm{r}) + (d_\mathrm{m} - d_\mathrm{r}) \Delta) A_\mathrm{r}^* \mathrm{d}\Omega}{\int_\Omega A_\mathrm{r}^{*2} \mathrm{d}\Omega}. \tag{A.10}$$

Assume further restrictions of the boundary conditions, so that if Γ is the boundary of Ω, then $\int_\Gamma A_\mathrm{r}^* \frac{\partial A_\mathrm{r}^*}{\partial \hat{n}} \mathrm{d}\Gamma = 0$, with $\partial/\partial\hat{n}$ being the derivative in the outwards direction $\hat{n}$ from the boundary. Under this assumption integration by parts can be used together with the fact that $\lambda_{\mathrm{d,r}} = 0$ to rewrite the above equation as

$$\lambda_{\mathrm{d,m}}^{(1)} = \frac{\int_\Omega A_\mathrm{r}^{*2}(G_\mathrm{m} - G_\mathrm{r})\mathrm{d}\Omega - (d_\mathrm{m} - d_\mathrm{r}) \int_\Omega |\nabla A_\mathrm{r}^*|^2 \mathrm{d}\Omega}{\int_\Omega A_\mathrm{r}^{*2} \mathrm{d}\Omega}. \tag{A.11}$$

Since the normalization factor $1/\int_\Omega A_\mathrm{r}^{*2}\mathrm{d}\Omega$ is always positive, this can be simplified if only the sign of the invasion fitness is to be calculated:

$$\lambda_{\mathrm{d,m}} > 0 \quad \Leftrightarrow \quad \int_\Omega A_\mathrm{r}^{*2} \left(G_\mathrm{m} - G_\mathrm{r}\right) - |\nabla A_\mathrm{r}^*|^2 (d_\mathrm{m} - d_\mathrm{r}) \mathrm{d}\Omega > 0. \tag{A.12}$$

It should be noted that Eqs. A.11 and A.12 due to the first-order nature of the approximation are only valid away from evolutionary singular points, i.e., when the selection gradient is not zero. We treat the case at evolutionary singular points in Appendix B.

Selection gradient

The selection gradient $D(\chi_\mathrm{r})$ is the quantity describing how the invasion fitness of a rare mutant changes with changes in trait around the resident trait, thus telling the strength and direction of directional selection for that resident. Since in a spatial system as in the previous section the invasion fitness is given by the dominant eigenvalue λ_d of a differential operator, to calculate the selection gradient we would want to compute

$$D(\chi_\mathrm{r}) = \left.\frac{\partial \lambda_\mathrm{d}}{\partial \chi_\mathrm{m}}\right|_{\chi_\mathrm{m}=\chi_\mathrm{r}}. \tag{A.13}$$

In general, this may not be possible, but for systems where Eq. A.11 is valid we can obtain the selection gradient by differentiating with respect to the mutant trait, and evaluating it at the resident trait:

$$\left.\frac{\partial \lambda_{\mathrm{d,m}}}{\partial \chi_\mathrm{m}}\right|_{\chi_\mathrm{m}=\chi_\mathrm{r}} = \frac{\partial}{\partial \chi_\mathrm{m}} \left[\frac{\int_\Omega A_\mathrm{r}^{*2}(G_\mathrm{m} - G_\mathrm{r})\mathrm{d}\Omega - (d_\mathrm{m} - d_\mathrm{r}) \int_\Omega |\nabla A_\mathrm{r}^*|^2 \mathrm{d}\Omega}{\int_\Omega A_\mathrm{r}^{*2} \mathrm{d}\Omega}\right]_{\chi_\mathrm{m}=\chi_\mathrm{r}}. \tag{A.14}$$

Since only G_m and d_m depend on χ_m this simplifies to:

$$D(\chi_\mathrm{r}) = \left.\frac{\partial \lambda_{\mathrm{d,m}}}{\partial \chi_\mathrm{m}}\right|_{\chi_\mathrm{m}=\chi_\mathrm{r}} = \frac{\int_\Omega A_\mathrm{r}^{*2} \left.\frac{\partial G_\mathrm{m}}{\partial \chi_\mathrm{m}}\right|_{\chi_\mathrm{r}} \mathrm{d}\Omega - \left.\frac{\partial d_\mathrm{m}}{\partial \chi_\mathrm{m}}\right|_{\chi_\mathrm{r}} \int_\Omega |\nabla A_\mathrm{r}^*|^2 \mathrm{d}\Omega}{\int_\Omega A_\mathrm{r}^{*2} \mathrm{d}\Omega}, \tag{A.15}$$

which recovers the expression in Eq. 10.

This result can be generalized to all self-adjoint operators. Define the derivative of a differential operator F with respect to a parameter χ, $\partial F/\partial\chi$ to be the differential operator whose terms have been differentiated with respect to that parameter, so that for instance the derivative of the reaction-diffusion operator $F_{\mathrm{RD}} = G(\mathbf{x},\chi) + d(\chi)\Delta$ with respect to χ is:

$$\frac{\partial F_{\mathrm{RD}}}{\partial\chi} = \frac{\partial G(\mathbf{x},\chi)}{\partial\chi} + \frac{\partial d(\chi)}{\partial\chi}\Delta. \tag{A.16}$$

Then, the selection gradient of the resident phenotype, whose equilibrium dynamics are given by

$$0 = \frac{\partial A^*_{\mathrm{r}}}{\partial t} = F(E^*_{\mathrm{r}},\chi_{\mathrm{r}})A^*_{\mathrm{r}}, \tag{A.17}$$

where F is self-adjoint, can be calculated as:

$$D(\chi_{\mathrm{r}}) = \left.\frac{\partial\lambda_{\mathrm{d}}}{\partial\chi}\right|_{\chi=\chi_{\mathrm{r}}} = \frac{\int_\Omega A^*_{\mathrm{r}}\left.\frac{\partial F}{\partial\chi}\right|_{\chi_{\mathrm{r}}} A^*_{\mathrm{r}}\mathrm{d}\Omega}{\int_\Omega A^{*2}_{\mathrm{r}}\mathrm{d}\Omega}. \tag{A.18}$$

Equation A.18 is in general, and particularly in quantum mechanics, referred to as the (first) Hellmann-Feynman theorem, after physicists Richard Feynman, and David Hellmann who among others proved the theorem (Hellmann 1933; Feynman 1939).

The above result also implies the following: If the growth and dispersal of the density of a morph can be described with a self-adjoint operator and dispersal is not under selection, then Eq. 11 can be used to calculate the selection gradient. An example would be integro-differential equations where nonlocal dispersal is described by a symmetric dispersal kernel.

Reproductive value

The reproductive value in a class-structured system of an individual in a given class is defined as the relative long-term contribution to a population by that individual (see e.g., Caswell 2001). In the case of continuous space, we may think of the class of an individual as being described by its spatial coordinate $\mathbf{x}$, and we may describe the distribution of individuals with unit density at a given location $\mathbf{x}_0$ as being $\delta(\mathbf{x}-\mathbf{x}_0)$, where δ is the Dirac delta distribution. We would like to show that the relative long-term reproductive contribution of such a distribution of individuals is proportional to $A^*(\mathbf{x}_0)$, the resident equilibrium distribution, so that the reproductive value is the same as that distribution.

Consider the equation:

$$\frac{\partial A(t,\mathbf{x})}{\partial t} = G(A,\mathbf{x})A + d\Delta A, \tag{A.19}$$

where we look at the linearized equation around the equilibrium solution $A^*(\mathbf{x})$, so that $G^*(\mathbf{x}) = G(A^*,\mathbf{x})$. We assume that the linear operator described by the linearized

equation is self-adjoint on a finite-sized landscape, and sufficiently well behaved (this will very nearly always be the case for ecological applications, but see Cantrell and Cosner 2004 for a full account). Given this, the operator has a complete set of orthogonal eigenfunctions $u_k(\mathbf{x})$ so that the solution to the equation

$$\frac{\partial A(t,\mathbf{x})}{\partial t} = G^*(\mathbf{x})A + d\Delta A \tag{A.20}$$

is given by $A(t,\mathbf{x}) = \sum_{k=0}^{\infty} c_k e^{\lambda_k t} u_k(\mathbf{x})$ and since u_k are orthogonal the constants c_k are given by $c_k = \langle A(0,\mathbf{x}), u_k(\mathbf{x})\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product given by $\langle u(\mathbf{x}), v(\mathbf{x})\rangle = \int u(\mathbf{x})v(\mathbf{x})\mathrm{d}\mathbf{x}$. If A^* is the stable equilibrium solution to the above equation then the dominant eigenvalue $\lambda_\mathrm{d} = 0$, and has an associated eigenfunction $u_\mathrm{d}(\mathbf{x}) = A^*(\mathbf{x})/\|A^*\|$, where $\|X(\mathbf{x})\| := \sqrt{\langle X, X\rangle}$, and for long times we have:

$$\lim_{t\to\infty} A(t,\mathbf{x}) = \langle A(0,\mathbf{x}), u_\mathrm{d}(\mathbf{x})\rangle u_\mathrm{d}(\mathbf{x}). \tag{A.21}$$

Since we are interested in the long-term contribution of an individual at coordinate $\mathbf{x_0}$ we let $A(0,\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}_0)$, and hence get

$$\lim_{t\to\infty} A(t,\mathbf{x}) = \langle A(0,\mathbf{x}), u_\mathrm{d}(\mathbf{x})\rangle u_\mathrm{d}(\mathbf{x}) = \frac{A^*(\mathbf{x}_0)}{\|A^*\|}\frac{A^*(\mathbf{x})}{\|A^*\|}. \tag{A.22}$$

Thus, for all $\mathbf{x}$ the relative long-term contribution of an individual at location $\mathbf{x}_0$ is given by $A^*(\mathbf{x}_0)$ in the sense that if we consider individuals at locations $\mathbf{x}_0^{(1)}$ and $\mathbf{x}_0^{(2)}$, then

$$\frac{\lim_{t\to\infty}(A(t,\mathbf{x}) \,|\, A(0,\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}_0^{(1)}))}{\lim_{t\to\infty}(A(t,\mathbf{x}) \,|\, A(0,\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}_0^{(2)}))} = \frac{A^*(\mathbf{x}_0^{(1)})}{A^*(\mathbf{x}_0^{(2)})}. \tag{A.23}$$

Hence, $A^*(\mathbf{x}_0)$ is the reproductive value of an individual at location $\mathbf{x}_0$.

Online Appendix B: Sympatric and parapatric evolutionary branchings

In the preceding appendix we derived an expression for the selection gradient, measuring the strength and direction of directional selection on a trait (Eq. A.18). When directional selection ceases and the system is at an evolutionarily singular point, however, additional information is required to determine whether the system is at an evolutionary branching point, or in an evolutionarily stable strategy. For a non-spatial system, using the notation of Eqs. 3 and 4, the second derivative of the invasion fitness

$$\left.\frac{\partial^2 G(E_\mathrm{r}^*,\chi)}{\partial\chi^2}\right|_{\chi=\chi_\mathrm{r}} \tag{B.1}$$

at an evolutionarily singular point will determine selection to be disruptive if the second derivative is positive, and stabilizing if it is negative (see e.g., Geritz et al. 1998).

Similarly, the second derivative of the invasion fitness of a spatial system whose dynamics are given by a partial differential equation with a self-adjoint operator can also be calculated. Let there be a system

$$\frac{\partial A(t,\mathbf{x})}{\partial t} = F(\chi)A, \tag{B.2}$$

where A is a phenotype density distribution, $\mathbf{x}$ is a spatial coordinate, t is the time, F a linear differential operator, and χ is the phenotype trait. Let λ_n be an eigenvalue of F, and A_n be the eigenfunction associated with λ_n. Furthermore, let λ_d be the principal eigenvalue of F, and let $A_\mathrm{d} = A_\mathrm{r}^*$ be the eigenfunction associated with λ_d, which is the equilibrium distribution of the resident. Furthermore introduce notation:

$$\partial_\chi X = \left.\frac{\partial X}{\partial \chi}\right|_{\chi=\chi_\mathrm{r}}, \quad \partial_\chi^2 X = \left.\frac{\partial^2 X}{\partial \chi^2}\right|_{\chi=\chi_\mathrm{r}}. \tag{B.3}$$

In Appendix A we calculated the selection gradient to be:

$$\partial_\chi \lambda_\mathrm{d} = \frac{\int_\Omega A_\mathrm{d}(\partial_\chi F)A_\mathrm{d}\mathrm{d}\Omega}{\int_\Omega A_\mathrm{d}^2\mathrm{d}\Omega}. \tag{B.4}$$

Applying the derivative with respect to the trait again, and using the Hellmann-Feynman theorems (Hellmann 1933; Feynman 1939), as well as the fact that the selection gradient is zero at an evolutionarily singular point yields an expression for the curvature of the fitness landscape around a resident trait in ecological equilibrium at an evolutionarily singular point:

$$\partial_\chi^2 \lambda_\mathrm{d} = \frac{\int_\Omega A_\mathrm{d}(\partial_\chi^2 F)A_\mathrm{d}\mathrm{d}\Omega}{\int_\Omega A_\mathrm{d}^2\mathrm{d}\Omega} + 2\sum_{n\neq\mathrm{d}} \frac{\left|\int_\Omega A_n(\partial_\chi F)A_\mathrm{d}\mathrm{d}\Omega\right|^2}{\int_\Omega A_\mathrm{d}^2\mathrm{d}\Omega \int_\Omega A_n^2\mathrm{d}\Omega(\lambda_\mathrm{d}-\lambda_n)}. \tag{B.5}$$

This expression for the curvature, as opposed to the selection gradient (Eq. A.18), depends not only on the dominant eigenvalue and eigenfunction, but also has contributions from all other eigenfunctions of F. The first term integrates the curvature of F over space weighted by resident equilibrium distribution, and the second term integrates the gradient of F, weighted by both the resident equilibrium density as well as other potential spatial distributions for the mutant. This allows us to interpret the two terms of the expression as contributions from the non-spatial and spatial dynamics respectively. The first term represents the ecological selection pressure on the phenotype, which can be either stabilizing or disruptive. The second term, representing the spatial dynamics is always positive, since the dominant eigenvalue λ_d is larger than all other eigenvalues per definition. This can interpreted as that spatial heterogeneity always contributes to diversification, all other things being equal.

If the first term is negative this means that there is no trait value around the resident's trait value which increases invasion fitness of a mutant having the same

specific spatial distribution as the resident, A_d. The curvature may still be positive if the magnitude of the second term is large enough, signifying that there exist other spatial distributions (eigenfunctions) A_n which for a small change in trait value would gain positive invasion fitness, putting the resident at a local fitness minimum. If on the other hand the first term is positive, this implies that there exist trait values close to the resident's trait value which would increase invasion fitness of a mutant having the same spatial distribution as the resident.

What this means is that we can classify evolutionary branching points into spatial and non-spatial branchings. We can call these cases sympatric and parapatric diversification, if we are careful to understand what really characterize them is the source of disruptive selection, or more precisely, the sign of the first term of Eq. B.5. The reason the names seem appropriate is that the first case corresponds to when a phenotype with a specific spatial distribution can branch into two phenotypes with the same spatial distributions, and the second case corresponds to a branching event where at least one of the new phenotypes must have a different spatial distribution from its progenitor.

If we consider the case when diffusivity is not under selection, then Eq. B.5 becomes:

$$\partial_\chi^2 \lambda_\mathrm{d} = \frac{\int_\Omega A_\mathrm{d}^2 (\partial_\chi^2 G) \mathrm{d}\Omega}{\int_\Omega A_\mathrm{d}^2 \mathrm{d}\Omega} + 2 \sum_{n \neq \mathrm{d}} \frac{\left| \int_\Omega A_n (\partial_\chi G) A_\mathrm{d} \mathrm{d}\Omega \right|^2}{\int_\Omega A_\mathrm{d}^2 \mathrm{d}\Omega \int_\Omega A_n^2 \mathrm{d}\Omega (\lambda_\mathrm{d} - \lambda_n)}. \tag{B.6}$$

The first term now describes the weighted mean of stabilizing/disruptive selection over all of space. The second term is still difficult to parse, but we can restrict ourselves to sufficiently high diffusivity, in the sense that diffusivity is high enough that the phenotype is close to or above the bifurcation in d from monomorphic to polymorphic, and certain domains. This allows us to make an estimate of the maximal amount of disruptive selection that offers some insight into the transition from a monomorphic community into a polymorphic one. If diffusivity is sufficiently high, we can approximate the eigenvalues of F by the eigenvalues of $d\Delta$. We consider only such bounded domains that the eigenvalues of Δ, $\hat{\lambda}_n$ are in a non-increasing sequence with only the largest eigenvalue being zero, and $\hat{\lambda}_n \to -\infty$ as $n \to \infty$. This will be so for many ecological applications, and is certainly true for the examples considered in this paper (see Cantrell and Cosner 2004, for a thorough discussion on what domains and boundary conditions are permissible). This means that the eigenvalues of $d\Delta$ are given by $d\hat{\lambda}_n$. Inserting this into the expression above and using Parseval's identity yields an upper bound for disruptive selection:

$$\partial_\chi^2 \lambda_\mathrm{d} \leq \frac{\int_\Omega A_\mathrm{d}^2 (\partial_\chi^2 G) \mathrm{d}\Omega}{\int_\Omega A_\mathrm{d}^2 \mathrm{d}\Omega} + 2 \frac{\int A_\mathrm{d}^2 (\partial_\chi G)^2 \mathrm{d}\Omega}{\int A_\mathrm{d}^2 \mathrm{d}\Omega} \frac{K}{d}. \tag{B.7}$$

Here, K is a constant that depends only on the shape and size of the domain that is such that $K\hat{\lambda}_n \leq -1$ for all $n > 0$. Since at an evolutionarily singular point it must be true that $\int_\Omega A_\mathrm{d}^2 \partial_\chi G \mathrm{d}\Omega = 0$, this means the expression $\int_\Omega A_d^2 (\partial_\chi G)^2 \mathrm{d}\Omega / \int_\Omega A_\mathrm{d}^2 \mathrm{d}\Omega$ can be interpreted as the weighted variance of local selection gradients in the region.

This means, that under sufficiently high diffusivity, the maximal amount of disruptive selection that spatial heterogeneity can contribute is proportional to the ratio between the weighted variance of local selection gradients and diffusivity.

To take an example consider the 4 cases of the model in Example 1, with $a_0 = 2.5$ or $a_0 = -2.5$, and $d = 1$ or $d = 10^{-4}$, representing cases with a weak and strong trade-off, and high and low diffusivity, respectively. Putting a single phenotype at the evolutionarily singular point $\chi_\mathrm{r} = 0$, we can look at the two terms for the selection curvature:

a_0 \ d	1	10^{-4}
2.5	$\partial^2_\chi \lambda_\mathrm{d} = -5.80 \cdot 10^{-3} + 2.19 \cdot 10^{-6}$	$\partial^2_\chi \lambda_\mathrm{d} = -5.80 \cdot 10^{-3} + 2.19 \cdot 10^{-2}$
-2.5	$\partial^2_\chi \lambda_\mathrm{d} = 6.87 \cdot 10^{-2} + 3.25 \cdot 10^{-4}$	$\partial^2_\chi \lambda_\mathrm{d} = 6.87 \cdot 10^{-2} + 3.25 \cdot 10^{0}$

We can see that the first terms are negative for $a_0 = 2.5$, similar to what non-spatial adaptive dynamics would predict for cases where a weak trade-off between two resources favors generalists (see de Mazancourt and Dieckmann (2004), for a treatment of trade-offs and evolutionarily singular points). For $d = 1$, which approaches the well-mixed case, the contribution from the spatial term is small, since the fitness contribution from different spatial distributions is weak. In contrast, sufficiently low rates of diffusion ($d = 10^{-4}$) allow the contribution of the second term to become large enough to render the net curvature positive, since at least one spatial distribution that differs from the resident phenotype's spatial distribution might yield positive invasion fitness for trait values close to the resident phenotype's trait value. For the strong trade-off ($a_0 = -2.5$, favoring two specialists) there is a sympatric branching for both the well-mixed and non-well-mixed case, again similar to what non-spatial adaptive-dynamics theory would predict.

Example 1 is also very well suited to analysis by Eq.B.7, as the system with one resident at the evolutionarily singular point, $\chi_\mathrm{r} = 0$ admits a solution that is constant in space. This means the solution can be computed analytically, as well as that the eigenvalues are exactly those of $d\Delta$. The eigenvalues of Δ are given by $\lambda_\mathrm{d} = \lambda_{00} = 0$ and $\hat{\lambda}_{nm} = -\pi^2(n^2+m^2)/L^2$, where $L = 2$ is the length of the side of the domain, which means that we can take $K = L^2/\pi^2$. However, due to the symmetry of the problem when $n = 0$ or $m = 0$ there will be no contribution from the integral over $\partial_x G$ and the eigenfunctions of $d\Delta$, and so we can get a better bound by only considering $n, m \geq 1$ in the sum, and so we can choose $K = L^2/(2\pi^2)$. Taken together this implies that:

$$\partial^2_\chi \lambda_\mathrm{d} \leq \frac{\int_\Omega \partial^2_\chi G \mathrm{d}\Omega}{L^2} + 2\frac{\int_\Omega (\partial_\chi G)^2 \mathrm{d}\Omega}{L^2}\frac{L^2}{2\pi^2 d}. \tag{B.8}$$

Here, G does not depend on d, and so we can calculate the smallest possible d for which we know that selection is stabilizing by setting the above expression to 0. Computing this yields $d = 3.86 \cdot 10^{-4}$, when we compute the variance of selection based on our numerical discretization, which agrees very well with what we can observe in Fig. 3.

When we compute the variance of selection gradients based on the analytical solutions for the continuous system we get $d = 3.44 \cdot 10^{-4}$. The small discrepancy is due to the increased variance arising from the slight patchiness of the discretization.

While useful for theoretical considerations, Eq. B.5 may in practice be less useful than direct numerical computation of the dominant eigenvalue of F for traits close to the resident to determine whether the resident is at an evolutionary branching point or ESS, since only the dominant eigenvalue has to be calculated, as opposed to all eigenvalues and eigenfunctions of F_{r}.

Online Appendix C: Population level selection in a spatial quantitative genetics model

We investigate to what extent our conclusions about directional selection derived in the setting of adaptive dynamics hold true also in a spatial quantitative genetics setting by using a model developed by Kirkpatrick and Barton (1997). The model describes trait evolution in a population with normally distributed trait values at each point in space. In contrast to the adaptive-dynamics framework introduced above, selection is frequency-independent as individual growth rates are allowed to depend only on local density, spatial location, and trait value. At the same time, the model is more permissive in that the average trait value can vary throughout the spatial domain, allowing local adaptation. Specifically, the spatial quantitative genetics model is described by the following equations:

$$\frac{\partial A(t,\mathbf{x})}{\partial t} = \left(G(A,\mathbf{x},z) + \frac{1}{2} V \left.\frac{\partial^2 G}{\partial \chi^2}\right|_{\chi=z} + d\Delta \right) A =: FA \tag{C.1a}$$

$$\frac{\partial z(t,\mathbf{x})}{\partial t} = V \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} + d\left(2\nabla \ln A \cdot \nabla z + \Delta z\right) \tag{C.1b}$$

$$\nabla A \cdot \hat{\mathbf{n}} = 0, \qquad \nabla z \cdot \hat{\mathbf{n}} = 0 \qquad \text{on the boundary } (\Gamma). \tag{C.1c}$$

In the limit of small standing variation of the trait distribution, i.e., for small values of V, this system becomes similar to the adaptive-dynamics case in several respects. First, low standing variation implies that the population is close to monomorphic at each point in space. Second, since the amount of standing variation sets the overall pace of the system's evolutionary dynamics, a population with low standing variation will be close to ecological equilibrium over evolutionary timescales. Third, at small standing variation, the diffusion term dominates the trait equation (Eq. C.1b), causing the local trait means z to be close to identical everywhere in space. This creates a situation where diffusion is dominant on ecological time-scales, homogenizing the trait distribution in space, which means that only the selection gradient will be acting on evolutionary time-scales. Below, we show that these similarities result in congruous equations for

the adaptive-dynamics and quantitative-genetics frameworks. Specifically, when the standing variation tends to zero, i.e. $V \to 0^+$, Eqs. C.1 reduce to:

$$\frac{\partial A(t,\mathbf{x})}{\partial t} = \left(G(A,\mathbf{x},z_c) + d\Delta\right) A \tag{C.2a}$$

$$\frac{\mathrm{d}z_c(t)}{\mathrm{d}t} = V \frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int A^2 \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z_c} \mathrm{d}\mathbf{x} \tag{C.2b}$$

$$\nabla A \cdot \hat{\mathbf{n}} = 0 \qquad \text{on the boundary } (\Gamma). \tag{C.2c}$$

Here $\hat{\mathbf{n}}$ is a unit vector pointing outwards from the boundary, and the rest of the notation is as in the main text. We assume equilibrium dynamics, both in the ecological sense that there exists an ecological equilibrium distribution $A^*(\mathbf{x})$ for each (reasonable) distribution $z(\mathbf{x})$ or z_c, and in the sense that the system will approach an eco-evolutionary equilibrium, i.e., $\partial A/\partial t \to 0$ and $\partial z/\partial t \to 0$ as $t \to \infty$. We shall also assume that this eco-evolutionary equilibrium is a unique attractor for all initial conditions under consideration. Furthermore we assume that for all $\mathbf{x}$, t, and V the density distribution of individuals A is strictly positive so that there exists a number $0 < A_{\min} \leq A$. Lastly, in order for it to be meaningful to talk about the full model approaching the reduced model we assume that the solutions to Eqs. C.1 are at least continuous in V.

When the system described by Eqs. C.1 is in equilibrium with $V > 0$ we can see from Eq. C.1b that

$$0 = V \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} + d\left(2\nabla \ln A \cdot \nabla z + \Delta z\right) \;\Rightarrow \tag{C.3}$$

$$A^2 V \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} = -A^2 d\left(\frac{2}{A}\nabla A \cdot \nabla z + \Delta z\right) \Rightarrow \tag{C.4}$$

$$V \int A^2 \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} \mathrm{d}\mathbf{x} = -d \int \nabla \cdot \left(A^2 \nabla z\right) \mathrm{d}\mathbf{x} \quad \Leftrightarrow \tag{C.5}$$

$$V \int A^2 \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} \mathrm{d}\mathbf{x} = -d \int A^2 \nabla z \cdot \hat{\mathbf{n}} \mathrm{d}\Gamma = 0. \tag{C.6}$$

The last line uses the divergence theorem and the boundary conditions for z. From this we see that $\int A^2 \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} \mathrm{d}\mathbf{x} = 0$ for all $V > 0$ and hence also in the limit $V \to 0^+$. On the other hand, when $V = 0$ Eq. C.1b becomes

$$2\nabla \ln A \cdot \nabla z + \Delta z = 0, \tag{C.7}$$

which for the given boundary conditions by the maximum principle and Hopf's lemma (see e.g., Evans 2010) has only a spatially constant solution z_c. Hence in the limit $V \to 0^+$ the equilibrium solution to Eqs. C.1 and Eqs. C.2 both need to satisfy

$$G(A,\mathbf{x},z) + d\Delta A = 0 \tag{C.8a}$$

$$\int A^2 \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} \mathrm{d}x = 0 \tag{C.8b}$$

$$z = z_c, \tag{C.8c}$$

in order to be a solution, and hence if (A, z_c) is a solution of Eqs. C.2 it will also be a solution of Eqs. C.1.

By the above arguments, and the assumption of unique eco-evolutionary equilibria, we know that the solutions of both systems will certainly converge to the same eco-evolutionary equlibrium distribution as $t \to \infty$. For the time-dependent system we shall argue that in the limit of small V the rate of change in time of a weighted spatial average z becomes close to z_c, and that z will be close to constant in the sense that the weighted spatial variance of z becomes small.

To make notation more compact and readable we introduced a spatial average weighted by the population densities of quantities denoted by $\langle\cdot\rangle$ given by:

$$\langle Y\rangle = \frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int AYA\mathrm{d}\mathbf{x}, \tag{C.9}$$

so that e.g.,

$$\langle z\rangle = \frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int A^2 z\mathrm{d}\mathbf{x}, \qquad \langle F\rangle = \frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int AFA\mathrm{d}\mathbf{x}, \tag{C.10}$$

where F is the differential operator in Eq. C.1a.

We calculate the time-derivative of this average of the mean trait distribution:

$$\frac{\mathrm{d}\langle z\rangle}{\mathrm{d}t} = V\left\langle \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} \right\rangle + 2(\langle zF\rangle - \langle z\rangle\langle F\rangle). \tag{C.11}$$

The first term is the one in Eq. C.2b, so in order to show that we approach the reduced model for small V, it is sufficient to show that the second term in the expression above goes to zero faster than linearly in V, and that the distribution of z in space becomes increasingly flat. We can rewrite the second term as

$$\langle zF\rangle - \langle z\rangle\langle F\rangle = \frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int (z - \langle z\rangle)A(F - \langle F\rangle)A\mathrm{d}\mathbf{x}, \tag{C.12}$$

to which we can apply the Cauchy-Swartz theorem to show that

$$|\langle zF\rangle - \langle z\rangle\langle F\rangle|^2 \leq \left(\frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int (FA)^2 \mathrm{d}\mathbf{x} - \langle F\rangle^2\right)\left(\langle z^2\rangle - \langle z\rangle^2\right). \tag{C.13}$$

We wish to show that both these factors depended on at least V^2, so that $|\langle zF\rangle - \langle z\rangle\langle F\rangle|$ will as well. To compare the full model to the reduced model we let the initial conditions

for the full model $z_0 := z(0,x)$, and $A_0 := A(0,x)$ be such that z_0 is constant in space, and A_0 is in ecological equilibrium, so that $FA_0 = 0$. This implies that:

$$\left.\frac{\partial A}{\partial t}\right|_{t=0} = 0 \tag{C.14}$$

$$\left.\frac{\partial z}{\partial t}\right|_{t=0} = V \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z_0}. \tag{C.15}$$

The idea is that from this initial state the pace of the ecological dynamics FA and the weighted variance $\langle z^2\rangle - \langle z\rangle^2$ are bounded by V and V^2 respectively. To see why this is so we will argue, somewhat formally, by expanding the solutions in power-series. Taking higher order derivatives in time at $t = 0$, by recursive application of Eqs. C.1, these will all have a factor V, and so Taylor-expanding FA and z in time we define

$$FA(t,x) =: V\alpha(t,x) \qquad\qquad z(t,x) =: z_0 + V\beta(t,x), \tag{C.16}$$

where α and β depend only on non-negative powers in V. We insert this into Eq. C.13 to get:

$$|\langle zF\rangle - \langle z\rangle\langle F\rangle| \leq V^2 \sqrt{\left(\frac{1}{\int A^2 \mathrm{d}\mathbf{x}}\int \alpha^2 \mathrm{d}\mathbf{x} - \left[\frac{1}{\int A^2 \mathrm{d}\mathbf{x}}\int A\alpha \mathrm{d}\mathbf{x}\right]^2\right)(\langle\beta^2\rangle - \langle\beta\rangle^2)} =: V^2\gamma(t) \tag{C.17}$$

where z_0 cancels in the second factor since it is constant in space. That FA will not grow uncontrollably in time follows from the assumption that the solution will approach an eco-evolutionary equilibrium over time, and so $FA \to 0$ as $t \to \infty$. To see why also the weighted variance will be controlled in time we can estimate the quantity $\langle z^2\rangle - \langle z\rangle^2$ at the eco-evolutionary equilibrium. Let $(A_{\mathrm{eq}}, z_{\mathrm{eq}})$ be the equilibrium solution to Eqs. C.1 and estimate

$$\begin{aligned}\langle z_{\mathrm{eq}}^2\rangle - \langle z_{\mathrm{eq}}\rangle^2 &= \frac{1}{\|A_{\mathrm{eq}}\|^2}\|(z_{\mathrm{eq}} - \langle z_{\mathrm{eq}}\rangle)A_{\mathrm{eq}}\|^2 \overset{(i)}{\leq} C_1\frac{1}{\|A_{\mathrm{eq}}\|^2}\|A_{\mathrm{eq}}\nabla z_{\mathrm{eq}}\|^2 \\ &\leq C_2\frac{1}{\|A_{\mathrm{eq}}\|^2 A_{\min}^2}\|A_{\mathrm{eq}}^2\nabla z_{\mathrm{eq}}\|^2 \overset{(ii)}{\leq} C_2\frac{1}{\|A_{\mathrm{eq}}\|^2 A_{\min}^2}\frac{V^2}{d^2}\|A_{\mathrm{eq}}^2 \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z}\|^2.\end{aligned} \tag{C.18}$$

Here, the norms are taken in L^2, inequality (i) is due to the weighted Poincaré inequality (see e.g., Pechstein and Scheichl 2011), and inequality (ii) is due to the bound for $A^2\nabla z$ in the divergence equation $\nabla\cdot(A^2\nabla z) = \frac{V}{d}A^2\left.\frac{\partial G}{\partial \chi}\right|_{\chi=z}$ as given in **?**. Hence $\langle z^2\rangle - \langle z\rangle^2$ is bounded by V^2 as $t \to \infty$, and will not grow uncontrollably in time.

Taken together this implies that as V gets small

$$\frac{\mathrm{d}\langle z\rangle}{\mathrm{d}t} \approx V\left\langle \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z}\right\rangle, \tag{C.19}$$

since the second term goes as V^2. More precisely we can rescale time as $\tau = Vt$ and see that

$$\left| \frac{\mathrm{d}\langle z \rangle}{\mathrm{d}\tau} - \left\langle \frac{\partial G}{\partial \chi}\Big|_{\chi=z} \right\rangle \right| \leq V\gamma \Rightarrow$$
$$\lim_{V\to 0^+} \frac{\mathrm{d}\langle z \rangle}{\mathrm{d}\tau} = \left\langle \frac{\partial G}{\partial \chi}\Big|_{\chi=z} \right\rangle. \tag{C.20}$$

Furthermore, since $\langle z^2 \rangle - \langle z \rangle^2$ is a measure of the variance of z, and goes to zero as V^2, z will be close to constant, and hence close to z_c as V gets small. Since the rate of change $\langle z \rangle$ will approach that of z_c, and the deviation from this mean will be small at all times, Eqs. C.1 will approach Eqs. C.2 as V becomes small.

It is worth pointing out that taking the square of the population density in the weighted average of local selection gradients is necessary, and that weighting directly with the densities will not produce the correct results, even in the limit of small variation. Mathematically, the reason why $\frac{1}{\int A \mathrm{d}\mathbf{x}} \int A \frac{\partial G}{\partial \chi}\big|_{\chi=z_c} \mathrm{d}\mathbf{x}$ cannot predict the evolution of a spatially constant trait is that to calculate the time derivative of $\langle z \rangle$ we use that

$$d\frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int A^2 \left[\frac{2}{A} \nabla A \cdot \nabla z + \Delta z \right] \mathrm{d}\mathbf{x} = d\frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int \nabla \cdot \left(A^2 \nabla z \right) \mathrm{d}\mathbf{x} = \tag{C.21}$$

$$= d\frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int_\Gamma A^2 \nabla z \cdot \hat{\mathbf{n}} \mathrm{d}\Gamma = 0, \tag{C.22}$$

where the last integral over Γ is over the boundary of the domain, which due to the boundary conditions for z is zero. This term however can only be integrated out due to the multiplication by A^2 in the integral. Only multiplying with a factor A, will leave a term that will make contributions on the order V to evolution of the spatially averaged trait, which is of the same order as the term $\frac{1}{\int A \mathrm{d}\mathbf{x}} \int A \frac{\partial G}{\partial \chi}\big|_{\chi=z_c} \mathrm{d}\mathbf{x}$.

Lastly we will examine to what extent the quantity $\left\langle \frac{\partial G}{\partial \chi}\big|_{\chi=z} \right\rangle$ holds similar interpretations to that of the population-level selection gradient in AD. In AD, the population level selection gradient measures the change in population invasion fitness with a small change in trait homogeneously across space. This case is similar for the QG model, given a specific definition of population-level fitness. One way to motivate this definition is to look at the per capita net growth rate of the entire population. To do this, a measure of the size of the total population is needed. The most obvious candidate would be $\int A \mathrm{d}\mathbf{x}$, but this turns out to be inconsistent with results above. Instead, they suggest that the better measure of total population size from a fitness perspective, denoted $\|A\|$, would be $\|A\| = \sqrt{\int A^2 \mathrm{d}\mathbf{x}}$. Using this definition we can calculate the per capita net growth rate:

$$\frac{1}{\|A\|} \frac{\mathrm{d}\|A\|}{\mathrm{d}t} = \frac{1}{\int A^2 \mathrm{d}\mathbf{x}} \int A \frac{\partial A}{\partial t} \mathrm{d}\mathbf{x} = \langle F \rangle, \tag{C.23}$$

which becomes the measure of population level fitness for the population.

To properly determine how this population level fitness changes with a small constant shift in mean trait z, we considered the equation from which the QG approximation is derived, namely:

$$\frac{\partial \mathcal{A}(t,\mathbf{x},\chi)}{\partial t} = G(A,\mathbf{x},\chi)\mathcal{A} + d\Delta\mathcal{A} =: \mathcal{F}\mathcal{A}, \tag{C.24}$$

which describes the growth and random movement of the density of individuals $\mathcal{A}$ at location $\mathbf{x}$ at time t with trait χ. The QG approximation is made by assuming a solution of the form

$$\mathcal{A}(t,\mathbf{x},\chi) = A(t,\mathbf{x})\eta(\chi, z(t,\mathbf{x})), \qquad \eta(\chi, z(t,\mathbf{x})) = \frac{1}{\sqrt{2\pi V}}\exp\left[-\frac{(\chi - z(t,\mathbf{x}))^2}{2V}\right]. \tag{C.25}$$

$G(A,\mathbf{x},\chi)$ is then Taylor expanded to order 2 around $G(A,\mathbf{x},z)$, and Eq. C.24 is integrated over all χ, which yields Eqs. C.1. Note that for this derivation to be valid the per capita net growth function G in Eq. C.24 may only depend on the total density of individuals $A = \int \mathcal{A}\mathrm{d}\chi$ at each point in space.

We can then use this formulation to calculate how the population level fitness changes with a small constant shift z_s across space by computing

$$\frac{\partial\langle F\rangle}{\partial\langle z\rangle} := \frac{\partial}{\partial z_s}\left[\frac{1}{\int A^2\mathrm{d}\mathbf{x}}\int A\int \mathcal{F}(A\eta(\chi, z+z_s))\mathrm{d}\chi\mathrm{d}\mathbf{x}\right]_{z_s=0}, \tag{C.26}$$

which after expanding the inner integral, and once again Taylor expanding G indeed yielded

$$\frac{\partial\langle F\rangle}{\partial\langle z\rangle} = \left\langle \left.\frac{\partial G}{\partial\chi}\right|_{\chi=z}\right\rangle = \frac{1}{\int A^2\mathrm{d}\mathbf{x}}\int A^2\left.\frac{\partial G}{\partial\chi}\right|_{\chi=z}\mathrm{d}\mathbf{x}. \tag{C.27}$$

This calculation does not depend on V being small, and so suggests a general definition for the population level selection gradient for this spatial QG model. It can, however, only fully predict the evolution of $\langle z\rangle$ in the small V limit.

Much like in the AD case, these conclusions are equally valid for certain types of metapopulation models. For such metapopulations the dynamics for the population density A_i, and mean trait z_i on patch $i = 1, 2, ..., I$ are given by:

$$\frac{\mathrm{d}A_i}{\mathrm{d}t} = G_i(A_i, z_i)A_i + \frac{1}{2}V\left.\frac{\partial^2 G_i}{\partial\chi^2}\right|_{\chi=z_i}A_i + \sum_{j=1}^{I} d_{ij}A_j \tag{C.28a}$$

$$\frac{\mathrm{d}z_i}{\mathrm{d}t} = V\left.\frac{\partial G_i}{\partial\chi}\right|_{\chi=z_i} + \frac{1}{A_i}\sum_{j=1}^{I} d_{ij}(z_j - z_i)A_j. \tag{C.28b}$$

Here G_i is the per capita net growth rate on patch i, and d_{ij} are the elements of a symmetric dispersal matrix denoting the dispersal rate from patch j to i, and the diagonal elements d_{ii} the dispersal out from patch i. If for this discrete case we define

$$\langle Y \rangle := \frac{1}{\sum_{i=1}^{I} A_i^2} \sum_{i=1}^{I} \sum_{j=1}^{I} A_i Y_{ij} A_j, \qquad \langle y \rangle := \frac{1}{\sum_{i=1}^{I} A_i^2} \sum_{i=1}^{I} A_i^2 y_i \tag{C.29}$$

for matrices and vectors respectively, then all the results derived for the continuous space model above hold for this metapopulation model as well. Hence, for small V the reduced model is

$$\frac{\mathrm{d}A_i}{\mathrm{d}t} = G_i(A_i, z_i)A_i + \sum_{j=1}^{I} d_{ij} A_j \tag{C.30a}$$

$$\frac{\mathrm{d}z_c}{\mathrm{d}t} = V \frac{1}{\sum_{i=1}^{I} A_i^2} \sum_{i=1}^{I} A_i^2 \left.\frac{\partial G_i}{\partial \chi}\right|_{\chi=z_c}, \tag{C.30b}$$

and the expression $\left\langle \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z_c} \right\rangle$ can be interpreted as the population level selection gradient measuring the change in population level fitness with a small change in trait.

Numerical example

To illustrate the behavior of the spatial quantitative genetics model as V gets small we use a simple 1-dimensional version of the model on the unit interval $x \in [0, 1]$, where each location has an optimal mean trait. The model is described by:

$$\frac{\partial A(t,x)}{\partial t} = G(A, x, z)A + \frac{1}{2} V \left.\frac{\partial^2 G}{\partial \chi^2}\right|_{\chi=z} A + d\frac{\partial^2 A}{\partial x^2} \tag{C.31a}$$

$$\frac{\partial z(t,x)}{\partial t} = V \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} + d\left(2\frac{\partial \ln A}{\partial x}\frac{\partial z}{\partial x} + \frac{\partial^2 z}{\partial x^2}\right) \tag{C.31b}$$

$$\left.\frac{\partial A}{\partial x}\right|_{x=0,x=1} = 0 \qquad \left.\frac{\partial z}{\partial x}\right|_{x=0,x=1} = 0 \tag{C.31c}$$

$$G(A, x, \chi) = 1 - (\chi - x^2)^2 - A, \qquad d = 10^{-5}. \tag{C.31d}$$

We compare the outcomes to that of the reduced model:

$$\frac{\partial A_c(t,x)}{\partial t} = G(A_c, x, z_c)A_c + d\frac{\partial^2 A_c}{\partial x^2} \tag{C.32a}$$

$$\frac{\mathrm{d}z_c(t)}{\mathrm{d}t} = V \left\langle \left.\frac{\partial G}{\partial \chi}\right|_{\chi=z} \right\rangle \tag{C.32b}$$

$$\left.\frac{\partial A_c}{\partial x}\right|_{x=0,x=1} = 0. \tag{C.32c}$$

The models are discretized by the metapopulation models Eqs. C.28 and Eqs. C.30 respectively. We integrate the models to eco-evolutionary equilibrium for $V = 10^{-5}$ and $V = 10^{-7}$, after rescaling time as $\tau = Vt$ in order for the outcomes to reach equilibrium in roughly the same scaled time. We initialize the model with $z = z_c = 0.9$ and A and A_c such that they are in ecological equilibrium at $t = 0$. The outcomes can be observed in Fig. C.1. As seen there, for $V = 10^{-7}$ the outcomes of the reduced and full model are nearly identical. Realizations for a series of different values of V indicate that the maximal error in space between A and A_c, and z and z_c is proportional to V for small enough V.

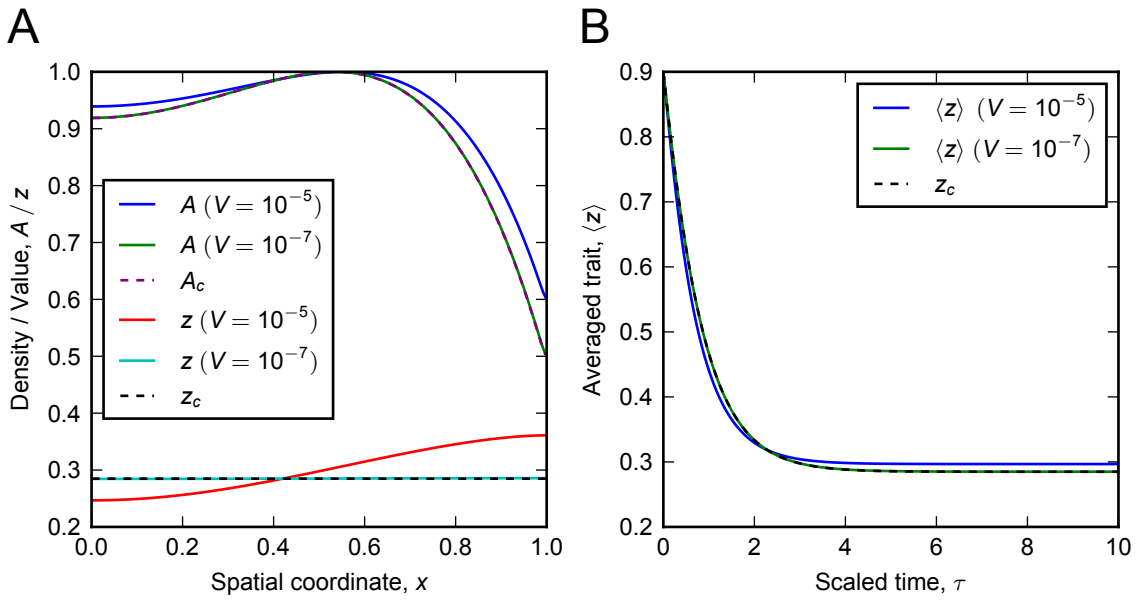

Figure C.1: Outcomes of the example quantitative genetics model. (A) Eco-evolutionary equilibrium densities in space for A and z for the full and reduced model. Note that the quantities for the reduced model depicted as broken lines are nearly directly on top of those for the full model with $V = 10^{-7}$. (B) Time evolution of the spatially averaged trait $\langle z \rangle$ for the full and reduced model.

Online Appendix D: Application to metacommunity models and numerical implementation in discretized space

The results derived in Appendix A and B are easily adapted to settings in which space is discrete, such as an evolving metacommunity on N patches. Discretization of space into a lattice is also a necessary step in the numerical implementation of any continuous space model. Below, we develop the discrete-space formalism in terms of a metacommunity model, but the same formalism is directly applicable to the discretization of a continuous space model into a lattice.

Let $A_{i,n}$ describe the density of phenotypes $i \in \{1, 2, ..., I\}$ on patch $n \in \{1, 2, ..., N\}$. Furthermore, let $g_{i,n}$ be the net growth rate of phenotype i on patch n, and let $d_{i,nm}$ be the density flow of phenotype i from patch m to patch n, and $d_{i,nn}$ be the outflow from patch n. Then the population dynamics of the system can be described by a set of coupled ordinary differential equations, so that

$$\frac{\mathrm{d}A_{i,n}}{\mathrm{d}t} = g_{i,n}A_{i,n} + \sum_{m=1}^{N} d_{i,nm}A_{i,m}. \tag{D.1}$$

If we reformulate this so that each phenotype is described by a column vector of densities on each patch $\mathbf{A}_i$, then we can let G_i be the $N \times N$ matrix that has $g_{i,n}$ as its diagonal elements, and D_i be a $N \times N$ matrix that has $d_{i,nm}$ as its elements, and get a set of vector-valued differential equations:

$$\frac{\mathrm{d}\mathbf{A}_i}{\mathrm{d}t} = G_i\mathbf{A}_i + D\mathbf{A}_i. \tag{D.2}$$

Under the conditions that all D_i are symmetric so that $d_{i,nm} = d_{i,mn}$, we can reformulate the main results of this paper for metacommunities. Let $\mathbf{A}_i^*$ be the equilibrium density distribution for phenotype i, and χ_i the trait of that phenotype. Using the same methods as in Appendix A and B, but with symmetric matrices instead of self-adjoint differential operators the selection gradient will be:

$$\left.\frac{\partial\lambda_i}{\partial\chi}\right|_{\chi=\chi_i} = \frac{\mathbf{A}_i^{*T}\left(\left.\frac{\partial G_i}{\partial\chi}\right|_{\chi=\chi_i} + \left.\frac{\partial D_i}{\partial\chi}\right|_{\chi=\chi_i}\right)\mathbf{A}_i^*}{\mathbf{A}_i^{*T}\mathbf{A}_i^*}. \tag{D.3}$$

Here the superscript T denotes the transpose. Letting $F_i = G_i + D_i$, the strength of disruptive selection, i.e., the curvature of the fitness landscape around a resident trait will be:

$$\left.\frac{\partial^2\lambda_i}{\partial\chi^2}\right|_{\chi=\chi_i} = \frac{\mathbf{A}_i^{*T}\left.\frac{\partial^2 F_i}{\partial\chi^2}\right|_{\chi=\chi_i}\mathbf{A}_i^*}{\mathbf{A}_i^{*T}\mathbf{A}_i^*} + 2\sum_{n\neq\mathrm{d}}\frac{\left(\mathbf{A}_n^T\left.\frac{\partial F_i}{\partial\chi}\right|_{\chi=\chi_i}\mathbf{A}_\mathrm{d}\right)^2}{\mathbf{A}_n^T\mathbf{A}_n\cdot\mathbf{A}_\mathrm{d}^T\mathbf{A}_\mathrm{d}\cdot(\lambda_\mathrm{d}-\lambda_n)}, \tag{D.4}$$

where $\mathbf{A}_\mathrm{d}$ is the eigenvector corresponding to the dominant eigenvalue λ_d of F_i, and the sum is over all other eigenvectors $\mathbf{A}_n$ corresponding to eigenvalues λ_n of F_i. The

eigenvector corresponding to the dominant eigenvalue is the same as the equilibrium distribution of the phenotype, i.e., $\mathbf{A}_\mathrm{d} = \mathbf{A}_i^*$. This expression can be used to differentiate sympatric and parapatric diversification in the sense of Appendix B. In the discrete patch case the situation is similar to the well known case of eigenvalue sensitivity analysis for matrix models (see Caswell 2001). However what both permits and makes useful the expressions Eq. D.3 and Eq. D.4 is that for symmetric dispersal matrices, the whole matrix F is symmetric and so has equal left and right eigenvectors.

The matrix formulation of the equations is also the basis of the numerical solutions to the partial differential equation models in this paper, as they naturally arise as approximations to the PDE-systems.

Online Appendix E: Details of the sinking phytoplankton model

The model for the population dynamics of sinking algae competing for light and nutrients is the same as the fixed stoichiometry version of a model used by Jäger et al. (2010). We reproduce a brief description of the system and its equations here, as well as how evolutionary dynamics are introduced to the system.

The rates of change of algae and dissolved nutrients are described by the equations

$$\frac{\partial A_j(z,t)}{\partial t} = G_j(R,I)A_j - l_\mathrm{bg}A_j - v\frac{\partial A_j}{\partial z} + d\frac{\partial^2 A_j}{\partial z^2} \tag{E.1a}$$

$$\frac{\partial R(z,t)}{\partial t} = -q\sum_j G_j(R,I)A_j + ql_\mathrm{bg}\sum_j A_j + d\frac{\partial^2 R}{\partial z^2}, \tag{E.1b}$$

where A_j is the density of phenotype j, and R is the concentration of dissolved nutrients. Light intensity at depth z is described by Beer-Lambert's law

$$I(z) = I_0 \exp\left(-k\int_0^z \sum_j A_j(s,t)\mathrm{d}s - k_\mathrm{bg}z\right), \tag{E.2}$$

with algal light attenuation coefficient k, background attenuation k_bg, and incoming light intensity at the surface of the water column I_0. The light and nutrients are used for algal growth in a multiplicative manner described by two Monod functions,

$$G_j(R,I) = G_\mathrm{max}\frac{R}{M_j+R}\frac{I}{H_j+I}, \tag{E.3}$$

with half-saturation constants for nutrients and light M_j and H_j. The algae suffer background losses l_bg, whose relative nutrient content, q, is instantly remineralized into dissolved nutrients. Furthermore, algae diffuse at rate d, and sink with speed v. The water column has reflective boundary conditions at the top described by

$$\left[vA_j - d\frac{\partial A_j}{\partial z}\right]_{z=0} = 0, \quad \left.\frac{\partial R}{\partial z}\right|_{z=0} = 0. \tag{E.4}$$

Algae at the bottom of the water column ($z_{\max}$) sink into the sediment as described by

$$\left[vA_j - d\frac{\partial A_j}{\partial z}\right]_{z=z_{\max}} = vA_j(z_{\max}), \tag{E.5}$$

where algal nutrients become part of the sediment nutrient stock R_s. Sedimented nutrients are mineralized and returned to the water column at rate r, yielding the rate of change

$$\frac{\mathrm{d}R_s}{\mathrm{d}t} = -rR_s + qvA(z_{\max}), \tag{E.6}$$

and the boundary condition for dissolved nutrients at the bottom of the water column

$$\left.\frac{\partial R}{\partial z}\right|_{z=z_{\max}} = \frac{r}{d}R_s. \tag{E.7}$$

All state variables and parameters are listed with units in Table E.1.

Table E.1: Parameters and state variables in Example 2.

Quantity	Definition	Value/range	units
A_j	Concentration of phenotype j	10*	mg C m^{-3}
R	Nutrient concentration	10*	mg R m^{-3}
R_s	Sedimented nutrients	100*	mg R m^{-2}
I	Light intensity		μmol photons m^{-2} s^{-1}
l_{bg}	Background mortality	0.2	day^{-1}
v	Sinking speed	(0.01, 3.0)	m day^{-1}
d	Turbulent diffusion coefficient	10	m^2 day^{-1}
q	Algal nutrient quota	0.02	mg R mg C^{-1}
r	Specific mineralization rate of sedimented nutrients	0.02	day^{-1}
$G_{\max}$	Maximum specific algal production	1.08	day^{-1}
M_j	Nutrient uptake half-saturation constant of phenotype j	$(0, \infty)$	mg R m^{-3}
H_j	Light uptake half-saturation constant of phenotype j	$(0, \infty)$	μmol photons m^{-2} s^{-1}
I_0	Light intensity at the surface	1400	μmol photons m^{-2} s^{-1}
k	Algal light attenuation coefficient	0.0003	m^2 mg C^{-1}
k_{bg}	Background light attenuation coefficient	0.1	m^{-1}
$z_{\max}$	Water column depth	50	m
M_0	Baseline nutrient half-saturation constant	0.5	mg R m^{-3}
H_0	Baseline light half-saturation constant	50	μmol photons m^{-2} s^{-1}
χ_j	Trait value of phenotype j	$(0, \infty)$	-

Note: Values marked with * are initial conditions. The values of the parameters are chosen to provide a good illustration, rather than to represent any specific natural conditions.

We introduce evolutionary dynamics to the system by letting the half-saturation constants for light and nutrient uptake be functions of an evolvable trait χ_j so that $M_j = M_0/\chi_j$ and $H_j = H_0\chi_j$. In order to study the evolutionary dynamics of the system, one would want to use Eq. 10 to calculate the selection gradient, but the advection term in the operator $F_j = G_j - l_{\mathrm{bg}} - v\partial/\partial z + d\partial^2/\partial z^2$ prevents it from being on reaction-diffusion form. It is however possible to transform equations E.1a, E.4, and E.5 into the right form using the variable transform used by e.g., Ryabov and Blasius

(2008). To do this we note that a reaction-advection-diffusion equation can be written in the general form

$$\frac{\partial A(\mathbf{x},t)}{\partial t} = FA = (G - \mathbf{v}\cdot\nabla + d\Delta)A, \tag{E.8}$$

where the advection term $-\mathbf{v}\cdot\nabla$, describing a directional flow, makes the operator F not self-adjoint. Yet, for constant velocity fields $\mathbf{v}$, it can be transformed into reaction-diffusion form by introducing the variable transformation

$$A = \tilde{A}\exp\left(\frac{\mathbf{v}\cdot\mathbf{x}}{2d}\right), \tag{E.9}$$

which results in the transformed equation

$$\frac{\partial \tilde{A}}{\partial t} = (G - \frac{|\mathbf{v}|^2}{4d} + d\Delta)\tilde{A} = \tilde{F}\tilde{A}. \tag{E.10}$$

This equation is now on the right form. The boundary conditions of the system must be transformed using the same variable transform, and the operator $\tilde{F}$ must be checked to be self-adjoint with the new boundary conditions.

Using this transform for the sinking algae system we get:

$$\frac{\partial \tilde{A}_j(z,t)}{\partial t} = G_j\tilde{A}_j - l_{\mathrm{bg}}\tilde{A}_j - \frac{v^2}{4d}\tilde{A}_j + d\frac{\partial^2\tilde{A}_j}{\partial z^2} = \tilde{F}_j\tilde{A}_j \tag{E.11a}$$

$$\left.\frac{\partial \tilde{A}_j}{\partial z}\right|_{z=0} = \frac{v}{2d}\tilde{A}_j(0), \quad \left.\frac{\partial \tilde{A}_j}{\partial z}\right|_{z=z_{\max}} = -\frac{v}{2d}\tilde{A}_j(z_{\max}). \tag{E.11b}$$

Under these transformed boundary conditions the operator $\tilde{F}_j$ is of the right form, and Eq. 10 can be used to investigate the evolutionary dynamics of the sinking algae. One can check that the transformed operator fulfills the criterion of self-adjointedness (Appendix A). In this example, there are also auxiliary equations describing the dissolved and sedimented nutrient dynamics. However, these equations still depend on the untransformed density A, and need not be transformed. Alternatively, they can be rewritten using Eq. E.9 so that e.g., the dissolved nutrient dynamics become

$$\frac{\partial R(z,t)}{\partial t} = -q\sum_j G_j(R,I)\tilde{A}_j\exp\left(\frac{vz}{2d}\right) + ql_{\mathrm{bg}}\sum_j \tilde{A}_j\exp\left(\frac{vz}{2d}\right) + d\frac{\partial^2 R}{\partial z^2}. \tag{E.12}$$

Online Appendix F: An analytically treatable two-patch model

This example is designed to be so simple as to be analytically tractable. Consequently, our method yields an analytical expression for the selection gradient from which both the direction of selection and the endpoint of trait evolution can be directly inferred.

The purpose of the example is to provide the reader interested in understanding the application of the techniques a minimal example to which they are applicable.

The system consists of two discrete patches containing S phenotypes, of which phenotype j has the densities $\mathbf{y}_j = (y_{j1}, y_{j2})$ in patches 1 and 2, respectively. We assume that patch 1 is habitable and thus a potential source habitat, that patch 2 is an uninhabitable sink habitat, and that the two patches are linked by passive transport. This yields the dynamic equations:

$$\frac{\mathrm{d}y_{j1}}{\mathrm{d}t} = g_j y_{j1}(1 - \sum_{k=1}^{S} y_{k1}) + d(y_{j2} - y_{j1}) \tag{F.1}$$

$$\frac{\mathrm{d}y_{j2}}{\mathrm{d}t} = -\mu_j y_{j2} + d(y_{j1} - y_{j2}). \tag{F.2}$$

Here g_j is the intrinsic growth rate of phenotype j in patch 1, which is discounted by summed intra- and inter-phenotype competition, μ_j is the mortality rate of phenotype j in patch 2, and d is an environmentally determined (i.e., non-evolving) rate of passive transport between the two patches, assumed to be equal for all phenotypes. When the system contains a single resident phenotype with density $\mathbf{y}_\mathrm{r}$ the equations can be expressed in matrix form as:

$$\frac{\mathrm{d}\mathbf{y}_\mathrm{r}}{\mathrm{d}t} = F_\mathrm{r}\mathbf{y}_\mathrm{r} = \begin{bmatrix} g_\mathrm{r}(1 - y_{\mathrm{r}1}) - d & d \\ d & -\mu_\mathrm{r} - d \end{bmatrix} \begin{bmatrix} y_{\mathrm{r}1} \\ y_{\mathrm{r}2} \end{bmatrix} \tag{F.3}$$

For this single morph the system has two equilibria, which are locally stable for different conditions:

$$\mathbf{y}^* = \begin{bmatrix} 1 - \frac{d\mu}{g(d+\mu)} \\ \frac{d}{d+\mu}\left(1 - \frac{d\mu}{g(d+\mu)}\right) \end{bmatrix} \qquad \text{Stable when} \qquad g > \mu \text{ or } g < \mu,\ d < \frac{g\mu}{\mu - g} \tag{F.4}$$

$$\mathbf{y}_0^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \text{Stable when} \qquad g < \mu,\ d > \frac{g\mu}{\mu - g}. \tag{F.5}$$

The stability was determined by eigenvalue analysis of the system's Jacobian matrix. We introduce evolutionary dynamics by letting the growth and mortality rates depend on an evolvable trait χ so that $\mu(\chi) = \chi$ and $g(\chi) = 1 - (\chi - 1)^2$. The evolving trait can take on any value in the range $0 < \chi < 1$, yielding a trade-off between growth rate in patch 1 and mortality rate in patch 2. For any trait value in this range, there exists an interior equilibrium of the resident phenotype $\mathbf{y}_\mathrm{r}^* = (y_{\mathrm{r}1}^*, y_{\mathrm{r}2}^*)$ that is ecologically stable for the entire evolutionary process, since $g - \mu = \chi - \chi^2 > 0$. We can thus express the equilibrium densities $y_{\mathrm{r}1}^*$ and $y_{\mathrm{r}2}^*$ of the resident phenotype (and subsequently the selection gradient acting on trait χ_r) as functions of only two parameters: the passive transport rate d and the resident trait value χ_r:

$$\mathbf{y}_\mathrm{r}^* = \begin{bmatrix} y_{\mathrm{r}1}^* \\ y_{\mathrm{r}2}^* \end{bmatrix} = \begin{bmatrix} 1 - \frac{d}{(d+\chi_\mathrm{r})(2-\chi_\mathrm{r})} \\ \frac{d}{d+\chi_\mathrm{r}}\left(1 - \frac{d}{(d+\chi_\mathrm{r})(2-\chi_\mathrm{r})}\right) \end{bmatrix}. \tag{F.6}$$

The dynamics of a rare mutant with density $\mathbf{y}_\mathrm{m}$ and trait χ will be governed by the equation $\mathrm{d}\mathbf{y}_\mathrm{m}/\mathrm{d}t = F_\mathrm{m}\mathbf{y}_\mathrm{m}$ with

$$F_\mathrm{m} = \begin{bmatrix} g_\mathrm{m}(\chi)(1-y^*_{\mathrm{r}1}) - d & d \\ d & -\mu_\mathrm{m}(\chi) - d \end{bmatrix}. \tag{F.7}$$

The invasion fitness of the mutant could now be determined by calculating the eigenvalues of the matrix F_m and finding the largest eigenvalue. This is certainly feasible for two patches, as we show below, but the problem quickly becomes computationally intractable as the number of patches increases, and the matrix consequently grows in size. The perturbation techniques introduced in this manuscript sidesteps this difficulty by offering a straightforward formula for the selection gradient that can be evaluated based on model ingredients and equilibrium densities. To determine the selection gradient based on this formula, we first express Eq. F.7 as a sum of two matrices representing population growth and passive transport:

$$F_\mathrm{m} = \underbrace{\begin{bmatrix} g(\chi)(1-y^*_1(\chi_\mathrm{r})) & 0 \\ 0 & -\mu(\chi) \end{bmatrix}}_{G} + d\underbrace{\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}}_{\delta}. \tag{F.8}$$

Here matrices G and δ are discrete 2-patch analogues of the spatially continuous function $G(\mathbf{x})$ and the Laplacian Δ, respectively, in Eq. 6. Because the trait χ does not affect the passive transport rate d, selection acts only on the local growth rates in patches 1 and 2. Hence, we can use a matrix version of Eq. 11 to average the local selection gradients $\partial G_{11}/\partial\chi|_{\chi_\mathrm{r}}$ and $\partial G_{22}/\partial\chi|_{\chi_\mathrm{r}}$ on patches 1 and 2, respectively, where we sum over all patches rather than integrating over continuous space (see Appendices A and D for details):

$$D(\chi_\mathrm{r}) = \frac{y^{*2}_{\mathrm{r}1}\left(\frac{\partial G_{11}}{\partial\chi}\Big|_{\chi_\mathrm{r}}\right) + y^{*2}_{\mathrm{r}2}\left(\frac{\partial G_{22}}{\partial\chi}\Big|_{\chi_\mathrm{r}}\right)}{y^{*2}_{\mathrm{r}1} + y^{*2}_{\mathrm{r}2}} = \frac{\left(1-\frac{d}{2}-\chi_\mathrm{r}\right)2d\chi_\mathrm{r}}{(2-\chi_\mathrm{r})(2d^2+2d\chi_\mathrm{r}+\chi^2_\mathrm{r})}. \tag{F.9}$$

This equation shows that the selection gradient is positive for $0 < \chi_\mathrm{r} < 1 - \frac{d}{2}$ and negative for $1 - \frac{d}{2} < \chi_\mathrm{r} < 1$. Through gradual evolution, a monomorphic population will thus gradually approach the singular trait value $\chi_\mathrm{r} = 1 - d/2$, provided that $0 < d < 2$. This singular trait value can be shown to be evolutionarily stable and thus the endpoint of the evolutionary process. Hence, with increasing rate of passive transport phenotypes become increasingly adapted to survival in the sink habitat at the expense of a reduced growth rate in the source habitat.

Note that even in this highly simplified example, the analytical expressions for the local equilibrium densities (Eq. F.6) and the spatially integrated selection gradient (Eq. F.9) are rather complex. In realistic ecological and evolutionary scenarios these expressions will usually have to be evaluated numerically, but the calculation proceeds through the very same steps as in this example.

As a comparison, we compute the selection gradient directly from the dominant eigenvalue. The invasion fitness of the rare mutant is given by the dominant eigenvalue of $F_{\mathrm{m}}(\chi_{\mathrm{r}}, \chi)$, $\lambda_{\mathrm{d}}(\chi_{\mathrm{r}}, \chi)$. Since F_{m} is a 2×2 matrix we can compute this eigenvalue directly:

$$\begin{aligned}\lambda_{\mathrm{d}}(\chi_{\mathrm{r}}, \chi) = d + \chi\frac{1 + (\chi - 2)y^*_{\mathrm{r1}} - \chi}{2} + \\ + \frac{\sqrt{\chi^2(\chi-2)(y^*_{\mathrm{r1}}-1)\left[(y^*_{\mathrm{r1}}-1)(\chi-2)+2\right] + 4d^2 + \chi^2}}{2},\end{aligned} \tag{F.10}$$

where the χ_{r} dependence is in y^*_{r1}. We can then compute the selection gradient as

$$D(\chi_{\mathrm{r}}) = \frac{\partial \lambda_{\mathrm{d}}}{\partial \chi}\Big|_{\chi_{\mathrm{r}}} = \frac{d\chi_{\mathrm{r}}(d + 2\chi_{\mathrm{r}} - 2)}{(\chi_{\mathrm{r}} - 2)(2d^2 + 2d\chi_{\mathrm{r}} + \chi_{\mathrm{r}}^2)}. \tag{F.11}$$

Though not shown, these calculations are significantly lengthier than those of the method used in the main text, and for larger matrices this difference becomes increasingly pronounced.

Online Appendix G: Why there are gaps in figures 3 and 4B

Consider the bifurcation diagrams of Fig. 3 and to a lesser degree Fig. 4B of how the eco-evolutionary equilibrium distributions of traits change with changing diffusivity or sinking speeds. These both exhibit gaps of various sizes as the system transition to a state with a different number of resident phenotypes. These gaps are not the result of lacking numerical resolution, but are the consequence of one of two types of bifurcation in trait-space that occur when diffusivity or sinking speed changes. The first type, which leads to smooth transitions without gaps, is observed in Fig. 3, e.g., when the initial single phenotype transitions into two at around $d \approx 3.8 \cdot 10^{-4}$. Such transitions occur when the negative curvature of the fitness landscape around the resident phenotype changes to positive curvature as diffusivity decreases. This permits the new phenotypes to be in principle arbitrarily close to one another in trait-space for an appropriate value of d, leading to a bifurcation without gaps.

The second type of bifurcation, observed e.g., when two phenotypes transition into three in Fig. 3 at around $d \approx 2.9 \cdot 10^{-4}$, occurs when niche space opens up in-between two phenotypes as the fitness landscape is deformed. Fig. G.1 depicts a sequence of fitness landscapes around that bifurcation point showing how the deformation of the fitness landscape leads to the sudden availability of niche space as the central peak reaches above an invasion fitness of zero. For further details, see Geritz et al. (1999).

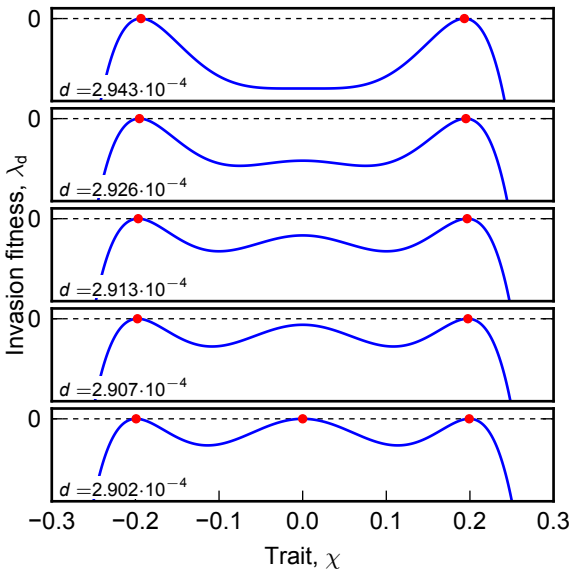

Figure G.1: Fitness landscapes of consumer ensembles at eco-evolutionary equilibrium for values of the diffusivity around the bifurcation point between two and three consumers in Fig. 3. The value of the rate of diffusion is indicated at the bottom left in each panel, and the red dots mark the trait values of the resident consumers in the system. The sequence of fitness landscapes show how as diffusivity decreases from the top to the bottom landscape, niche space suddenly opens up in the middle of the resident consumers as the fitness landscape is deformed and the peak in the middle reaches above $\lambda_d = 0$. This phenomenon explains the many gaps seen in the bifurcation diagram in Fig. 3. The same mechanism is responsible for the red phenotype in Fig. 4B appearing and disappearing in the ensemble.

References

Ajar, É. 2003. Analysis of disruptive selection in subdivided populations. BMC Evolutionary Biology 3:22.

Anderson, K. E., R. M. Nisbet, S. Diehl, and S. D. Cooper. 2005. Scaling population responses to spatial environmental variability in advection-dominated systems. Ecology Letters 8:933–943.

Barton, N. H. 1983. Multilocus clines. Evolution 37:454–471.

Barton, N. H. 1999. Clines in polygenic traits. Genetical research 74:223–236.

Birand, A., A. Vose, and S. Gavrilets. 2012. Patterns of species ranges, speciation, and extinction. The American Naturalist 179:1–21.

Blondel, J., D. W. Thomas, A. Charmantier, P. Perret, P. Bourgault, and M. M. Lambrechts. 2006. A thirty-year study of phenotypic and genetic variation of blue tits in mediterranean habitat mosaics. Bioscience 56:661–673.

Bolker, B., and S. W. Pacala. 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. Theoretical population biology 52:179–197.

Britton, N. F., et al. 1986. Reaction-diffusion equations and their applications to biology. Academic Press.

Cantrell, R. S., and C. Cosner. 2004. Spatial ecology via reaction-diffusion equations. John Wiley & Sons.

Case, T. J., and M. L. Taper. 2000. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. The American Naturalist 155:583–605.

Caswell, H. 2001. Matrix Population Models. 2nd ed. Sinauer Associates.

Champagnat, N. 2003. Convergence of adaptive dynamics n-morphic jump processes to the canonical equation and degenerate diffusion approximation. Prépublication de l'Université de Nanterre (Paris X) no 3.

Day, T. 2001. Population structure inhibits evolutionary diversification under competition for resources. Pages 71–86 *in* Microevolution Rate, Pattern, Process. Springer.

de Mazancourt, C., and U. Dieckmann. 2004. Trade-off geometries and frequency-dependent selection. The American Naturalist 164:765–778.

Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of mathematical biology 34:579–612.

Dockery, J., V. Hutson, K. Mischaikow, and M. Pernarowski. 1998. The evolution of slow dispersal rates: a reaction diffusion model. Journal of Mathematical Biology 37:61–83.

Doebeli, M., and U. Dieckmann. 2003. Speciation Along Environmental Gradients. Nature 421:259–264.

Durrett, R., and S. Levin. 1994. The importance of being discrete (and spatial). Theoretical population biology 46:363–394.

Eberl, H. J., D. F. Parker, and M. Van Loosdrecht. 2001. A new deterministic spatio-temporal continuum model for biofilm development. Computational and Mathematical Methods in Medicine 3:161–175.

Eldredge, N., J. N. Thompson, P. M. Brakefield, S. Gavrilets, D. Jablonski, J. B. Jackson, R. E. Lenski, B. S. Lieberman, M. A. McPeek, and W. Miller III. 2005. The dynamics of evolutionary stasis. Paleobiology 31:133–145.

Evans, L. C. 2010. Partial Differential Equations. 2nd ed. American Mathematical Society.

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. Annual review of genetics 10:253–280.

Ferrière, R., and J.-F. Le Galliard. 2001. Invasion Fitness and Adaptive Dynamics in Spatial Population Models. Pages 57–79 *in* J. Clobert, E. Danchin, A. A. Dhondt, and J. D. Nichols, eds. Dispersal. Oxford University Press.

Feynman, R. 1939. Forces in molecules. Physical Review 56:340.

Fisher, R. A. 1937. The wave of advance of advantageous genes. Annals of eugenics 7:355–369.

Fisher, R. A. 1950. Gene frequencies in a cline determined by selection and diffusion. Biometrics 6:353–361.

Gavrilets, S., and A. Vose. 2005. Dynamic patterns of adaptive radiation. Proceedings of the National academy of Sciences of the United States of America 102:18040–18045.

Geritz, S., E. Kisdi, G. Meszéna, and J. Metz. 1998. Evolutionarily Singular Strategies and the Adaptive Growth and Branching of the Evolutionary Tree. Evolutionary Ecology 12:35–57.

Geritz, S. A., E. van der Meijden, and J. A. Metz. 1999. Evolutionary dynamics of seed size and seedling competitive ability. Theoretical population biology 55:324–343.

Haldane, J. 1948. The theory of a cline. Journal of genetics 48:277–284.

Haller, B. C., R. Mazzucco, and U. Dieckmann. 2013. Evolutionary branching in complex landscapes. The American Naturalist 182:E127–E141.

Hastings, A. 1983. Can spatial variation alone lead to selection for dispersal? Theoretical Population Biology 24:244–251.

Hellmann, H. 1933. Zur rolle der kinetischen elektronenenergie für die zwischenatomaren kräfte. Zeitschrift für Physik 85:180–190.

Holmes, E. E., M. A. Lewis, J. Banks, and R. Veit. 1994. Partial differential equations in ecology: spatial interactions and population dynamics. Ecology 75:17–29.

Huisman, J., M. Arrayás, U. Ebert, and B. Sommeijer. 2002. How do sinking phytoplankton species manage to persist? The American Naturalist 159:245–254.

Jäger, C. G., S. Diehl, and M. Emans. 2010. Physical Determinants of Phytoplankton Production, Algal Stoichiometry, and Vertical Nutrient Fluxes. The American Naturalist 175:E91–E104.

Kingsolver, J. G., and S. E. Diamond. 2011. Phenotypic selection in natural populations: what limits directional selection? The American Naturalist 177:346–357.

Kirkpatrick, M., and N. H. Barton. 1997. Evolution of a species' range. The American Naturalist 150:1–23.

Kooijman, S. 1998. The synthesizing unit as model for the stoichiometric fusion and branching of metabolic fluxes. Biophysical chemistry 73:179–188.

Kubisch, A., R. D. Holt, H.-J. Poethke, and E. A. Fronhofer. 2014. Where am i and why? synthesizing range biology and the eco-evolutionary dynamics of dispersal. Oikos 123:5–22.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. Evolution 33:402–416.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Law, R., and U. Dieckmann. 2000. Moment approximations of individual-based models. Pages 252–270 *in* U. Dieckmann, R. Law, and J. Metz, eds. The geometry of ecological interactions: simplifying spatial complexity. Cambridge University Press.

Le Galliard, J.-F., R. Ferrière, and U. Dieckmann. 2003. The adaptive dynamics of altruism in spatially heterogeneous populations. Evolution 57:1–17.

Lehmann, L., N. Perrin, and F. Rousset. 2006. Population demography and the evolution of helping behaviors. Evolution 60:1137–1151.

Lion, S. 2015. Moment equations in spatial evolutionary ecology. Journal of theoretical biology .

Lion, S., and M. van Baalen. 2008. Self-structuring in spatial evolutionary ecology. Ecology Letters 11:277–295.

Lion, S., and M. van Baalen. 2009. The Evolution of Juvenile-Adult Interactions in Populations Structured in Age and Space. Theoretical Population Biology 76:132–145.

Merilä, J., B. Sheldon, and L. Kruuk. 2001. Explaining stasis: microevolutionary studies in natural populations. Genetica 112:199–222.

Metz, J., R. Nisbet, and S. Geritz. 1992. How Should We Define 'Fitness' for General Ecological Scenarios? Trends in Ecology & Evolution 7:198–202.

Mágori, K., P. Szabó, F. Mizera, and G. Meszéna. 2005. Adaptive Dynamics on a Lattice: Role of Spatiality in Competition, Co-Existence and Evolutionary Branching. Evolutionary Ecology Research 7:1–21.

Mizera, F., and G. Meszéna. 2003. Spatial Niche Packing, Character Displacement and Adaptive Speciation along an Environmental Gradient. Evolutionary Ecology Research 5:1–20.

Mollison, D. 1991. Dependence of epidemic and population velocities on basic parameters. Mathematical biosciences 107:255–287.

Monod, J. 1949. The growth of bacterial cultures. Annual Reviews in Microbiology 3:371–394.

Nagylaki, T. 1975. Conditions for the existence of clines. Genetics 80:595–615.

Norberg, J., M. C. Urban, M. Vellend, C. A. Klausmeier, and N. Loeuille. 2012. Eco-evolutionary responses of biodiversity to climate change. Nature Climate Change 2:747–751.

North, A., S. Cornell, and O. Ovaskainen. 2011. Evolutionary responses of dispersal distance to landscape structure and habitat loss. Evolution 65:1739–1751.

Ohtsuki, H., C. Hauert, E. Lieberman, and M. A. Nowak. 2006. A simple rule for the evolution of cooperation on graphs and social networks. Nature 441:502–505.

Pechstein, C., and R. Scheichl. 2011. Weighted poincaré inequalities and applications in domain decomposition. Pages 197–204 *in* Domain decomposition methods in science and engineering XIX. Springer.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press.

Rousset, F. 2004. Genetic structure and selection in subdivided populations. No. 40 in Monographs in Population Biology. Princeton University Press.

Rousset, F., and O. Ronce. 2004. Inclusive fitness for traits affecting metapopulation demography. Theoretical population biology 65:127–141.

Ryabov, A., and B. Blasius. 2008. Population growth and persistence in a heterogeneous environment: the role of diffusion and advection. Math. Model. Nat. Phenom 3:42–86.

Sakurai, J. J., and S. F. Tuan. 1985. Modern quantum mechanics, vol. 1. Addison-Wesley Reading, Massachusetts.

Servedio, M. R., and S. Gavrilets. 2004. The evolution of premating isolation: local adaptation and natural and sexual selection against hybrids. Evolution 58:913–924.

Siepielski, A. M., K. M. Gotanda, M. B. Morrissey, S. E. Diamond, J. D. DiBattista, and S. M. Carlson. 2013. The spatial patterns of directional phenotypic selection. Ecology letters 16:1382–1392.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

Slatkin, M. 1975. Gene flow and selection in a two-locus system. Genetics 81:787–802.

Slatkin, M. 1978. Spatial patterns in the distributions of polygenic characters. Journal of Theoretical Biology 70:213–228.

Speirs, D. C., and W. S. Gurney. 2001. Population persistence in rivers and estuaries. Ecology 82:1219–1237.

Tao, Y., and M. Winkler. 2013. Locally bounded global solutions in a three-dimensional chemotaxis-stokes system with nonlinear diffusion. Pages 157–178 *in* Annales de l'Institut Henri Poincare (C) Non Linear Analysis. Vol. 30. Elsevier.

Thompson, J. N. 1999. Specific hypotheses on the geographic mosaic of coevolution. The American Naturalist 153:S1–S14.

Tilman, D. 1980. Resources: A Graphical-Mechanistic Approach to Competition and Predation. The American Naturalist 116:362–393.

Tilman, D. 1982. Resource Competition and Community Structure. Monographs in population biology. Princeton University Press.

Troost, T. A., B. W. Kooi, and S. A. L. M. Kooijman. 2005. Ecological specialization of mixotrophic plankton in a mixed water column. The American Naturalist 166:E45–E61.

Turchin, P. 1989. Population consequences of aggregative movement. Journal of Animal Ecology 58:75–100.

Van Baalen, M., and D. A. Rand. 1998. The unit of selection in viscous populations and the evolution of altruism. Journal of theoretical biology 193:631–648.

Waxman, D., and S. Gavrilets. 2005. Issues of terminology, gradient dynamics and the ease of sympatric speciation in adaptive dynamics. Journal of Evolutionary Biology 18:1214–1219.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.