The evolution of immigration strategies facilitates niche expansion by divergent adaptation in a structured metapopulation model[†]

Éva Kisdi*, Helene C. Weigang & Mats Gyllenberg

Department of Mathematics and Statistics
FIN-00014 University of Helsinki, Finland

*Corresponding author; eva.kisdi@helsinki.fi.

[†]This paper is dedicated to the memory of Ilkka Hanski, who inspired us to embark on this research.

Accepted for publication by The American Naturalist on 28 August 2019.

DOI: 10.1086/706258.

Please cite the journal publication:

https://www.journals.uchicago.edu/doi/pdfplus/10.1086/706258

Keywords: adaptive dynamics; dispersal; function-valued trait; habitat choice; local adaptation; peripheral niche

Abstract

Local adaptation and habitat choice are two key factors that control the distribution and diversification of species. Here we model habitat choice mechanistically as the outcome of dispersal with non-random immigration. We consider a structured metapopulation with a continuous distribution of patch types, and determine the evolutionarily stable immigration strategy as the function linking patch type to the probability of settling in the patch upon encounter. We uncover a novel mechanism whereby coexisting strains that only slightly differ in their local adaptation trait can evolve substantially different immigration strategies. In turn, different habitat use selects for divergent adaptations in the two strains. We propose that the joint evolution of immigration and local adaptation can facilitate diversification, and discuss our results in the light of niche conservatism versus niche expansion.

Introduction

Dispersal is widely recognized as a key mechanism for adaptation in fragmented, patchy habitats. Accordingly, a vast body of literature examines the evolution of dispersal (see Ronce 2007; Bonte et al. 2012; Travis et al. 2012; Clobert et al. 2012; Cote et al. 2017 for reviews). Most models of dispersal evolution, however, focus only on one aspect of dispersal, the decision to leave a patch (emigration). It is commonly assumed that if they survive the hazards of transience, the dispersers settle in patches (immigrate) randomly over the entire landscape (Hamilton and May 1977; Levin et al. 1984; Olivieri et al. 1995; Mathias et al. 2001; Parvinen 2002; Massol et al. 2011; Cotto et al. 2013; Weigang and Kisdi 2015; and many others) or over some neighbourhood of the source patch defined by spatial proximity or connectivity (e.g. Comins 1982; Harada 1999; Heino and Hanski 2001; Rousset and Gandon 2002; North et al. 2011; Karisto and Kisdi 2017). Also most models exploring conditional dispersal focus on how emigration depends on local population size (Jánosi and Scheuring 1997; Travis et al. 1999; Gyllenberg and Metz 2001; Poethke and Hovestadt 2002; Kun and Scheuring 2006; Bocedi et al. 2012; Parvinen et al. 2012), kin competition (Ronce et al. 1998) or on the body size, competitive ability, or fecundity of the individual (Bonte and de la Peña 2009; Gyllenberg et al. 2008, 2011a,b; Kisdi et al. 2012).

The common assumption of random immigration contrasts sharply with the increasing realization that settlement after dispersal may often be non-random (Bowler and Benton 2005; Clobert et al. 2009; Matthysen 2012). Immigration may depend on local population size and reproductive success (e.g. Doligez et al. 2002; Garant et al. 2005; Parejo et al. 2007), habitat quality and available mates (Matter and Roland 2002), or the presence of the preferred host plant (Singer and Thomas 1996; Hanski and Singer 2001). Importantly, dispersal decisions correlate with traits under contrasting selection in different habitats. In sticklebacks, individuals of lake and stream populations that move to the opposite habitat resemble phenotypically the individuals of the target habitat (Bolnick et al. 2009). Dispersal behaviour correlates with individual preference for different temperatures in lizards (Bestion et al. 2015). The ciliate *Tetrahymena thermophila* chooses between habitats of different temperatures according to which habitat it is better adapted to, and this process enhances local adaptation (Jacob et al. 2017).

We shall focus on non-random immigration due to habitat choice, i.e., due to dispersers favouring certain patches over others. We envisage habitat choice as a process of individuals searching for habitat patches, and when a patch is found, making a decision on settlement after evaluating the local environment in the patch. For example, the butterflies *Melitaea cinxia* and *Euphrydryas editha* locate dry/barren meadows as habitat patches. When in a patch, they search for their preferred host plants; in case they cannot find the preferred host, they may either choose to accept a less preferred plant (presumably to escape dispersal-related costs) or to leave the patch and disperse further. Finding habitat patches and making the decision to stay or to leave appear to be separate, consecutive steps (Singer 2015).

Even though habitat choice is the result of searching for a suitable habitat and a decision to settle, many models investigating the effects of habitat choice ignore the process of finding the patches. Instead, they simply assume that a certain fraction of individuals end up in a certain habitat (Egas et al. 2004; Beltman et al. 2005; Ravigné et al. 2004, 2009) or individuals spend a certain fraction of their time in a certain habitat (Schreiber 2012). The same assumption was made also in population genetic models of habitat choice, with some corrections to accommodate the difficulty of finding rare habitats (e.g. Rauscher 1984; Garcia-Dorado 1986; Hedrick 1990; but see de Meeûs et al. 1993 for a more mechanistic model). Assuming that the probability of living in a certain habitat is the trait directly under selection is, however, misleading, because it ignores the uncertainty and the cost of finding the preferred habitat. In other words, substituting the mechanistic modelling of dispersal, as the process of emigration, transience, and immigration, with its presumed result, the distribution of individuals over habitats, excludes many factors that shape the evolution of habitat choice.

Some models capturing the process of transience and selective immigration assume that habitat patches are immutably good or bad for survival and reproduction (Ward 1987; Baker and Rao 2004; Stamps et al. 2005; Gyllenberg et al. 2016; Nurmi et al. 2017; Weigang 2017; Crowley et al. 2019). As several empirical studies cited above show, however, whether or not a habitat is favourable to an individual often depends on the phenotype of the individual, i.e., how well it is adapted to the local environment. Habitat choice via dispersal with non-random immigration should therefore be investigated also jointly with the evolution of local adaptation.

Traditionally, local adaptation is seen as a factor selecting against dispersal (Balkau and Feldman 1973). When immigration is random, dispersal creates a net flux from habitats where individuals are well adapted into habitats where they are not, and such source-sink dynamics disfavours dispersal (Hastings 1983). When environmental stochasticity or kin competition selects for dispersal, the joint evolution of emigration and local adaptation exhibits rich dynamics, including alternative stable states (Kisdi 2002; Nurmi and Parvinen 2011; Berdahl et al. 2015). As expected, high dispersal favours generalists, whereas low dispersal facilitates the diversification of the local adaptation trait. However, diversification can also start with evolutionary branching of the emigration strategy, not of the local adaptation trait (Kisdi 2002; Nurmi and Parvinen 2011). Dispersal distance can also diversify when different resource patches have contrasting spatial configurations (Cenzer and M'Gonigle 2019).

For non-random immigration, matching habitat choice (Edelaar et al. 2008; Jacob et al. 2015) is a plastic strategy whereby individuals choose habitats based on their own phenotypes. This assumes good prospecting capabilities including controlled movement and memory. Models vary how matching habitat choice is implemented. A simple possibility is that each individual chooses, from a certain number of patches, the patch where the locally optimal phenotype is closest to its own (Edelaar et al. 2017; Mortier et al. 2019); but this choice can be suboptimal due to crowding if many individuals choose the same patch. In other simulation models, the probability of dispersal is a prescribed function of the difference between the departure and target patches, with only its parameters evolving (Armsworth and Roughgarden 2005; Enfjäll and Leimar 2009; Berner and Thibert-Plante 2015); this function however ought to evolve freely.

In this paper, we investigate analytically how a fully flexible immigration strategy evolves jointly with a local adaptation trait in a structured metapopulation model. Habitat choice emerges mechanistically from the behaviour of dispersers, who encounter patches randomly and immigrate with an evolving probability specific to some fixed property of the patch (e.g. temperature or humidity). For a given local adaptation trait, we derive the corresponding evolutionarily stable immigration strategy, which balances the effects of the individual's match with the local environment and of local crowding. For the joint evolution of local adaptation and the immigration strategy, we show that once the metapopulation becomes dimorphic, competition drives the two strains to evolve substantially different immigration strategies even if they differ only slightly in their local adaptation trait. Different habitat choice then facilitates the divergence of the local adaptation trait as well. We suggest that by this mechanism, considerable levels of polymorphism may build up in a landscape where the patches vary continuously, and discuss the relevance of these results for niche evolution.

The model

We consider a structured metapopulation model with a large number of habitat patches (Gyllenberg and Hanski 1992; Hanski and Gyllenberg 1993; Gyllenberg et al. 1997; Gyllenberg and Metz 2001). The patches differ in a permanent quantitative property y (such as mean temperature, humidity, etc.), which we call the type of the patch, and which varies in an interval $Y = [y_{min}, y_{max}]$. The local population does not affect the patch type and thus the patch-type distribution n (obeying $\int_Y n(y)dy = 1$) is given and does not change with time.

Table 1: Important parameters and variables

Symbol	Definition
α	*per capita* patch encounter rate of dispersers
D	number of individuals in the dispersal pool per patch at steady state
$g(N, y, \theta)$	*per capita* growth rate of a population with trait θ in a patch of type y with local population density N
γ	*per capita* emigration rate
$I(y)$	immigration rate into a patch of type y at steady state
μ	per patch catastrophe rate
$n(y)$	patch type distribution
ν	*per capita* death rate during dispersal
$N(\tau, y)$	local population density in a patch of type y and age τ
$\psi(y)$	probability of immigration into a patch of type y upon encounter (evolving)
$q(\tau)$	density of patches of age τ
θ	local adaptation trait (evolving)
y	patch type
Y (Y_0, Y_1, Y_f)	set of patch types (subsets described in the text, together cover Y)

Local catastrophes kill all individuals in a patch but leave the patch habitable. We refer to the time elapsed since the last catastrophe, τ, as the age of the patch. For simplicity, we assume that catastrophes occur at a constant rate μ, up to a maximum age τ_{max}, when the local population is destroyed. At the metapopulation steady-state, therefore, patch age follows the truncated exponential distribution $q(\tau) = \mu e^{-\mu\tau}/(1 - e^{-\mu\tau_{max}})$. We introduce the truncation at τ_{max} to ensure that no immigrant can produce infinitely many descendants in one local population, but set τ_{max} high enough so that the truncation has no appreciable effect on the dynamics of the resident metapopulation.

Each individual has a heritable trait θ, which affects how well it is adapted to various patch types. Further, each individual has a heritable function ψ for its immigration strategy, which gives, for each patch type y, the probability $\psi(y)$ to settle in the patch when it is encountered during dispersal. We assume that inheritance is clonal and a mutation affects either ψ or θ but not both.

We describe the resident metapopulation at its steady-state; see section S.1 of the appendix (supplemental appendix, available online) for the full model and Table 1 for notation. Between catastrophes, a local population of the resident strategy (ψ, θ) in a patch of type y follows the dynamics

$$\frac{dN}{d\tau} = g(N, y, \theta)\, N - \gamma\, N + I(y) \tag{1}$$

with the initial condition $N(0, y) = 0$ set by the last catastrophe. Here γ is the constant emigration rate and $I(y)$ is immigration specified below. We choose the density-dependent *per capita* growth rate of θ-individuals in a patch of type y to be

$$g(N, y, \theta) = r\left(1 - \frac{N}{k}\right) - c(\theta - y)^2 \tag{2}$$

such that local populations grow logistically with an extra death rate $c(\theta - y)^2$, which increases as θ deviates from y.

Individuals who leave their patch enter a global dispersal pool, where they die at a *per capita* rate ν and encounter patches randomly at a rate α. When an individual encounters a patch, it immigrates to settle down in the patch with probability $\psi(y)$; otherwise it returns to the dispersal pool. Note that the decision depends on the type of the patch but not on the age of the patch. In other words, the individuals can judge the permanent physical properties (such as temperature or humidity) of the patch, but not the size of the local population or any other cue that changes with time since the last catastrophe and therefore informs about local population size.

At the metapopulation steady state, individuals enter the dispersal pool at the rate $E = \int_Y \int_0^{\tau_{max}} \gamma\, N(\tau, y)\, q(\tau) n(y)\, d\tau\, dy$ per patch. With D denoting the number of individuals in the dispersal pool per patch and $\bar{\psi} = \int_Y \psi(y) n(y) dy$ the average probability of settlement, individuals are removed from the dispersal pool at the rate $(\alpha\bar{\psi} + \nu)D$. Since the influx equals the removal at steady-state, the size of the dispersal pool is

$$D = \frac{E}{\alpha \bar{\psi} + \nu} \tag{3}$$

and the immigration rate into a patch of type y is $I(y) = \alpha \psi(y) D$. Notice that via $I(y)$, the solution $N(\tau, y)$ of equation (1) depends on D. With substituting $N(\tau, y)$ into E, equation (3) is an equation for D as unknown. Once this equation is solved, $N(\tau, y)$ and hence the metapopulation steady-state follow easily.

When two resident strategies (ψ_1, θ_1) and (ψ_2, θ_2) inhabit the metapopulation, the local population dynamics are given by

$$\begin{aligned} \frac{dN_1}{d\tau} &= g(N_1 + N_2, y, \theta_1)\, N_1 - \gamma\, N_1 + I_1(y) \\ \frac{dN_2}{d\tau} &= g(N_1 + N_2, y, \theta_2)\, N_2 - \gamma\, N_2 + I_2(y) \end{aligned} \tag{4}$$

where the immigration rates are calculated analogously to the monomorphic case. For our choice of g in equation (2), the local dynamics in equations (1) and (4) can be solved explicitly (see section S.2), which greatly facilitates the analysis.

To analyse the long-term evolution of the immigration strategy ψ and the local adaptation trait θ, we use the framework of adaptive dynamics. We assume that the resident metapopulation is at its steady state; due to the recurrent catastrophes, individual populations are always growing towards their local equilibrium density, but the distribution of local population sizes is constant. Mutations occur infrequently and alter the traits only slightly. A mutant is able to increase in numbers, i.e., to invade, if its basic reproduction number is greater than 1 (Gyllenberg and Metz 2001; Metz and Gyllenberg 2001).

The basic reproduction number of a mutant

To calculate the basic reproduction number $R_m(\psi_m, \theta_m, \psi, \theta)$ of a mutant strategy (ψ_m, θ_m) in the resident metapopulation with strategy (ψ, θ), we consider dispersal generations. A dispersal generation starts when a mutant individual enters the dispersal pool. If it survives to settle into a patch, this mutant will start a local population, from which emigrants return to the dispersal pool. The dispersal generation ends with the extinction of the mutant's local population (or with the death of the mutant if it failed to settle into a patch). The basic reproduction number R_m is the expected number of emigrants who enter the dispersal pool from the local population started by the initial mutant individual.

Let $\rho(y, \theta_m, \theta, \psi(y)D)$ denote the expected number of emigrants produced by a mutant with local adaptation trait θ_m upon settling into a patch of type y, given that the local population dynamics is governed by the resident's local adaptation θ and immigration rate $\alpha\psi(y)D$. In section S.3 we detail how to calculate $\rho(y, \theta_m, \theta, \psi(y)D)$ based on the population dynamic model outlined above. Obviously, the expected number of mutant emigrants increases with decreasing $|\theta_m - y|$, i.e., the better the mutant is adapted to the patch, the more descendants it will produce. Importantly, $\rho(y, \theta_m, \theta, \psi(y)D)$ decreases with $\psi(y)D$. This is because the more individuals immigrate into a patch, the higher the local population density is, which decreases the *per capita* growth rate due to competition and therefore fewer dispersers are produced by an immigrant (see section S.4 for a formal proof). For the same reason, $\rho(y, \theta_m, \theta, \psi(y)D)$ is lower if the resident is better adapted to the patch, i.e., it decreases with decreasing $|\theta - y|$.

From the dispersal pool, a mutant individual with immigration strategy ψ_m settles into a patch of type between y and $y + dy$ with probability $\alpha\psi_m(y)n(y)dy/[\alpha\bar{\psi}_m + \nu]$, and then produces on average $\rho(y, \theta_m, \theta, \psi(y)D)$ emigrants who return to the dispersal pool. The basic reproduction number of the mutant is therefore

$$R_m(\psi_m, \theta_m, \psi, \theta) = \frac{\alpha \int_Y \psi_m(y) n(y) \rho(y, \theta_m, \theta, \psi(y)D)\, dy}{\alpha \bar{\psi}_m + \nu}. \tag{5}$$

The basic reproduction number of the resident is 1; the mutant can invade if $R_m > 1$.

Evolutionarily stable immigration strategy

In this section, we consider the evolution of the immigration strategy only, i.e., we assume that mutants differ from the resident in ψ, but $\theta_m = \theta$ is fixed.

We call a patch of type y *not worthwhile* for a population with local adaptation trait θ if $\rho(y, \theta, \theta, 0) < 1$. An individual entering such a patch would not replace itself in the dispersal pool even in the absence of competition (when $\psi(y)D = 0$, the patch has no resident population), and therefore such patches should never be settled in. We denote the set of all non-worthwhile patch types by Y_0. These are the patches where the individuals are poorly adapted because y is too different from θ; the exact condition is given in Proposition 3 of section S.4. Note that while the set Y_0 depends on the local adaptation trait, it does not depend on the immigration strategy.

The worthwhile patches can be divided further. Recall that $\rho(y, \theta, \theta, \psi(y)D)$ is decreasing as a function of $\psi(y)D$. In some of the worthwhile patches, the individuals are so well adapted that $\rho(y, \theta, \theta, D) > 1$, i.e., an individual entering here will more than replace itself in the dispersal pool even if every resident who encounters this patch immigrates here ($\psi(y) = 1$). These patches should always be accepted. We denote the set of these highly beneficial patches with Y_1. Which patches belong to the set Y_1 depends not only on the local adaptation trait θ but also on the resident immigration strategy ψ, because the latter also affects the size of the dispersal pool, D.

In the remaining worthwhile patches we have $\rho(y, \theta, \theta, 0) \geqslant 1$ but $\rho(y, \theta, \theta, D) \leqslant 1$. In section

S.4, we prove that in these patches, an intermediate immigration strategy $\psi(y)$ is selected, which solves the equation $\rho(y,\theta,\theta,\psi(y)D) = 1$. We denote this set of patch types with Y_f, where 'f' is to mark that only in a fraction of encounters an individual will settle in the patch. To explain this result heuristically, suppose that the resident population is such that $\rho(y,\theta,\theta,\psi(y)D) > 1$ holds in some patch type y. In this case, a mutant that immigrates into this patch more often than the resident ($\psi_m(y) > \psi(y)$) will produce more dispersers, and hence will have a higher basic reproduction number than the resident (i.e., $R_m > 1$), which means that the mutant can invade. Conversely, if $\rho(y,\theta,\theta,\psi(y)D) < 1$, then a mutant with $\psi_m(y) < \psi(y)$ can invade the resident.

In summary, we prove in section S.4 that the strategy

$$\psi(y) = \begin{cases} 0 & \text{if } y \in Y_0 \\ f(y) & \text{if } y \in Y_f \\ 1 & \text{if } y \in Y_1, \end{cases} \tag{6}$$

where $f(y)$ is such that $\rho(y,\theta,\theta,f(y)D) = 1$ with D fulfilling $R_m(\psi,\theta,\psi,\theta) = 1$, is a (weak) evolutionarily stable strategy (ESS). The set Y_0 contains the patches with large $|y-\theta|$ (patches the individuals are least adapted to), whereas the set Y_1 has the patches with small $|y-\theta|$ (patches the individuals are best adapted to), with Y_f inbetween. Hence the shape of the ESS immigration strategy is hat-like, as shown in the examples of Figure 1.

Evolution of local adaptation

Suppose now that only the local adaptation trait θ evolves and $\psi_m = \psi$ is fixed. For brevity, we write the basic reproduction number as $R_m(\theta_m,\theta)$, suppressing its dependence on the constant immigration strategy. Since θ is a scalar, it is straightforward to characterize the evolutionarily stable local adaptation trait (Maynard Smith 1982; Geritz et al. 1998). Monomorphic evolution

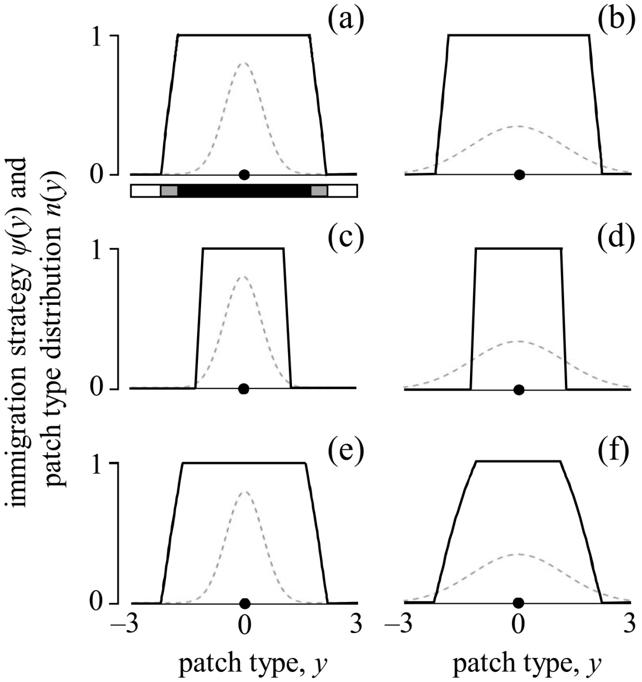

Figure 1: Evolutionarily stable immigration strategies. Black line: $\psi(y)$, the probability of immigration upon encountering a patch of type y; dashed grey line: $n(y)$, the distribution of patch types (truncated normal with variance σ^2). The local adaptation trait is $\theta = 0$ in each panel, i.e., it matches the most common patch type (black dot). In panel (a), the horizontal bar along the y-axis marks the non-worthwhile patches Y_0 with white and the worthwhile patches Y_f and Y_1 respectively with grey and black. In the Discussion, we refer to the grey part as the peripheral niche and the grey and black parts together as the fundamental niche. Parameter values: (a) $r = 5$, $k = 80$, $c = 1$, $\gamma = 2$, $\alpha = 1$, $\nu = 1$; $\mu = 0.1$, $\tau_{max} = 200$, $\sigma = 0.5$ ($D = 56.96$); (b) as in (a) except $\sigma = 1.2$ ($D = 44.82$); (c) as in (a) except $c = 3$ ($D = 50.51$); (d) as in (a) except $c = 3$ and $\sigma = 1.2$ ($D = 32.70$); (e) as in (a) except $\gamma = 0.5$ ($D = 16.82$); (f) as in (a) except $\nu = 0.1$ and $\sigma = 1.2$ ($D = 116.52$). Set of worthwhile patch types ($\psi(y) > 0$): (a), (b), (e), (f) $(-2.21, 2.21)$; (c), (d) $(-1.28, 1.28)$. Set of patch types with $\psi(y) = 1$: (a) $(-1.76, 1.76)$; (b) $(-1.86, 1.86)$; (c) $(-1.05, 1.05)$; (d) $(-1.13, 1.13)$; (e) $(-1.64, 1.64)$; (f) $(-1.09, 1.09)$.

ceases when the local adaptation trait θ satisfies the singularity condition

$$\left.\frac{\partial R_m(\theta_m, \theta)}{\partial \theta_m}\right|_{\theta_m=\theta} = 0, \tag{7}$$

and the singular trait value is locally evolutionarily stable (ESS) if

$$\left.\frac{\partial^2 R_m(\theta_m, \theta)}{\partial \theta_m^2}\right|_{\theta_m=\theta} < 0. \tag{8}$$

We discuss convergence stability in section S.5.

Monomorphic evolutionary singularities

The immigration strategy and the local adaptation trait attain a joint evolutionary singularity (ψ^*, θ^*) if equations (6) and (7) hold simultaneously. The practical difficulty in finding the singularity is that in order to obtain the ESS immigration strategy in equation (6), we need to know D, the size of the dispersal pool; but D depends on the resident immigration strategy itself. Further, the ESS immigration strategy depends on the local adaptation trait and *vice versa*. In section S.6, we describe a numerical procedure to find the joint evolutionary singularity (ψ^*, θ^*).

Since we assume that a mutation affects either the immigration strategy or the local adaptation trait but not both, we can establish trait-wise whether the singularity (ψ^*, θ^*) is evolutionarily stable. The immigration strategy ψ^* is always a weak ESS, i.e., no mutant with a different immigration strategy ψ has a positive invasion fitness, but mutants differing only for patch types in Y_f are neutral (see Proposition 2 in section S.4). If condition (8) holds, then the local adaptation trait is also an ESS, and therefore the singularity (ψ^*, θ^*) is evolutionarily stable (but see below the section Diversification for an important caveat).

To provide concrete examples, we assume the distribution of patch types is normal with mean 0 and variance σ^2, truncated to the interval $Y = [-3, 3]$. Due to symmetry, $\theta^* = 0$ is then a singular local adaptation trait, and therefore the strategies shown in Figure 1 correspond to joint monomorphic singularities, (ψ^*, θ^*) (see section S.7 on asymmetric patch type distributions).

In the example of Figure 1a, condition (8) is satisfied, and therefore we have an evolutionarily stable singularity. In Figure 1b-f, however, condition (8) is violated, so that the singularity is not stable against mutants that have a different local adaptation trait, and the local adaptation trait undergoes evolutionary branching if the immigration strategy is held fixed at ψ^*. This happens when there is a wide distribution of patch types (high σ; Figure 1b) or when mortality quickly increases with the difference from the locally best adapted trait, $|\theta - y|$ (high c; Figure 1c). In both cases, immigrants often find themselves in a patch where the generalist strategy $\theta^* = 0$ is not well adapted, so that the local adaptation trait is under disruptive selection (Meszéna et al. 1997; Kisdi and Geritz 1999; Kisdi 2002). Evolutionary stability is lost also if the emigration rate γ is low so that individuals remain in the same patch for a long time (Figure 1e), or if the patch encounter rate α is low so that many dispersers die before they could settle (not shown). In the latter two cases, the patches are more isolated, and therefore θ diversifies to match the local optimum (Meszéna et al. 1997).

In the examples of Figure 1c and 1d, the parameters are like in Figure 1a and 1b, respectively, except the value of c is higher, i.e., the within-patch mortality increases faster with the difference between the local adaptation trait θ^* and its optimum y. With high c, many patch types are not worthwhile ($\psi^*(y) = 0$ for broad intervals of y); in Figure 1d, a considerable fraction of patches are not used by the population. Despite the range of patch types that harbour a local population is narrower, the singularity is still not evolutionarily stable, because relatively small differences in y mean a large difference in mortality, and therefore the patch type variation can be partitioned between coexisting strains on a finer scale.

Low mortality in the dispersal pool might seem to imply that individuals can afford to be more choosy, i.e., accept only patches to which they are better adapted. The set of worthwhile patches, i.e., where immigration happens at all, is however independent of dispersal mortality (Proposition 3 in section S.4). The evolutionarily stable immigration strategy in equation (6) depends on dispersal mortality exclusively via D, the size of the dispersal pool. With low mortality, the size of the dispersal pool is large, so that the worthwhile patches receive many immigrants and competition is strong within the patches. This decreases the benefit from immigration and therefore decreases $\psi^*(y)$ preserving its positivity; and also narrows the set of patch types with $\psi^*(y) = 1$ (compare Figure 1f with Figure 1b). Dispersal mortality thus acts by influencing within-patch competition rather than directly.

The same holds for the distribution of patch types. With a wider distribution of patches, one might expect the dispersers to accept more patch types to avoid a long stay in the dispersal pool with its risk of mortality. Once again, we find that the set of worthwhile patches is independent of the patch type distribution, and the evolutionarily stable immigration strategy is influenced only via D, i.e., only via within-patch competition. A wider distribution means that many patches have only low population density, because even though they are worthwhile, the population is not well adapted to the patch. As a consequence, there are fewer emigrants and the dispersal pool is smaller. Fewer immigrants alleviate competition and therefore increase $\psi^*(y)$ in the worthwhile patches, as well as broaden the set of patches with $\psi^*(y) = 1$. In Figure 1, however, the difference is small.

Dimorphic populations

Suppose that the metapopulation contains two resident strategies. To find the ESS immigration strategies ψ_1 and ψ_2 for fixed values of θ_1 and θ_2, our starting point is that the first resident produces more dispersers than the second resident in a patch of type y if it is better adapted to this patch type, i.e., if $|y - \theta_1| < |y - \theta_2|$. Since one or the other resident is always better adapted to the patch (except for the single point $y = (\theta_1 + \theta_2)/2$), they cannot both have $\rho = 1$ emigrants from the patch, and therefore they cannot both have intermediate probabilities ($0 < \psi_{1,2}(y) < 1$) of immigration into a patch of type y at the ESS (cf. the section "Evolutionarily stable immigra-

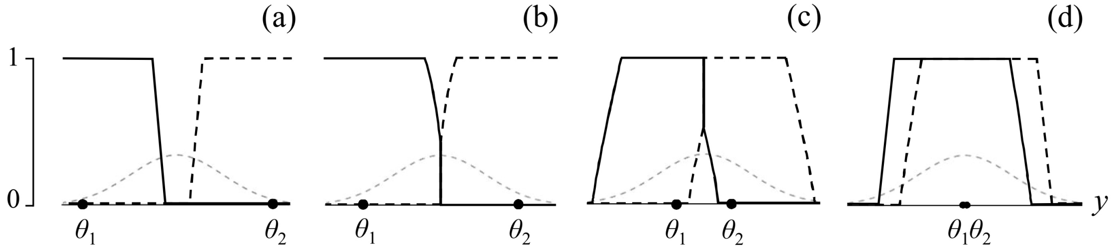

Figure 2: Evolutionarily stable immigration strategies in dimorphic populations. Black line: $\psi_1(y)$; dashed black line: $\psi_2(y)$; dashed grey line: the distribution of patch types, $n(y)$ (truncated normal with $\sigma = 1.2$). Parameter values: $r = 5$, $k = 80$, $c = 1$, $\gamma = 5$, $\alpha = 1$, $\nu = 0.1$, $\mu = 0.1$, $\tau_{max} = 200$. Resident strategies: (a) $\theta_{1,2} = \pm 2.5$ ($D_{1,2} = 126.7$); (b) $\theta_{1,2} = \pm 2$ ($D_{1,2} = 178.9$); (c) $\theta_{1,2} = \pm 0.7$ ($D_{1,2} = 227.5$); (d) $\theta_{1,2} = \pm 0.05$ ($D_{1,2} = 135.7$).

tion strategy").

With this in mind, we can proceed similarly to the monomorphic case and partition the set of patch types Y as follows. Patch types that are not worthwhile for either of the two residents are in the subset Y_{00}. For the rest, consider the patches where the first resident is better adapted. If the first resident achieves $\rho = 1$ at an immigration probability less than 1, then, given the immigration of the first resident, the second (less adapted) resident would receive $\rho < 1$ and therefore rejects the patch; these patch types are in Y_{f0}. Patches where the first resident immigrates with probability 1 (and hence has $\rho > 1$) can be, given the presence of the first resident, not worthwhile for the second resident (Y_{10}), or worthwhile but with $\rho = 1$ for the second resident at an intermediate probability of immigration (Y_{1f}), or good enough also for the second resident to achieve $\rho > 1$ with full immigration (Y_{11}). The roles are reversed in patches where the second resident is better adapted.

We give the formulas of the evolutionarily stable immigration strategies in dimorphic populations, along with a practical way of finding the ESS, in section S.8. Figure 2 shows examples with local adaptation traits fixed at symmetric values ($\theta_2 = -\theta_1$). If θ_1 and θ_2 are far from each other, then the two residents' worthwhile patch types do not overlap even in their monomorphic populations. The two residents thus never immigrate in the same patch and therefore they are independent, each behaving as in its monomorphic population (the middle of the patch type range belongs to Y_{00}; Figure 2a). If θ_1 and θ_2 are somewhat closer, then the residents' worthwhile patch types start to overlap; but since they cannot both have an intermediate immigration probability, the distribution of the two residents is abutting such that the first resident immigrates only into patches with $y < 0$, whereas the second resident only into patches with $y > 0$ (the subsets Y_{f0} an Y_{0f} are adjacent; Figure 2b). With θ_1 and θ_2 even closer to each other, the range of patch types used by the two residents overlap, but the probability of immigration drops at $y = 0$ discontinuously from 1 to a lower positive value (the subsets Y_{1f} and Y_{f1} are adjacent; Figure 2c). If θ_1 and θ_2 are sufficiently close to each other, then there is a range of patches where both residents immigrate with probability 1 (the subset Y_{11} occupies the middle of the range; Figure 2d).

As Figure 2d suggests, the difference between the evolutionarily stable immigration strategies of the two residents does not vanish even if the two residents have virtually the same local adaptation trait. This is because the two residents cannot have the same intermediate immigration probability, no matter by how little they differ in their local adaptation trait. Figure 3a shows what happens when θ_1 and θ_2 become identical. In the bottom panel, we plot the evolutionarily stable immigration strategies when the two residents differ only by an infinitesimal amount (this is very similar to Figure 2d), compared to the immigration strategy of the monomorphic ESS in the top panel. The two residents of the dimorphic population together are equivalent to the single resident of the monomorphic population. The patch types where both residents immigrate with probability 1 are the same where the single resident immigrates with probability 1 (i.e., Y_{11} coincides with Y_1 in the middle of the range); and the patch types where neither resident immigrates into are the same that are rejected by the single resident (i.e., Y_{00} coincides with Y_0 at the far left and far right of the range). The difference between the dimorphic and monomorphic populations is in the range where the monomorphic ESS has intermediate probability of immigration (i.e., in Y_f, which consists of two intervals on each sides of the plot, between the long vertical guide lines). In the dimorphic population, the resident with $\theta^* - \epsilon$ is marginally better adapted to patches with small y, and therefore its immigration probability becomes positive first as y increases in the bottom panel of Figure 3a; as explained above, the immigration probability of the other resident can become positive only when that of the first has reached 1. In other words, the left interval of Y_f of the monomorphic population is divided into Y_{f0} and Y_{1f} in the dimorphic population. In Y_{f0}, the first resident has twice as high immigration probability as the monomorphic population, and since each of the two residents has half the size of the monomorphic dispersal pool, this means that the overall immigration rate is the same as in the monomorphic population. In Y_{1f}, the immigration probability of the second resident is such that the total immigration matches that of the monomorphic population. The same pattern occurs in the right interval of Y_f, which is split into Y_{f1} and Y_{0f} in the dimorphic population.

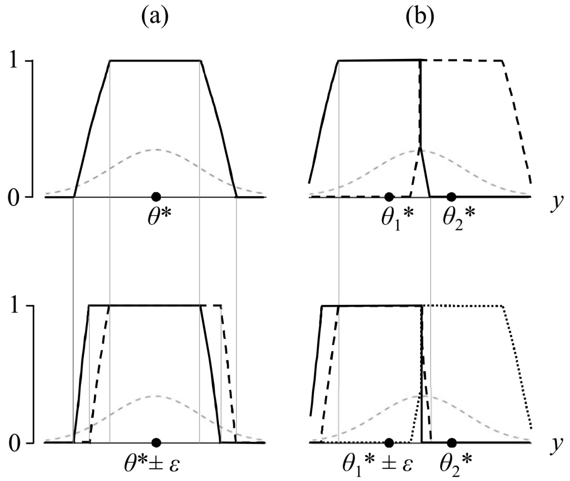

Figure 3: Comparison of the evolutionarily stable immigration strategies (parameters as in Figure 2). (a) Top: the monomorphic ESS at $\theta^* = 0$ and the corresponding ψ^* ($D = 267.4$); bottom: dimorphic population with $\theta_{1,2} = \pm\epsilon$, $\epsilon \to 0$ ($D_{1,2} = 267.4/2 = 133.7$). (b) Top: the dimorphic ESS at $\theta^*_{1,2} = \pm 0.855$ and the corresponding $\psi^*_{1,2}$ ($D_{1,2} = 235.5$); bottom: the trimorphic population of $\theta^*_1 \pm \epsilon$ and θ^*_2, with $\epsilon \to 0$. Line styles as in Figure 2, except in the bottom of panel (b), where black line: immigration strategy of the resident with $\theta^*_1 - \epsilon$; black dashed line: immigration strategy of the resident with $\theta^*_1 + \epsilon$; thin dashed line: immigration strategy of the resident with θ^*_2.

Diversification

The discontinuous change of the evolutionarily stable immigration strategy, shown in Figure 3a and described in the previous section, has far-reaching consequences for diversification.

The monomorphic singularity in Figure 3a is trait-wise an ESS, i.e., no mutant that differs in only θ can invade, and no mutant that differs in only ψ can invade (the same holds also in Figure 1a). However, when the population is not yet at the ESS, it is possible for two resident strategies with different local adaptation traits to coexist; and such coexistence can arise also with mutations of small effect when the population is close to the monomorphic singularity (Geritz et al. 1998; Geritz 2005).

Assume now that the immigration strategy ψ evolves fast to its ESS relative to the speed of evolution of the local adaptation trait θ. Suppose that θ is near, but not yet at, its evolutionarily stable value θ^*, and the immigration strategy is at the ESS corresponding to θ. In section S.9 we argue that, at least if the patch size distribution is sufficiently symmetric, there is a positive probability that an invading mutant with different θ will coexist with the former resident, forming a dimorphic population. If so, then the immigration strategies of the two residents quickly evolve to the distinctly different shapes shown in the bottom panel of Figure 3a.

This changes selection on the local adaptation trait. The resident with the smaller local adaptation trait immigrates into patches with (relatively) small y, and hence experiences selection for decreasing θ_1; and likewise, the resident with the higher value of the local adaptation trait experiences selection for increasing θ_2. Because the immigration strategy changes discontinuously from a single function ψ to the two distinctly different functions ψ_1 and ψ_2, the change in the selection gradient on θ is also discontinuous, from vanishing near the monomorphic singularity (i.e., zero for $\theta = \theta^* + \epsilon$, $\epsilon \to 0$) to a distinctly negative value for the smaller of the two residents and a distinctly positive value for the higher one (in the example of Figure 3a, the selection gradients in the dimorphic population are ± 0.04 for $\epsilon \to 0$). The opposite selection gradients drive the two residents to evolve away from each other.

We therefore conclude that the population will diversify irrespectively of whether the monomorphic singularity is trait-wise evolutionarily stable or not. Trait-wise evolutionary stability is irrelevant even though we assume that a mutation affects only the local adaptation trait or only the immigration strategy but not both. What matters is the coexistence of strains with different local adaptation traits, which can emerge also in the vicinity of an ESS; once this coexistence occurs, disruptive selection is generated by the contrasting immigration strategies. Note that our argument assumes that the immigration strategy evolves fast to its ESS, and when two coexisting residents evolve their immigration strategies, this happens such that coexistence is not lost (cf. Geritz et al. 2016). We investigate the details of the adaptive dynamics in a simplified model in section S.10, and demonstrate that diversification can happen also without a real time-scale separation between the evolution of immigration and the evolution of local adaptation.

The two residents continue to evolve until they arrive at a dimorphic evolutionary singularity (top panel of Figure 3b, a situation not far from Figure 2c; see section S.8 on how to locate the joint singularity (ψ_1^*, θ_1^*), (ψ_2^*, θ_2^*)). In this example, also the dimorphic singularity is trait-wise evolutionarily stable. If however the population becomes trimorphic as an invading mutant coexists with the former two residents, then the above scenario of diversification plays out once more. As the bottom panel of Figure 3b illustrates, two residents near θ_1^* evolve distinctly different immigration strategies, which leads to opposite nonvanishing selection gradients on them. As they evolve apart from each other, the population may proceed to a trimorphic singularity.

The present model is too complicated for us to prove that polymorphic singularities are attracting and that coexistence occurs also at polymorphic singularities or under significant asymmetries (i.e., to generalize sections S.5 and S.9). Intuitively, however, these seem likely. The argument behind a single immigration strategy splitting into two distinct strategies for residents with almost identical local adaptation traits (Figure 3a) holds for any level of polymorphism (such as trimorphism in Figure 3b), and this leads to disruptive selection on the residents with initially similar local adaptation traits. We therefore expect that high levels of polymorphism may build up by this mechanism.

Discussion

Dispersal is often seen as homogenisation. This has two important consequences for adaptation in heterogeneous landscapes. First, high dispersal is expected to favour generalists, who can cope in any environment they may arrive at (Brown and Pavlovic 1992; Meszéna et al. 1997). Second, if the organism is initially adapted to one habitat but not to others, then it may remain ’trapped’ within its original niche. Niche conservatism (Holt and Gaines 1992; Wiens et al. 2010) occurs because most individuals experience selection in the environment where population density is high, which is the environment the species is already adapted to (cf. the “multiplier effect” of McNamara and Dall 2011); deviations that would benefit the few individuals living elsewhere, at the cost of harming the majority, are selected against. In other words, gene flow from the high-abundance habitats prevents adaptation in novel habitats, which can also constrain the spatial expansion of a species (Kirkpatrick and Barton 1997; see also Polechová and Barton 2015). Alternatively, the population can undergo evolutionary branching and thereby fill the available niches (Brown and Pavlovic 1992; Meszéna et al. 1997; Kisdi and Geritz 1999), but generically only if dispersal and the difference between habitats are within certain intervals (Meszéna et al. 1997).

With habitat choice, dispersal does not lead to homogenisation. Habitat choice eliminates the selective advantage of generalists; in an extreme case, there is no gene flow across contrasting environments and therefore locally adapted specialists can freely evolve (Edelaar and Bolnick 2012). On the other hand, one could expect habitat choice only to strengthen niche conservatism, because individuals would never choose to settle in habitats where they are not yet adapted (Holt 1987). When avoiding demographic sinks, there is no selection for expanding the fundamental niche (i.e., the set of habitats where the species can maintain a viable population).

Our model demonstrates that the fundamental niche can expand also under adaptive habitat choice. Diversification is the key to niche expansion. As expected, in monomorphic populations, the evolutionarily stable immigration strategy rejects the patches where less than one emigrant would be produced. These non-worthwhile patches are the metapopulation equivalents of demographic sinks outside the fundamental niche, whereas the worthwhile patch types form the fundamental niche. Crucially, some of the worthwhile patches are accepted with a probability less than one (Figure 1, marked with a grey bar in panel (a)). These patches are demographic sources at low population density, but should they be accepted more often, they would become overcrowded and turn into sinks. The immigration strategy therefore evolves such that these patches balance on the edge, producing exactly one emigrant on average. Below we shall refer to these patch types as the *peripheral niche*, because they are not sources but also not sinks. (Recall that sources must exist even if there is no immigration into sinks, because population growth in the sources compensates for dispersal mortality. Note also that 'peripheral' is meant in the niche space, here on the horizontal axis of our figures, not necessarily in geographic space.)

Diversification starts with the coexistence of two strains with different local adaptation traits; with mutations of small phenotypic effect, these strains differ only slightly. The monomorphic singularity may or may not be an evolutionary branching point for the local adaptation trait (i.e., with the immigration strategy fixed), but this is largely irrelevant because coexistence is possible also if the local adaptation trait is not under disruptive selection as in case of evolutionary branching. We assume that when two strains coexist, they quickly evolve their immigration strategies to their respective ESSs. Since the strains have similar local adaptation traits, their fundamental niches are virtually the same. What differs is their behaviour in the peripheral niche, which they partition such that in the more extreme part of the peripheral niche, only one strain is present (Figure 3a). The partitioning of the peripheral niche breaks the initial symmetry between the two strains. The realised niche of a strain is skewed (as the bottom panel of Figure 3a shows, each strain accepts more patch types on one side than on the other), and the opposite skews of the two strains induce disruptive selection on the local adaptation trait. The skew of the realized niche is non-vanishing even if the initial two strains are infinitesimally close to each other (Figure 3a) so that, unlike during evolutionary branching, here the selection gradients are not small even at the onset of diversification. In other words, the fast evolution of the immigration strategy creates a discontinuity between selection in monomorphic and in dimorphic populations.

As the local adaptation traits of the two strains diverge, their fundamental niches move apart, such that the metapopulation of the two strains occupies a wider part of the niche space (top panel of Figure 3b) than the initial monomorphic metapopulation (top panel of Figure 3a). As long as the two strains are regarded as one species, the fundamental niche of the species is expanded. Note that there are several examples for within-species variation such that different individuals prefer different parts of the species' niche. Females of the butterfly *Melitaea cinxia* have opposing preferences for ovipositing on one or the other of the two host plants used by the species (Kuussaari et al. 2000; Hanski and Singer 2001), and host plant preference is heritable in *Euphydryas editha* (Singer and Thomas 1996). In a similar vein, individual bees have been found to prefer one or the other of two columbine species, *Aquilegia formosa* and *A. pubescens* (Fulton and Hodges 1999). Generalist species may fill their wide niche due to within-species variability under contrasting selection in different habitats, as shown for tree lizards by Taylor et al. (2018). Whether reproductive isolation evolves and the strains become separate species may correlate negatively with the overlap between their evolved habitat use.

The complexity of the present model does not allow us to reach analytical results beyond monomorphic singularities, but the above mechanism suggests that diversification can happen also at dimorphic and higher-morphic singularities (cf. Figure 3b); hence we conjecture that a considerable level of polymorphism can build up.

As mentioned above, we assume that the immigration strategy evolves fast relative to the evolution of local adaptation. In section S.10, we analyse a simplified model (with a discrete patch type distribution and no catastrophes) where we can derive the canonical equation of adaptive dynamics (Dieckmann and Law 1996; Durinx et al. 2008), and thereby explore the consequences of changing the relative speed of evolution. Assuming that the local adaptation trait attains an ESS under fixed immigration, we show that immigration indeed has to evolve fast enough for diversification. Full time scale separation is however not necessary, and the required speed seems realistic in our examples (behavioural strategies like the immigration strategy can evolve faster than morphological traits involved in local adaptation). The analysis in section S.10 provides further insight into how and why diversification fails when the immigration strategy evolves too slowly.

The partitioning of the peripheral niche via fast evolution of the immigration strategy is not the only source of disruptive selection. Depending on parameter values, the local adaptation trait may have an evolutionary branching point also with the immigration strategy fixed; and if the local adaptation trait diversifies, the two branches will evolve different immigration strategies, expanding the fundamental niche. At the onset of diversification, however, disruptive selection from within the patches accepted by both strains is much weaker than disruptive selection from the contrasting use of the peripheral niche. We thus predict that partitioning the peripheral niche is an important driver of diversification even if diversification could also happen without it.

Our model assumes a continuous distribution of patch types. This is important insofar as it naturally includes the peripheral niche; there are patches in the peripheral niche whenever the continuous distribution of patch types covers an interval as wide as the fundamental niche. With only a few patch types, the same mechanism of diversification works if some patch types are in the peripheral niche (cf. section S.10); but if each of the few patch types happens to fall outside the peripheral niche, then evolving immigration does not facilitate diversification.

Surprisingly, Ravigné et al. (2009) obtained similar results to ours in models with only two patch types. They assumed saturated patches and cost-free habitat choice, whereby a genetically determined fraction of the individuals settle in a given habitat. In their models, joint evolution of local adaptation and habitat choice results in evolutionary branching whenever the singularity is convergence stable. Compared to our model, however, the models of Ravigné et al. (2009) represent a very special case. Fecundity is assumed to be large enough to saturate the patches no matter how maladapted the individuals are, which means that a patch is never outside the fundamental niche. Further, at the monomorphic singularity, each habitat produces exactly one surviving offspring per parent (there is no net population growth in the habitats because with cost-free habitat choice, there is no need to compensate for dispersal-related mortality). Both habitats therefore behave as if they were patches in our peripheral niche. The evolutionarily stable habitat choice is only a weak ESS (as in our model, mutants differing from the ESS only in their immigration into patches of the peripheral niche are neutral), and as soon as the population becomes dimorphic, the two strains with different local adaptation traits evolve opposing habitat preferences in the same way as in our model. Costly habitat choice would however upset the balance of exactly one offspring per parent in each habitat, and would turn the weak ESS of habitat choice into a strong one, which is harder to destabilize.

To see how the results of Ravigné et al. (2009) depend on habitat choice being cost-free, it is instructive to compare their Model 1 with the soft-selection version of the population genetic model of de Meeûs et al. (1993). Both are extensions of Levene's (1953) model, but de Meeûs et al. (1993) include a costly search process for habitat choice. They show that habitat choice is disadvantageous in the sense that the ”choosy” allele is lost in every population that is monomorphic for the local adaptation trait. This is because in their model, passive dispersal without habitat choice leads to the ESS proportions found by Ravigné et al. (2009), and, due to the cost, alternative strategies are not neutral (as at the weak ESS of Ravigné et al. 2009) but selected against. If one implemented the costly search process in the adaptive dynamics Model 1 of Ravigné et al. (2009), it would prevent evolutionary branching unless the local adaptation trait alone has a branching point. It is less clear how the cost of habitat choice would change diversification in Model 3 of Ravigné et al. (2009), but since Model 3 also has a weak ESS, it can be sensitive to changing the assumption of cost-free habitat choice. In contrast, in our model the peripheral niche always exists and the ESS is always weak, so that the model predicts diversification irrespectively of the cost (as long as the metapopulation is viable). What changes with the cost is which patch types are in the peripheral niche (see Figure 1b,f), not whether the peripheral niche exists.

Both our model and those of Ravigné et al. (2009) assume clonal inheritance, i.e., no recombination between the loci that determine local adaptation and the immigration strategy. By destroying the association between preference and performance, frequent recombination would prevent the evolution of habitat choice (see Beltman et al. 2005 for a simulation study including recombination). A possible resolution lies in the evolution of genetic architecture, whereby variation can be concentrated in only a few loci (Kisdi and Geritz 1999; Kopp and Hermisson 2006; Van Doorn and Dieckmann 2006), and these loci can form supergenes with no recombination. The latter is of particular interest in the context of speciation (mate choice in place of habitat choice; see Butlin 2005 for a review and Merrill et al. 2011 for an example in *Heliconius* butterflies). For local adaptation and habitat choice, Hawthorne and Via (2001) showed that QTLs affecting preference and performance are linked in the pea aphid races specialising on clover and alfalfa.

An alternative to the heritable, genetically determined immigration strategy is that individuals choose their habitat depending on their phenotypes. Matching habitat choice (reviewed in Edelaar et al. 2008; Jacob et al. 2015) can help the maintenance genetic polymorphism (Garcia-Dorado 1986; Ravigné et al. 2004), the emergence of polymorphism via evolutionary branching (Egas et al. 2004), and the evolution of narrow-niche specialists (Mortier et al. 2019); it can also substitute for phenotypic plasticity (Edelaar et al. 2017). Importantly, matching habitat choice can facilitate the colonisation of new environments (Berner and Thibert-Plante 2015) even if these are initially sinks, provided there is enough genetic variation in the population (Holt and Barfield 2008, 2015).

In some aspects, matching habitat choice is similar to the fast-evolving immigration strategy of our model; upon a change in the local adaptation trait, matching habitat choice instantaneously adjusts where an individual settles. There are, however, important differences. If matching habitat choice is based only on the type of the patch such that individuals minimise the difference between the local environment and the environment they are best adapted to (as in the models of Edelaar et al. 2017; Mortier et al. 2019), then the “popular” patches are crowded whereas other, still suitable patches can remain unused, which is suboptimal. Local population size is an important factor for habitat choice (see e.g. Garant et al. 2005); Jacob et al. (2015) even suggest that individuals may vary in what local population density is best for them (they may be “colonisers” or “joiners”). On the other hand, habitat choice based on all relevant information may assume more knowledge than what is available in reality, especially when decisions have to be made in a short time (for example, Doligez et al. 2002 showed that immigrants use less information than residents). Our model strikes a middle ground assuming that individuals perceive only the local environment (patch type) but not local density, yet they can evolutionarily adapt to the levels of crowding associated with different patch types. Because an evolved heritable immigration strategy can associate patch types with typical population densities (or other relevant factors not directly observed), it might be even a better option than matching habitat choice that ignores density effects. As explained above, avoiding overcrowding creates the peripheral niche, which is instrumental in the new mechanism we propose for diversification.

With the local adaptation trait fixed, a patch is equally favourable or less favourable to all individuals of the population. Without the local adaptation trait evolving, our model is thus related to models of dispersal with selective immigration into good habitats. These models often assume that dispersal is a separate stage of life, which must be completed within a given period of time, but otherwise the dispersers perform a sequential search as in our model. With a finite time horizon, individuals initially immigrate only into good-quality patches, but as time runs out, they will immigrate into any patch they encounter. If fitness within a habitat is fixed, then finding the best times when an individual should start accepting inferior habitats is an optimization problem (Ward 1987; Baker and Rao 2004; Stamps et al. 2005; Crowley et al. 2019). If, however, reproduction within a patch depends on local population density as in our model, then the fitness of an individual depends on the immigration strategy followed by the rest of the population, and therefore selection is frequency-dependent. This case has previously been considered only with two patch types (rather than a continuous distribution) and a single individual per patch (rather than each patch harbouring a population with local dynamics). Nurmi et al. (2017) found the evolutionarily stable immigration strategy, i.e., the time after which dispersers should accept also the worse of the two patch types. Gyllenberg et al. (2016) and Weigang (2017) considered a variant where the individuals cannot keep track of time but accept the worse habitat with some probability that remains constant during the dispersal season; frequency-dependent selection can then yield evolutionary branching of the immigration strategy.

We assumed no time limit for dispersal and a constant rate of mortality in the dispersal pool. In reality, the mortality rate of dispersers likely increases with time since emigration, as dispersers run out of their energy reserves. This in practice limits the time spent dispersing, and, as predicted by the models described above, we expect that the dispersers should become increasingly ready to settle into any patch. Moreover, we suggest that the same pattern could arise also if dispersal mortality is in fact constant – which means that from the pattern of increasing readiness to settle, one cannot infer increasing costs of dispersal. Our model predicts that patches of the peripheral niche should be accepted with some probability less than 1. To realize this immigration strategy, the dispersers need to randomize their behaviour (i.e., sometimes immigrate into the patch and sometimes not), which can be done by using a random cue (the analogue of tossing a coin). An easily available random cue is time spent since emigration. If dispersers accept patches in the peripheral niche only after some time threshold, the result will be that only a fraction of encounters with these patches lead to immigration. We note, however, that mathematically it is not a trivial task to find the appropriate, patch-type dependent time thresholds that yield the evolutionarily stable immigration strategy of our model.

In the present model, we assumed that emigration occurs at a constant rate; to concentrate on the joint evolution of immigration and local adaptation, we did not consider the evolution of emigration strategies. The model does include catastrophes destroying local populations, which necessarily maintains positive emigration (a strain with no dispersal is not viable, because its local populations are all taken out by catastrophes). Dispersal is costly if there is a danger of settling into an inferior habitat (or the ”wrong” habitat for a habitat specialist), but the evolved immigration strategy alleviates this cost. We therefore expect that emigration is readily maintained by natural selection.

If the emigration rate is genetically fixed, then in a heterogeneous landscape of different patch types it may undergo evolutionary branching (Mathias et al. 2001, Parvinen 2002). However, if prospective immigrants are aware of the type of the patch they encounter as we assume in the present model, then the same information must also be available for the individuals who live in the patch, and therefore the emigration rate must depend at least on patch type. The emigration strategy is therefore given by a function, similarly to our immigration strategy. In a simpler model with only two types of patches and no local adaptation, Gyllenberg et al. (2016) investigated the evolution of patch-type dependent immigration strategies with fixed emigration, whereas Weigang (2017) extended this model with evolving patch-type dependent emigration; the qualitative conclusions were similar.

The residents of a patch are also likely to be able to judge the local population size. In this case, we expect that emigration evolves to a threshold strategy, so that instantaneous emigration occurs when the local population is above a critical size (Gyllenberg and Metz 2001; Metz and Gyllenberg 2001); the critical size depends on the patch type, i.e., the emigration strategy is again a function. If such a threshold strategy can be realised without error and delay, then the evolutionarily stable immigration strategy can accept every patch in the fundamental niche with probability 1. In our model, patches of the peripheral niche are accepted less often because a higher probability of settling would overcrowd the patch and would therefore select against settling there. If emigration prevents that a patch would be overcrowded, then nothing selects against settling. This would, however, not prevent diversification driven by a niche partitioning mechanism analogous to the one in our model. Coexisting strains differ in the threshold density above which they emigrate from a given patch. The locally better adapted strain can maintain a higher population density and thereby induces the emigration of the less adapted strain. This results in a split of habitat use as in our model, which implies disruptive selection on the local adaptation trait.

A technical advance in our work is that in section S.2, we explicitly solve the local population dynamics of two resident strains, assuming constant immigration and linear density-dependence (logistic model) with identical slope (identical r/k in equation (2)). The explicit solution greatly facilitated our numerical analyses, and since many metapopulation models use the same local population dynamics as we did, we believe that the explicit solution will be useful for other modellers as well.

Finally, we highlight that important mathematical tools are lacking for the analysis of evolving function-valued traits, and this impinges on the completeness of our results. We can determine the singular immigration strategy and its evolutionary stability (section S.4; see Parvinen et al. 2006, 2013 for the general theory). However, we had to assume that the immigration strategy evolves fast, it converges to its ESS, and when two strains co-evolve their immigration strategies at the beginning of diversification, this happens via a transient such that the population remains dimorphic (e.g. the transient does not include a point where one immigration strategy is distinctly inferior). To establish convergence stability, we have the canonical equation of adaptive dynamics also for function-valued traits (Metz et al. 2016 and references therein), but it contains an unknown covariance function. For vector-valued traits, strong convergence stability (Leimar 2009) ensures that the trait vector evolves to the singularity irrespectively of the covariances between its components; however, no analogue is available for function-valued traits. The sufficient conditions for evolutionary branching given by Geritz et al. (2016) for vector-valued traits ensure that the dimorphism is not lost, but these conditions are also not available for function-valued traits. While our assumptions on the convergence of the immigration strategy and the preservation of dimorphism are intuitively appealing, a mathematical proof must be left for future research.

Acknowledgements

This paper is dedicated to the memory of Ilkka Hanski, who was an inspiration for us all and a collaborator of M.G. for almost thirty years. The research presented in this paper was initiated by discussions with Ilkka. We thank Kalle Parvinen, Hans Metz, Åke Brännström, Daniel Bolnick, Joachim Hermisson, and an anonymous reviewer for comments on earlier versions, and acknowledge financial support by the Finnish Doctoral Programme in Computational Sciences (FICS), the Doctoral Programme in Mathematics and Statistics (DOMAST) and the Academy of Finland.

References

Armsworth, P. R. and J. E. Roughgarden. 2005. Disturbance induces the contrasting evolution of reinforcement and dispersiveness in directed and random movers. Evolution 59: 2083–2096.

Baker, M. B. and S. Rao. 2004. Incremental costs and benefits shape natal dispersal: theory and example with *Hemilepistus reaumuri*. Ecology 85: 1039–1051.

Balkau, B. and M. W. Feldman. 1973. Selection for migration modification. Genetics 74: 171–174.

Beltman, J. B. and J. A. J. Metz. 2005. Speciation: more likely through a genetic or through a learned habitat preference? Proc. R. Soc. B 272: 1455–1463.

Berdahl, A., C. J. Torney, E. Schertzer and S. A. Levin. 2015. On the evolutionary interplay between dispersal and local adaptation in heterogeneous environments. Evolution 69: 1390–1405.

Berner D. and X. Thibert-Plante. 2015. How mechanisms of habitat preference evolve and promote divergence with gene flow. J. evol. Biol. 28: 1641–1655.

Bestion, E., J. Clobert and J. Cote. 2015. Dispersal response to climate change: scaling down to intraspecific variation. Ecology Letters 18: 1226–1233.

Bocedi, G., J. Heinonen and J. M. J. Travis. 2012. Uncertainty and the role of information acquisition in the evolution of context-dependent emigration. Am. Nat. 179: 606–620.

Bolnick, D. I., L. K. Snowberg, C. Patenia, W. E. Stutz, T. Ingram and O. L. Lau. 2009. Phenotype-dependent native habitat preference facilitates divergence between parapatrick lake and stream stickleback. Evolution 63: 2004–2016.

Bonte, D. and E. de la Peña. 2009. Evolution of body condition-dependent dispersal in metapopulations. J. evol. Biol. 22: 1242–1251.

Bonte, D., H. Van Dyck, J. M. Bullock, A. Coulon, M. M. Delgado, M. Gibbs, V. Lehouck, E. Matthysen, K. Mustin, M. Säästamöinen, N. Schtickzelle, V. M. Stevens, S. Vandewoestijne, M. Baguette, K. A. Barton, T. G. Benton, A. Chaput-Bardy, J. Clobert, C. Dytham, T. Hovestadt, C. M. Meier, S. C. F. Palmer, C. Turlure and J. M. J. Travis. 2012. Costs of dispersal. Biological Reviews 87: 290–312.

Bowler, D. E. and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biological Reviews 80: 205–225.

Brown J. S. and N. B. Pavlovic. 1992. Evolution in heterogeneous environments: Effects of migration on habitat specialization. Evol. Ecol. 6: 360–382.

Butlin, R. K. 2005. Recombination and speciation. Mol. Ecol. 14: 2621–2635.

Cenzer M. and L. K. M'Gonigle. 2019. Local adaptation in dispersal in multi-resource landscapes. Evolution 73: 648–660.

Clobert, J., M. Baguette, T. G. Benton and J. M. Bullock (eds). 2012. Dispersal ecology and evolution. Oxford University Press.

Clobert, J., J.-F. Le Galliard, J. Cote, S. Meylan and M. Massot. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations.

Ecology Letters 12: 197–209.

Comins H. N. 1982. Evolutionarily stable strategies for localized dispersal in two dimensions. J. theor. Biol. 94: 579–606.

Cote J., E. Bestion, S. Jacob, J. Travis, D. Legrand and M. Baguette. 2017. Evolution of dispersal strategies and dispersal syndromes in fragmented landscapes. Ecography 40: 56–73.

Cotto O., I. Olivieri and O. Ronce. 2013. Optimal life-history schedule in a metapopulation with juvenile dispersal. J. evol. Biol. 26: 944–954.

Crowley P. H., P. C. Trimmer, O. Spiegel, S. M. Ehlman, W. S. Cuello and A. Sih. 2019. Predicting habitat choice after rapid environmental change. Am. Nat. 193: 619–632.

de Meeûs, T., Y. Michalakis, F. Renaud and I. Olivieri. 1993. Polymorphism in heterogeneous environments, evolution of habitat selection and sympatric speciation: soft and hard selection models. Evol. Ecol. 7: 175–198.

Dieckmann U. and R. Law. 1996. The dynamical theory of coevolution: A derivation from stochastic ecological processes. J. Math. Biol. 34: 579–612.

Doligez B., E. Danchin and J. Clobert. 2002. Public information and breeding habitat selection in a wild bird population. Science 297: 1168–1170.

Durinx M., J. A. J. Metz and G. Meszéna. 2008. Adaptive dynamics for physiologically structured population models. J. Math. Biol. 56: 673–742.

Edelaar, P. and D. I. Bolnick. 2012. Non-random gene flow: an underappreciated force in evolution and ecology. TREE 27: 659–665.

Edelaar P., R. Jovani and I. Gomez-Mestre. 2017. Should I change or should I go? Phenotypic plasticity and matching habitat choice in the adaptation to environmental heterogeneity. Am. Nat. 190: 506–520.

Edelaar, P., A. M. Siepielski and J. Clobert. 2008. Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. Evolution 62: 2462–2472.

Enfjäll, K., and O. Leimar. 2009. The evolution of dispersal – the importance of information about population density and habitat characteristics. Oikos 118: 291–299.

Egas, M., U. Dieckmann and M. W. Sabelis. 2004. Evolution restricts the coexistence of specialists and generalists: the role of the trade-off structure. Am. Nat. 163: 518–531.

Fulton, M. and S. A. Scott. 1999. Floral isolation between *Aquilegia formosa* and *Aquilegia pubescens*. Proc. R. Soc. Lond. B 266: 2247–2252.

Garant, D., L. E. B. Kruuk, T. A. Wilkin, R. H. McCleery and B. C. Sheldon. 2005. Evolution driven by differential dispersal within a wild bird poppulation. Nature 433: 60–65.

Garcia-Dorado, A. 1986. The effect of niche preference on polymorphism protection in a heterogeneous environment. Evolution 40: 936–945.

Geritz, S. A. H. 2005. Resident-invader dynamics and the coexistence of similar strategies. J. Math. Biol. 50: 67–82.

Geritz, S. A. H., E. Kisdi, G. Meszéna and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12: 35–57.

Geritz, S. A. H., J. A. J. Metz and C. Rueffler. 2016. Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. J. Math. Biol. 72: 1081–1099.

Gyllenberg, M. and I. Hanski. 1992. Single-species metapopulation dynamics: a structured model. Theor. Pop. Biol. 42: 35–61.

Gyllenberg, M., A. Hastings and I. Hanski. 1997. Structured metapopulation models. In: I. Hanski and M. E. Gilpin (eds.): Metapopulation Biology, Academic Press, London, pp. 93–122.

Gyllenberg, M., E. Kisdi and M. Utz. 2008. Evolution of condition-dependent dispersal under kin competition. J. Math. Biol. 57: 285–307.

Gyllenberg, M., E. Kisdi and M. Utz. 2011a. Body condition dependent dispersal in a heterogeneous environment. Theor. Pop. Biol. 79: 139–154.

Gyllenberg, M., E. Kisdi and M. Utz. 2011b. Variability within families and the evolution of body-condition-dependent dispersal. Journal of Biological Dynamics 5: 191–211.

Gyllenberg, M., E. Kisdi and H. C. Weigang. 2016. On the evolution of patch-type dependent immigration. J. theor. Biol. 395: 115–125.

Gyllenberg, M. and J. A. J. Metz. 2001. On fitness in structured metapopulations. J. Math. Biol.

268: 545–560.

Hamilton, W. and R. M. May. 1977. Dispersal in stable habitats. Nature 269: 578–581.

Hanski, I. and M. Gyllenberg. 1993. Two general metapopulation models and the core-satellite species hypothesis. Am. Nat. 142: 17–41.

Hanski, I. and M. Singer. 2001. Extinction-colonization dynamics and host-plant choice in butterfly metapopulations. Am. Nat. 158: 341–353.

Harada, Y. 1999. Short- vs. long-range disperser: The evolutionarily stable allocation in a lattice-structured habitat. J. theor. Biol. 201: 171–187.

Hastings, A. 1983. Can spatial variation alone lead to selection for dispersal? Theor. Pop. Biol. 24: 244–251.

Hawthorne, D. J. and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412: 904–907.

Hedrick, P. W. 1990. Genotypic-specific habitat selection: a new model and its application. Heredity 65: 145–149.

Heino, M. and I. Hanski. 2001. Evolution of migration rate in a spatially realistic metapopulation model. Am. Nat. 157: 495–512.

Holt, R. D. 1987. Population dynamics and evolutionary processes: the manifold roles of habitat selection. Evol. Ecol. 1: 331–347.

Holt R. D. and M. Barfield. 2008. Habitat selection and niche conservatism. Isr. J. Ecol. Evol. 54: 295–309.

Holt R. D. and M. Barfield. 2015. The influence of imperfect matching habitat choice on evolution in source-sink environments. Evol. Ecol. 29: 887–904.

Holt, R. D. and M. S. Gaines. 1992. Analysis of adaptation in heterogeneous landscapes: Implications for the evolution of fundamental niches. Evol. Ecol. 6: 433–447.

Jacob, S., E. Bestion, D. Legrand, J. Clobert and J. Cote. 2015. Habitat matching and spatial heterogeneity of phenotypes: implications for metapopulation and metacommunity functioning. Evol. Ecol. 29: 851–871.

Jacob, S., D. Legrand, A. S. Chaine, D. Bonte, N. Schtickzelle, M. Huet and J. Clobert. 2017. Gene flow favours local adaptation under habitat choice in ciliate microcosms. Nature Ecology and Evolution 1: 1407–1410.

Jánosi, I. and I. Scheuring. 1997. On the evolution of density dependent dispersal in a spatially structured population model. J. theor. Biol. 187: 397–408.

Karisto, P. and E. Kisdi. 2017. Evolution of dispersal under variable connectivity. J. theor. Biol. 419: 52–65.

Kirkpatrick, M. and N. H. Barton. 1997. Evolution of a species’ range. Am Nat 150: 1–23.

Kisdi, E. 2002. Dispersal: risk spreading versus local adaptation. Am. Nat. 159: 579–596.

Kisdi, E. and S. A. H. Geritz. 1999. Adaptive dynamics in allele space: Evolution of genetic polymorphism by small mutations in a heterogeneous environment. Evolution 53: 993–1008.

Kisdi, E., M. Utz and M. Gyllenberg. 2012. Evolution of condition-dependent dispersal. In: J. Clobert, M. Baguette, T. G. Benton and J. M. Bullock (eds): Dispersal ecology and evolution, Oxford University Press, pp. 139-151.

Kopp, M. and J. Hermisson. 2006. The evolution of genetic architecture under frequency-dependent disruptive selection. Evolution 60 :1537–1550.

Kun, Á. and I. Scheuring. 2006. The evolution of density-dependent dispersal in a noisy spatial population model. Oikos 115: 308–320.

Kuussaari, M., M. Singer and I. Hanski. 2000. Local specialization and landscape-level influence on host use in an herbivorous insect. Ecology 81: 2177–2187.

Leimar, O. 2009. Multidimensional convergence stability. Evol. Ecol. Res. 11: 191–208.

Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. Am. Nat. 87: 331–333.

Levin S. A., D. Cohen and A. Hastings. 1984. Dispersal strategies in patchy environments. Theor. Pop. Biol. 26: 165–191.

Massol F., A. Duputié, P. David, and P. Jarne. 2011. Asymmetric patch size distribution leads to disruptive selection on dispersal. Evolution 65: 490–500.

Mathias A., E. Kisdi and I. Olivieri. 2001. Divergent evolution of dispersal in a heterogeneous and variable landscape. Evolution 55: 246–259.

Matter, S. F. and J. Roland. 2002. An experimental examination of the effects of habitat quality on the dispersal and local abundance of the butterfly *Parnassius smintheus*. Ecol. Entomol. 27: 308–316.

Matthysen E. 2012. Multicausality of dispersal: A review. In: J. Clobert, M. Baguette, T. G. Benton and J. M. Bullock (eds): Dispersal ecology and evolution. Oxford University Press, Oxford, pp. 3–18.

Maynard Smith, J. 1982. Evolution and the theory of games. Cambridge University Press.

McNamara J. M. and S. R. X. Dall. 2011. The evolution of unconditional strategies via the 'multiplier effect'. Ecol. Letters 14: 237–243.

Merrill, R. M., B. Van Schooten, J. A. Scott and C. D. Jiggins. 2011. Pervasive genetic associations between traits causing reproductive isolation in Heliconius butterflies. Proc. R. Soc. B 278: 511–518.

Meszéna, G., I. Czibula and S. A. H. Geritz. 1997. Adaptive dynamics in a 2-patch environment: A toy model for allopatric and parapatric speciation. Journal of Biological Systems 5: 265–284.

Metz, J. A. J. and M. Gyllenberg. 2001. How should we define fitness in structured metapopulation models? Including an application to the calculation of evolutionarily stable dispersal strategies. Proc. R. Soc. B 268: 499–508.

Metz, J. A. J., K. Stanková and J. Johansson. 2016. The canonical equation of adaptive dynamics for life histories: From fitness-returns to selection gradients and Pontryagin's maximum principle. J. Math. Biol. 72: 1125–1152.

Mortier F., S. Jacob, M. L. Vandegehuchte and D. Bonte. 2019. Habitat choice stabilizes metapopulation dynamics by enabling ecological specialization.Oikos 128: 529–539.

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

Nurmi, T. and K. Parvinen. 2011. Joint evolution of specialization and dispersal in structured metapopulations. J. theor. Biol. 275: 78–92.

Nurmi, T., K. Parvinen and V. Selonen. 2017. The evolution of site-selection strategy during dispersal. J. theor. Biol. 425: 11–22.

Olivieri, I., Y. Michalakis and P.-H. Gouyon. 1995. Metapopulation genetics and the evolution of dispersal. Am. Nat. 146: 202–228.

Parejo, D., J. White and E. Danchin. 2007. Settlement decisions in blue tits: difference inthe use of social information according to age and individual success. Naturwissenschaften 94: 749–757.

Parvinen, K. 2002. Evolutionary branching of dispersal strategies in structured metapopulations. J. Math. Biol. 45: 106–124.

Parvinen, K., U. Dieckmann and M. Heino. 2006. Function-valued adaptive dynamics and the calculus of variations. J. Math. Biol. 52: 1–26.

Parvinen, K., M. Heino and U. Dieckmann. 2013. Function-valued adaptive dynamics and optimal control theory. J. Math. Biol. 67: 509–533.

Parvinen, K., A. Seppänen and J. D. Nagy. 2012. Evolution of complex density-dependent dispersal strategies. Bull. Math. Biol. 74: 2622–2649.

Poethke, H. J. and T. Hovestadt. 2002. Evolution of density- and patch-size-dependent dispersal rates. Proc. R. Soc. Lond. B 269: 637–645.

Polechová J. and N. H. Barton. 2015. Limits to adaptation along environmental gradients. Proc. Nat. Acad. Sci. USA 112: 6401–6406.

Rausher, M. D. 1984. The evolution of habitat preference in subdivided populations. Evolution 38: 596–608.

Ravigné, V., U. Dieckmann and I. Olivieri. 2009. Live where you thrive: joint evolution of habitat choice and local adaptation facilitates specialization and promotes diversity. Am. Nat. 174: E141–E169.

Ravigné, V., I. Olivieri and U. Dieckmann. 2004. Implications of habitat choice for protected polymorphisms. Evol. Ecol. Res. 6: 125–145.

Ronce, O. 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. Annual Review of Ecology, Evolution, and Systematics 38: 231–253.

Ronce, O., J. Clobert and M. Massot. 1998. Natal dispersal and senescence. Proc. Natl. Acad. Sci. USA 95: 600–605.

Rousset, F. and S. Gandon. 2002. Evolution of the distribution of dispersal distance under distance-dependent cost of dispersal. J. evol. Biol. 15: 515–523.

Schreiber, S. 2012. The evolution of patch selection in stochastic environments. Am. Nat. 180: 17–34.

Singer M. C. 2015. Adaptive and maladaptive consequences of "matching habitat choice": Lessons from a rapidly-evolving butterfly metapopulation. Evol. Ecol. 29: 905–925.

Singer M. C. and C. D. Thomas. 1996. Evolutionary responses of a butterfly metapopulation to human- and climate-caused environmental variation. Am. Nat. 148: S9–S39.

Stamps, J. A., V. Krishnan and M. Reid. 2005. Search costs and habitat selection by dispersers. Ecology 86: 510–518.

Taylor J. N., W. M. Ternes and M. S. Lattanzio. 2018. Natural selection favors local specialization in a widespread habitat generalist. Evolution 72: 2090–2099.

Travis, J. M. J., D. J. Murrell and C. Dytham. 1999. The evolution of density-dependent dispersal. Proc. R. Soc. Lond. B 266: 1837–1842.

Travis, J. M. J., K. Mustin, K. A. Barton, T. G. Benton, J. Clobert, M. M. Delgado, C. Dytham, T. Hovestadt, S. C. F. Palmer, H. Van Dyck and D. Bonte. 2012. Modelling dispersal: an eco-evolutionary framework incorporating emigration, movement, settlement behaviour and the multiple costs involved. Methods in Ecology and Evolution 3: 628–641.

Van Doorn, S. and U. Dieckmann. 2006. The long-term evolution of multi-locus traits under frequency-dependent disruptive selection. Evolution 60: 2226–2238.

Ward, S. 1987. Optimal habitat selection in time-limited dispersers. Am. Nat. 129: 568–579.

Weigang, H. C. 2017. Coevolution of patch-type dependent emigration and patch-type dependent immigration. J. theor. Biol. 426: 140–151.

Weigang H. C. and E. Kisdi. 2015. Evolution of dispersal under a fecundity-dispersal trade-off. J. theor. Biol. 371: 145–153.

Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, T. J. Davies, J.-A. Grytnes, S. P. Harrison, B. A. Hawkins, R. D. Holt, C. M. McCain and P. R. Stephens. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology Letters 13: 1310–1324.

Electronic Appendix to
The evolution of immigration strategies facilitates niche expansion by divergent adaptation in a structured metapopulation model

Éva Kisdi*, Helene C. Weigang & Mats Gyllenberg
Department of Mathematics and Statistics, FIN-00014 University of Helsinki, Finland
*corresponding author, eva.kisdi@helsinki.fi

The main article is published in The American Naturalist

S.1 The full model

Here we present our structured metapopulation model in detail. First we follow the lines of the classic theory of Gyllenberg and Hanksi (1992) and Hanski and Gyllenberg (1993), and then present an alternative formulation based on the patch age distribution. For simplicity, here we assume no maximum age for the patches ($\tau_{max} = \infty$), and truncate the patch age distribution at a finite τ_{max} only at the end of this section.

Let $p(t, N, y)$ denote the density of patches of type y with local population density N at time t. Consistency requires that

$$\int_0^\infty p(t, N, y) dN = n(y)$$

for all time t and for all $y \in Y$. The dynamics of the metapopulation is described by the following partial differential equation for the size distribution of local populations:

$$\frac{\partial}{\partial t} p(t, N, y) + \frac{\partial}{\partial N}[G(t, N, y) p(t, N, y)] = -\mu p(t, N, y) \tag{S.1.1}$$

where $G(t, N, y) = g(N, y, \theta)N - \gamma N + I(t, y)$ abbreviates the right hand side of equation (1) of the main text for local population dynamics with time-varying immigration $I(t,y)$, and with *per capita* growth rate $g(N,y,\theta)$ and emigration rate γ. Since $G(t,0,y) = I(t,y)$, the boundary condition of the PDE in (S.1.1) is

$$I(t,y)p(t,0,y) = \mu \int_0^\infty p(t,N,y)dN. \tag{S.1.2}$$

Note that the patch type y is not a dynamical variable but merely a label in equation (S.1.1) and in the accompanying boundary condition (S.1.2).

Let $\Omega \to \infty$ denote the total number of patches and $\Omega D(t)$ the total number of individuals in the dispersal pool, such that $D(t)$ is the number of dispersing individuals per patch as defined in the main text. A dispersing individual encounters patches at a rate α, so that it encounters one specific patch at the rate α/Ω. Multiplying with the total number of dispersers, a specific patch is encountered at the rate $\alpha D(t)$, and immigration occurs at the rate $I(t,y) = \alpha\psi(y)D(t)$. The size of the dispersal pool changes according to

$$\frac{dD(t)}{dt} = \int_Y \int_0^\infty \gamma N p(t,N,y)dNdy - \left(\alpha \int_Y \psi(y)n(y)dy + \nu\right) D(t), \tag{S.1.3}$$

where ν is the *per capita* death rate of dispersers.

Steady-state solutions are found by putting the time derivatives in equations (S.1.1) and (S.1.3) equal to zero. From equation (S.1.1), we obtain

$$\hat{p}(N,y) = \hat{p}(0,y)\frac{\hat{I}(y)}{\hat{G}(N,y)} \exp\left[-\int_0^N \frac{\mu}{\hat{G}(\xi,y)}d\xi\right]$$

for the steady-state distribution, where $\hat{I}(y)$ and $\hat{G}(N,y)$ are respectively the steady-state values of $I(t,y)$ and $G(t,N,y)$. For a given y, the support of $\hat{p}(\cdot,y)$ is the interval between 0 and $\bar{N}(y)$, the smallest root of $\hat{G}(N,y) = 0$; after a catastrophe, the local population density cannot grow higher than $\bar{N}(y)$. Substituting $\hat{G}(N,y) = g(N,y,\theta)N - \gamma N + \hat{I}(y)$ and using the normalization $\int_0^{\bar{N}(y)} p(t,N,y)dN = n(y)$ to find $\hat{p}(0,y)$, we arrive at

$$\hat{p}(N,y) = n(y)\frac{\mu}{g(N,y,\theta)N - \gamma N + \hat{I}(y)} \exp\left[-\int_0^N \frac{\mu}{g(\xi,y,\theta)\xi - \gamma\xi + \hat{I}(y)}d\xi\right]. \tag{S.1.4}$$

From equation (S.1.3), the steady-state size of the dispersal pool is

$$\hat{D} = \frac{\hat{E}}{\alpha\bar{\psi} + \nu}, \tag{S.1.5}$$

where $\hat{E} = \int_Y \int_0^{\bar{N}(y)} \gamma N \hat{p}(N,y)dNdy$ is the average emigration rate at steady state and $\bar{\psi} = \int_Y \psi(y)n(y)dy$ is the mean probability of immigration upon encountering a patch.

We now switch to using patch age, i.e., the time τ elapsed since the last catastrophe, as the structuring variable instead of local population size N. This formulation is perhaps simpler, and has advantages in the numerical analyses especially if the catastrophe rate is low, so that for each y, $\hat{p}(N,y)$ is concentrated near $\bar{N}(y)$.

Since the catastrophes occur at the constant rate μ independently of patch type and age (and therefore of population size), the steady-state patch age distribution is the exponential distribution $q(\tau) = \mu e^{-\mu\tau}$. The local population size $N(\tau,y)$ is obtained by solving $\frac{dN}{d\tau} = \hat{G}(N,y)$, i.e., equation (1) of the main text; see section S.2 below for the explicit solution. To obtain the steady-state patch distribution, note that a fraction $q(\tau)d\tau$ of the patches of type y are between ages τ and $\tau + d\tau$; the same patches have local population sizes between $N(\tau,y)$ and $N(\tau + d\tau, y)$, i.e., in the infinitesimal interval $dN = N(\tau + d\tau, y) - N(\tau, y) = \hat{G}(N(\tau,y),y)d\tau$, so that

$$\hat{p}(N,y)\hat{G}(N,y) = n(y)q(\tau)$$

when evaluated at $N = N(\tau,y)$. Since $\frac{d\tau}{dN} = \frac{1}{\hat{G}(N,y)}$, the age of the patch can be expressed with local population size as $\tau = \int_0^{N(\tau,y)} \frac{1}{\hat{G}(\xi,y)} d\xi$, which we use to write $q(\tau) = \mu \exp\left(-\int_0^{N(\tau,y)} \frac{\mu}{\hat{G}(\xi,y)} d\xi\right)$. Substituting this and $\hat{G}(N,y)$ into the above equation, we arrive at equation (S.1.4).

An advantage of the formulation based on the patch age distribution is that we do not need equation (S.1.4) to express the quantities we need in the main text, $\hat{I}(y)$, $\hat{E}$ and $\hat{D}$. Indeed, we have the steady-state average emigration rate per patch directly as

$$\hat{E} = \int_Y \int_0^\infty \gamma\, N(\tau,y)\, q(\tau) n(y)\, d\tau\, dy, \tag{S.1.6}$$

with $q(\tau) = \mu e^{-\mu\tau}$, and $\hat{D} = \frac{\hat{E}}{\alpha\bar{\psi}+\nu}$, $\hat{I}(y) = \alpha\psi(y)\hat{D}$ as before. To assume a maximum patch age τ_{max}, we simply truncate the exponential distribution as in the main text and replace infinity with τ_{max} in equation (S.1.6). In the main text, we drop the hats of $\hat{I}$, $\hat{E}$ and $\hat{D}$.

S.2 Local population dynamics

Here we derive the explicit solution of the local population dynamics for monomorphic and dimorphic populations in equations (1) and (4) of the main text, respectively, with the *per capita* growth rate given in equation (2) of the main text. We start with the dimorphic case,

$$\begin{aligned} \frac{dN_1}{d\tau} &= g(N_1 + N_2, y, \theta_1)\, N_1 - \gamma\, N_1 + I_1(y) \\ \frac{dN_2}{d\tau} &= g(N_1 + N_2, y, \theta_2)\, N_2 - \gamma\, N_2 + I_2(y) \end{aligned}$$

where $I_i(y) = \alpha\psi_i(y) D_i$ (for $i = 1, 2$), and the sizes of the dispersal pools are given by

$$D_i = \frac{E_i}{\alpha\bar{\psi}_i + \nu}$$

with

$$\begin{aligned} E_1 &= \int_Y \int_0^{\tau_{max}} \gamma N_1(\tau, y) q(\tau) n(y)\, d\tau\, dy \\ E_2 &= \int_Y \int_0^{\tau_{max}} \gamma N_2(\tau, y) q(\tau) n(y)\, d\tau\, dy. \end{aligned}$$

The solution of the monomorphic dynamics can be obtained analogously and is given explicitly in equation (S.2.9) at the end of this section.

Since we focus on the dynamics in a single patch, y is a fixed number; therefore in this section, we shorten the notation by dropping y from the arguments of $N_i(\tau, y)$ and $I_i(y)$ ($i = 1, 2$). Further, we write $\beta_i = r - c(\theta_i - y)^2 - \gamma$ and $\delta = r/k$, and introduce $N(\tau) = N_1(\tau) + N_2(\tau)$ for total population size. With these abbreviations, the within-patch population dynamics are given by

$$\begin{aligned} \frac{dN_1(\tau)}{d\tau} &= (\beta_1 - \delta N(\tau)) N_1(\tau) + I_1, \\ \frac{dN_2(\tau)}{d\tau} &= (\beta_2 - \delta N(\tau)) N_2(\tau) + I_2 \end{aligned} \tag{S.2.1}$$

with $N_1(0) = 0$, $N_2(0) = 0$.

If we assume that $N(\tau)$ is known, we can solve the above equations and get

$$\begin{aligned} N_1(\tau) &= I_1 \int_0^\tau \mathrm{e}^{\beta_1(\tau - s)} \mathrm{e}^{-\delta \int_s^\tau N(\sigma)\, d\sigma} ds, \\ N_2(\tau) &= I_2 \int_0^\tau \mathrm{e}^{\beta_2(\tau - s)} \mathrm{e}^{-\delta \int_s^\tau N(\sigma)\, d\sigma} ds. \end{aligned} \tag{S.2.2}$$

Then the total population size in the patch at time τ since the last catastrophe is

$$N(\tau) = \int_0^\tau \Big(I_1\, \mathrm{e}^{\beta_1(\tau - s)} + I_2\, \mathrm{e}^{\beta_2(\tau - s)} \Big) \mathrm{e}^{-\delta \int_s^\tau N(\sigma)\, d\sigma} ds.$$

We define

$$K(\tau) = I_1\,\mathrm{e}^{\beta_1 \tau} + I_2\,\mathrm{e}^{\beta_2 \tau} \quad \text{and}$$
$$X(\tau) = \mathrm{e}^{\delta \int_0^\tau N(\sigma)\,d\sigma}. \tag{S.2.3}$$

It follows that

$$\frac{dX(\tau)}{d\tau} = \delta N(\tau) X(\tau) \quad \text{with } X(0) = 1 \tag{S.2.4}$$

and

$$N(\tau) = \int_0^\tau K(\tau - s) \frac{X(s)}{X(\tau)}\,ds. \tag{S.2.5}$$

Substituting the latter expression into equation (S.2.4) we obtain

$$\frac{dX(\tau)}{d\tau} = \delta \int_0^\tau K(\tau - s) X(s)\,ds. \tag{S.2.6}$$

Now we take the Laplace transform on both sides of equation (S.2.6). Denoting the Laplace transform as $\hat{X}(\lambda)$, we get

$$\lambda \hat{X}(\lambda) - X(0) = \delta \left(\frac{I_1}{\lambda - \beta_1} + \frac{I_2}{\lambda - \beta_2} \right) \hat{X}(\lambda),$$

which we solve for $\hat{X}(\lambda)$:

$$\hat{X}(\lambda) = \frac{1}{\lambda - \delta \left(\frac{I_1}{\lambda - \beta_1} + \frac{I_2}{\lambda - \beta_2} \right)} = \frac{(\lambda - \beta_1)(\lambda - \beta_2)}{\lambda(\lambda - \beta_1)(\lambda - \beta_2) - \delta I_1 (\lambda - \beta_2) - \delta I_2 (\lambda - \beta_1)}. \tag{S.2.7}$$

Because the Laplace transform is a rational function in λ, its inverse can be calculated explicitly. Let λ_1, λ_2 and λ_3 denote the three roots of the cubic expression in the denominator of equation (S.2.7); note that they are real because I_1 and I_2 are positive. For the inverse Laplace transform we obtain

$$X(\tau) = \frac{\mathrm{e}^{\lambda_1 \tau}(\lambda_1 - \beta_1)(\lambda_1 - \beta_2)}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)} - \frac{\mathrm{e}^{\lambda_2 \tau}(\lambda_2 - \beta_1)(\lambda_2 - \beta_2)}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)} + \frac{\mathrm{e}^{\lambda_3 \tau}(\lambda_3 - \beta_1)(\lambda_3 - \beta_2)}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}. \tag{S.2.8}$$

Since $\mathrm{e}^{-\delta \int_s^\tau N(\sigma)\,d\sigma} = X(s)/X(\tau)$, we can write the solution in equation (S.2.2) as

$$\begin{aligned} N_1(\tau) &= I_1 \int_0^\tau \mathrm{e}^{\beta_1(\tau-s)} X(s)ds \Big/ X(\tau), \\ N_2(\tau) &= I_2 \int_0^\tau \mathrm{e}^{\beta_2(\tau-s)} X(s)ds \Big/ X(\tau). \end{aligned}$$

Substituting $X(\tau)$ from equation (S.2.8), we arrive at the explicit solution of the dimorphic dynamics in equation (S.2.1),

$$N_1(\tau) = \frac{I_1(\mathrm{e}^{\lambda_1\tau}(\lambda_1-\beta_2)(\lambda_2-\lambda_3) - \mathrm{e}^{\lambda_2\tau}(\lambda_2-\beta_2)(\lambda_1-\lambda_3) + \mathrm{e}^{\lambda_3\tau}(\lambda_3-\beta_2)(\lambda_1-\lambda_2))}{\mathrm{e}^{\lambda_1\tau}(\lambda_1-\beta_1)(\lambda_1-\beta_2)(\lambda_2-\lambda_3) - \mathrm{e}^{\lambda_2\tau}(\lambda_2-\beta_1)(\lambda_2-\beta_2)(\lambda_1-\lambda_3) + \mathrm{e}^{\lambda_3\tau}(\lambda_3-\beta_1)(\lambda_3-\beta_2)(\lambda_1-\lambda_2)},$$

$$N_2(\tau) = \frac{I_2(\mathrm{e}^{\lambda_1\tau}(\lambda_1-\beta_1)(\lambda_2-\lambda_3) - \mathrm{e}^{\lambda_2\tau}(\lambda_2-\beta_1)(\lambda_1-\lambda_3) + \mathrm{e}^{\lambda_3\tau}(\lambda_3-\beta_1)(\lambda_1-\lambda_2))}{\mathrm{e}^{\lambda_1\tau}(\lambda_1-\beta_1)(\lambda_1-\beta_2)(\lambda_2-\lambda_3) - \mathrm{e}^{\lambda_2\tau}(\lambda_2-\beta_1)(\lambda_2-\beta_2)(\lambda_1-\lambda_3) + \mathrm{e}^{\lambda_3\tau}(\lambda_3-\beta_1)(\lambda_3-\beta_2)(\lambda_1-\lambda_2)}.$$

The solution of the monomorphic dynamics in equation (1) of the main text is analogous. Since $K(\tau)$ in equation (S.2.3) has only one term for the monomorphic dynamics, the denominator of equation (S.2.7) is quadratic rather than cubic, so that the roots λ_1, λ_2 are easy to obtain. This leads to the explicit solution

$$N(\tau, y) = \frac{2I(y)\left(\mathrm{e}^{\tau A(\theta,y)} - 1\right)}{A(\theta,y)\left(\mathrm{e}^{\tau A(\theta,y)} + 1\right) - \beta(\theta,y)\left(\mathrm{e}^{\tau A(\theta,y)} - 1\right)} \tag{S.2.9}$$

where $A(\theta,y) = \sqrt{\beta(\theta,y)^2 + 4I(y)r/k}$ and $\beta(\theta,y) = r - c(\theta-y)^2 - \gamma$. As τ goes to infinity, $N(\tau,y)$ goes monotonically to its equilibrium value. Note, however, that in dimorphic populations, $N_1(\tau,y)$ and $N_2(\tau,y)$ need not be both monotonic functions of τ.

S.3 The expected number of emigrants

In this section, we detail the calculation of the expected number of emigrants produced by an immigrant. Suppose that a mutant immigrates into a patch of type y and age T. Its local dynamics is subject to demographic stochasticity, but its expected population size $M(\tau,y)$ grows according to

$$\frac{dM(\tau,y)}{d\tau} = \big(g(N(\tau,y),y,\theta_m) - \gamma\big)M(\tau,y). \tag{S.3.1}$$

The linear dynamics can be averaged over patches of the same initial conditions, so that in a metapopulation of infinite size where the mutant is present in large numbers even when it has low frequency, the within-patch demographic stochasticity is irrelevant and the expected population size can be used for the mutant dynamics. Since the mutant is rare both locally and in the dispersal pool, the growth rate of the mutant depends on the local resident density $N(\tau,y)$. As we are interested in the number of dispersers descending from a single mutant, there is no immigration term in equation (S.3.1).

Solving equation (S.3.1), we obtain that for each immigrant into a patch of type y and age T, there are on average

$$\mathrm{e}^{\int_T^\tau (g(N(\zeta,y),y,\theta_m)-\gamma)d\zeta}$$

mutants by the time the patch has reached age τ, assuming that there was no catastrophe between ages T and τ. The expected number of dispersers produced over all patch ages is given by

$$F(T,y,\theta_m,\theta,\psi(y)D) = \gamma \int_T^{\tau_{max}} \mathrm{e}^{-\mu(\tau-T)} \cdot \mathrm{e}^{\int_T^\tau (g(N(\zeta,y),y,\theta_m)-\gamma)d\zeta}\, d\tau, \tag{S.3.2}$$

where $\mathrm{e}^{-\mu(\tau-T)}$ is the probability of no catastrophe between T and τ. With finite maximum patch age τ_{max}, this quantity is finite; but notice that if we replace τ_{max} with infinity, the integral in equation (S.3.2) can become infinite. This is because the mutant by assumption has no self-limitation (its dynamics is linear) and a sufficiently good mutant can outgrow the catastrophe rate (i.e., its growth in patches not yet destroyed can permanently exceed the loss due to catastrophes). Importantly, F depends on θ and on $\psi(y)D$ because $N(\cdot,y)$, the solution of equation (1) in the main text, depends on the resident's local adaptation trait and on the immigration term $I(y) = \alpha\psi(y)D$. Integrating over the distribution of T, we obtain the expected number of emigrants produced by a mutant that immigrates into a patch of type y,

$$\rho(y, \theta_m, \theta, \psi(y)D) = \int_0^{\tau_{max}} q(T)F(T, y, \theta_m, \theta, \psi(y)D)\, dT. \tag{S.3.3}$$

S.4 Evolutionarily stable immigration strategy

To derive the ESS immigration strategy in equation (6) of the main text, we assume that all patch types exist at a positive density ($n(y) > 0$ for all $y \in Y$); obviously, the choice whether an individual should immigrate into non-existing patches is neutral. Since in this section we focus on the immigration strategy ψ and assume that $\theta_m = \theta$ is fixed, we abbreviate $\rho(y, \theta_m, \theta, \psi(y)D)$ as $\rho(y, \psi(y)D)$ and $R_m(\psi_m, \theta, \psi, \theta)$ as $R_m(\psi_m, \psi)$.

As preparation, we show that $\rho(y, \psi(y)D)$ is a decreasing function of its second argument, $\psi(y)D$. Consider a given patch, i.e., let y be fixed. Recall that, by its definition, ρ increases with the *per capita* growth rate $g(N(\tau, y), y, \theta)$, and with our choice of the *per capita* growth rate in equation (2) of the main text, $g(N(\tau, y), y, \theta)$ always decreases with $N(\tau, y)$. Hence we need to show that $N(\tau, y)$ at any fixed τ increases with $\psi(y)D$. Recall that $I(y) = \alpha\psi(y)D$, and $\psi(y)D$ enters $N(\tau, u)$ in equation (S.2.9) only through $I(y)$. Using the definition $A(\theta, y) = \sqrt{\beta(\theta, y)^2 + 4I(y)r/k}$ and taking the inverse, we can rewrite equation (S.2.9) as

$$\frac{1}{N(\tau, y)} = \left(\frac{\mathrm{e}^{\tau A(\theta,y)} + 1}{\mathrm{e}^{\tau A(\theta,y)} - 1}\right) \sqrt{\left(\frac{\beta(\theta, y)}{2I(y)}\right)^2 + \frac{r/k}{I(y)}} - \frac{\beta(\theta, y)}{2I(y)}.$$

$(\mathrm{e}^x + 1)/(\mathrm{e}^x - 1)$ is a decreasing function and for positive x, its value is greater than 1. Therefore the first factor on the right hand side above decreases with $A(\theta, y)$ for any fixed τ, and since $A(\theta, y)$ increases with $I(y)$, the first factor also decreases with $I(y)$. It is an easy exercise to show that the square root above increases faster with $1/I(y)$ than the last (negative) term. The first term on the right hand side thus decreases faster with $I(y)$ than the second, so that $1/N(\tau, y)$ decreases and hence $N(\tau, y)$ increases with $I(y)$.

The following two propositions characterise the ESS immigration strategy. Proposition 1 formulates a necessary condition for an ESS by showing that any strategy that differs from equation (6) of the main text, here reproduced in equation (S.4.1), can be invaded by a mutant. Proposition 2 shows that the strategy in equation (6) is a (weak) ESS.

Proposition 1. *The evolutionarily stable immigration strategy must be of the form*

$$\psi^*(y) = \begin{cases} 0 & \text{if } y \in Y_0 = \{y : \rho(y,0) < 1\}, \\ f(y) & \text{if } y \in Y_f = \{y : \rho(y,0) \geqslant 1 \text{ and } \rho(y,D) \leqslant 1\}, \\ 1 & \text{if } y \in Y_1 = \{y : \rho(y,D) > 1\}, \end{cases} \tag{S.4.1}$$

for almost every y, where $f(y)$ solves $\rho(y, f(y)D) = 1$ and D solves $R_m(\psi^*, \psi^*) = 1$.

(By "almost every" y, we mean that the set of patch types where an ESS differs from the strategy given in equation (S.4.1) must not represent any positive fraction of the patches. With a continuous distribution of patch types, deviations from equation (S.4.1) are possible at individual points in Y.)

Proof. From equation (5) of the main text, the basic reproduction number of the mutant with strategy $\psi + h$ in an environment set up by the resident with strategy ψ and dispersal pool size D is

$$R_m(\psi + h, \psi) = \frac{\alpha \int_Y [\psi(y) + h(y)] n(y) \rho(y, \psi(y)D)\, dy}{\alpha \int_Y [\psi(y) + h(y)] n(y)\, dy + \nu}$$

(note that h need not be infinitesimal). The mutant can invade if $R_m(\psi + h, \psi) - R_m(\psi, \psi) > 0$, i.e., if

$$R_m(\psi + h, \psi) - R_m(\psi, \psi) =$$

$$= \frac{\alpha \int_Y [\psi(y) + h(y)] n(y) \rho(y, \psi(y)D)\, dy}{\alpha(\bar{\psi} + \bar{h}) + \nu} - \frac{\alpha \int_Y \psi(y) n(y) \rho(y, \psi(y)D)\, dy}{\alpha \bar{\psi} + \nu} =$$

$$
\begin{aligned}
&= \frac{(\alpha\bar{\psi}+\nu)\alpha\int_Y \psi(y)n(y)\rho(y,\psi(y)D)\,dy - (\alpha(\bar{\psi}+\bar{h})+\nu)\alpha\int_Y \psi(y)n(y)\rho(y,\psi(y)D)\,dy}{(\alpha(\bar{\psi}+\bar{h})+\nu)(\alpha\bar{\psi}+\nu)} \\
&\quad + \frac{\alpha\int_Y h(y)n(y)\rho(y,\psi(y)D)\,dy}{\alpha(\bar{\psi}+\bar{h})+\nu} = \\
&= \frac{-\alpha\bar{h}}{\alpha(\bar{\psi}+\bar{h})+\nu}\underbrace{\frac{\alpha\int_Y \psi(y)n(y)\rho(y,\psi(y)D)\,dy}{\alpha\bar{\psi}+\nu}}_{=1} + \frac{\alpha\int_Y h(y)n(y)\rho(y,\psi(y)D)\,dy}{\alpha(\bar{\psi}+\bar{h})+\nu} = \\
&= \frac{\alpha\int_Y h(y)n(y)(\rho(y,\psi(y)D)-1)\,dy}{\alpha(\bar{\psi}+\bar{h})+\nu} > 0 \qquad \text{(S.4.2)}
\end{aligned}
$$

with $\bar{\psi} = \int_Y \psi(y)n(y)\,dy$ and $\bar{h} = \int_Y h(y)n(y)\,dy$. The underbraced formula equals 1 because it is the basic reproduction number of the resident. The denominator in the last row of equation (S.4.2) is always strictly positive (h is admissible only if $\psi(y)+h(y)$ is non-negative for all y), hence the condition for invasion is equivalent to

$$\int_Y h(y)n(y)(\rho(y,\psi(y)D)-1)\,dy > 0. \qquad \text{(S.4.3)}$$

1. Suppose that $\rho(y,\psi(y)D) < 1$ holds in some patches where $\psi(y) > 0$. Denote the set of these patch types with Y_-, and assume $\int_{Y_-} n(y)\,dy > 0$. Consider the mutant strategy $\psi + h$ with

   $$h(y) = \begin{cases} -\epsilon\psi(y) & \text{if } y \in Y_-, \\ 0 & \text{otherwise} \end{cases}$$

   where $0 < \epsilon \leqslant 1$. This choice represents an admissible mutant strategy ($\psi(y)+h(y) \geqslant 0$ for all y). Substituting into the left hand side of condition (S.4.3) yields

   $$-\epsilon\int_{Y_-} \psi(y)n(y)(\rho(y,\psi(y)D)-1)\,dy, \qquad \text{(S.4.4)}$$

   which is positive, i.e., the mutant can invade and therefore the resident cannot be an ESS.

2. Suppose now that $\rho(y,\psi(y)D) > 1$ holds in some patches where $\psi(y) < 1$. Denote the set of these patch types with Y_+, and assume $\int_{Y_+} n(y)\,dy > 0$. Consider the mutant strategy

$\psi + h$ with

$$h(y) = \begin{cases} \epsilon(1-\psi(y)) & \text{if } y \in Y_+, \\ 0 & \text{otherwise} \end{cases}$$

where $0 < \epsilon \leqslant 1$. This choice represents an admissible mutant strategy ($\psi(y) + h(y) \leqslant 1$ for all y). Substituting into the left hand side of condition (S.4.3) yields

$$\epsilon \int_{Y_+} (1-\psi(y))n(y)(\rho(y, \psi(y)D) - 1)\,dy, \tag{S.4.5}$$

which is positive, i.e., the mutant can invade and therefore the resident cannot be an ESS.

Since $\rho(y, \psi(y)D)$ is a decreasing function of its second argument, $\rho(y, 0) < 1$ implies $\rho(y, \psi(y)D) < 1$ for any positive value of $\psi(y)D$. Hence point (1) above shows that a strategy that violates the first line of equation (S.4.1) for a nonzero fraction of the patches cannot be an ESS. Similarly, $\rho(y, D) > 1$ implies $\rho(y, \psi(y)D) > 1$, and therefore point (2) above shows that a strategy that violates the last line of equation (S.4.1) cannot be an ESS. Finally, consider the set of patch types where (given the resident's D) $\rho(y, 0) \geqslant 1$ and $\rho(y, D) \leqslant 1$. By by continuity, there exists a number $f(y) \in [0, 1]$ that solves the equation $\rho(y, f(y)D) = 1$. Since $\psi(y) > f(y)$ implies $\rho(y, \psi(y)D) < 1$, if $\psi(y) > f(y)$ holds in a nonzero fraction of these patches, then point (1) above shows that the strategy cannot be an ESS. Similarly, since $\psi(y) < f(y)$ implies $\rho(y, \psi(y)D) > 1$, if $\psi(y) < f(y)$ holds in a nonzero fraction of these patches, then point (2) above shows that the strategy cannot be an ESS. Hence a strategy that violates the middle line of equation (S.4.1) cannot be an ESS. □

Proposition 2. *The strategy* ψ^* *given in equation* (S.4.1) *is a weak ESS.*

Proof. Consider a mutant $\psi^* + h$ with arbitrary admissible h (i.e., such that $0 \leqslant \psi^*(y) + h(y) \leqslant 1$ for all y) in the resident population of strategy ψ^*. From condition (S.4.3), the mutant can invade if

$$\int_{Y_0} h(y)n(y)(\rho(y,0) - 1)\,dy + \int_{Y_f} h(y)n(y)(\rho(y, f(y)D) - 1)\,dy + \int_{Y_1} h(y)n(y)(\rho(y, D) - 1)\,dy > 0$$

In the first term on the left hand side, $h(y)$ must be positive or zero to be admissible; and since $\rho(y,0) < 1$ for $y \in Y_0$, the first term is negative unless $h(y)$ is zero for almost every y in Y_0, in which case the first term is zero. In the last term, $h(y)$ must be negative or zero to be admissible; and since $\rho(y,D) > 1$ for $y \in Y_1$, the last term is negative unless $h(y)$ is zero for almost every y in Y_1. The middle term is, however, zero, because $\rho(y, f(y)D) = 1$ for $y \in Y_f$. Hence a mutant that deviates from ψ^* in any positive fraction of the patches in Y_0 or in Y_1 will go extinct, but a mutant that differs from ψ^* only in Y_f is neutral. The strategy ψ^* in equation (S.4.1) cannot be invaded, but neutral mutants do exist, so that ψ^* is only a weak ESS. □

The sets Y_f and Y_1 depend on the immigration rate and therefore on D, so that they can be determined only numerically. In the remainder of this section, we derive the condition for a patch type y to be in Y_0, i.e., to be outside the set of worthwhile patches.

Proposition 3. *For θ given, a patch type y belongs to $Y_0 = \{y : \rho(y,0) < 1\}$ if*

$$\frac{\gamma}{a(\theta,y)}\left(1 - \frac{\mu}{\mu - a(\theta,y)} \cdot \frac{e^{-a(\theta,y)\tau_{max}} - e^{-\mu\tau_{max}}}{1 - e^{-\mu\tau_{max}}}\right) < 1 \tag{S.4.6}$$

holds with $a(\theta,y) = \gamma + \mu - g(0,y,\theta)$.

Proof. If $\psi(y)D = 0$ in $\rho(y,\psi(y)D)$, a patch of type y does not receive resident immigrants, so that $N(\tau,y) = 0$ for all $\tau \in [0,\tau_{max}]$ and an immigrant would establish a population that grows exponentially. Substituting $N(\tau,y) = 0$ into equation (S.3.2), we obtain

$$F(T,y,\theta,\theta,0) = \gamma \int_T^{\tau_{max}} e^{-a(\theta,y)(\tau-T)} d\tau = \frac{\gamma}{a(\theta,y)}\left(1 - e^{-a(\theta,y)(\tau_{max}-T)}\right)$$

so that

$$\begin{aligned}\rho(y,0) &= \frac{\gamma}{a(\theta,y)} \int_0^{\tau_{max}} \frac{\mu e^{-\mu T}}{1 - e^{-\mu\tau_{max}}}\left(1 - e^{-a(\theta,y)(\tau_{max}-T)}\right) dT = \\ &= \frac{\gamma}{a(\theta,y)}\left(1 - \frac{\mu}{\mu - a(\theta,y)} \cdot \frac{e^{-a(\theta,y)\tau_{max}} - e^{-\mu\tau_{max}}}{1 - e^{-\mu\tau_{max}}}\right).\end{aligned}$$

□

For finite maximum patch age τ_{max}, it can be determined only numerically for which values of $a(\theta, y)$, and therefore (with θ fixed) for which values of y the above expression yields $\rho(y, 0) < 1$. With $\tau_{max} \to \infty$, however, the above expression simplifies to

$$\rho(y,0) = \begin{cases} \infty & \text{if } a(\theta,y) \leqslant 0 \\ \gamma/a(\theta,y) & \text{if } a(\theta,y) > 0. \end{cases} \tag{S.4.7}$$

This result we can interpret heuristically. With constant catastrophe rate and no maximum patch age, the age of an empty patch is irrelevant for an immigrant. Consider a large number of patches of type y where solitary immigrants establish exponentially growing populations. The collective number of individuals grows exponentially at the rate $g(0, y, \theta) - \gamma - \mu$; for the collection of many patches, the catastrophes act simply as an extra mortality rate. If this growth rate is positive (i.e., $a(\theta, y)$ defined in the proposition is negative), then the collective number of individuals "outgrows" the catastrophes, resulting in unbounded growth and infinitely many emigrants. If $a(\theta, y)$ is positive, then the collective number of individuals changes in the same way as in a population that decays exponentially at the rate $a(\theta, y)$, with no births. There are $\gamma/a(\theta, y)$ entries in the dispersal pool for every removed individual, so that when the exponentially decaying population has vanished, each initial individual has contributed $\gamma/a(\theta, y)$ individuals to the dispersal pool.

Equation (S.4.7) yields $\rho(y, 0) < 1$ whenever $a(\theta, y) > \gamma$, i.e., $g(0, y, \theta) - \mu < 0$. With our choice of g in equation (2) of the main text, this is equivalent with

$$|y - \theta| > \sqrt{\frac{r - \mu}{c}}, \tag{S.4.8}$$

so that a patch type is not worthwhile ($y \in Y_0$) if y differs from θ more than by the threshold given on the right hand side of inequality (S.4.8). This simple result holds only for $\tau_{max} \to \infty$.

However, in our numerical examples τ_{max} is so high that the difference from condition (S.4.8) is negligible.

S.5 Convergence stability

In this section, we discuss the convergence stability of a monomorphic singularity. If the immigration strategy ψ is fixed, then the singular local adaptation trait θ (which satisfies equation (7) of the main text) is locally convergence stable (attracts) if

$$\left.\frac{\partial^2 R_m(\theta_m, \theta)}{\partial \theta_m^2} + \frac{\partial^2 R_m(\theta_m, \theta)}{\partial \theta_m \partial \theta}\right|_{\theta_m=\theta} < 0. \tag{S.5.1}$$

A convergence stable ESS is an evolutionary endpoint. If the singularity of a scalar trait is convergence stable but not an ESS, then it is an evolutionary branching point, where the population becomes dimorphic and the two coexisting strategies evolve away from each other (Geritz et al. 1998). Note, however, that this is true only for scalar traits (Geritz et al. 2016). Our current mathematical tools are insufficient to characterize the dynamics of evolution (in particular, evolutionary branching) when the local adaptation trait evolves jointly with the function-valued immigration strategy at arbitrary speeds.

We therefore assume that the immigration strategy evolves fast to its ESS, so that for any current value of the local adaptation trait θ, ψ is as given in equation (6) of the main text. Note that the ESS immigration strategy being an evolutionary attractor is an assumption, but in the present model it seems reasonable. Let ψ^θ denote the ESS immigration strategy for fixed θ. The selection gradient

$$G(\theta) = \lim_{\theta_m \to \theta} \frac{R_m(\psi^\theta, \theta_m, \psi^\theta, \theta) - 1}{\theta_m - \theta} \tag{S.5.2}$$

shows whether a mutant with θ_m slightly higher or lower than the resident θ can invade. For small mutations and away from singularities, invasion implies fixation (Geritz 2005). Once the mutant has spread so that the resident value of θ has changed, the fast-evolving immigration strategy changes accordingly. $G(\theta)$ shows in which direction θ evolves, and the singularity $(\psi^*, \theta^*) = (\psi^{\theta^*}, \theta^*)$ is convergence stable if $G'(\theta^*) < 0$. We evaluate this condition numerically by computing $G(\theta)$ for θ near θ^*. Although $G(\theta^*) = 0$ is equivalent to the singularity condition in equation (7) of the main text, the convergence stability condition $G'(\theta^*) < 0$ is different from condition (S.5.1) since here ψ^θ changes with θ. Unsurprisingly, all monomorphic singularities shown in Figure 1 satisfy $G'(\theta^*) < 0$.

S.6 How to find monomorphic evolutionary singularities

To find the joint monomorphic evolutionary singularity (ψ^*, θ^*), first we characterize the dispersal pool size D of an evolutionarily stable immigration strategy with an equation that is numerically more efficient to use.The size of the dispersal pool D can be determined from the equation $R_m(\psi, \theta, \psi, \theta) = 1$, which, by equation (5) of the main text, is equivalent to

$$\alpha \int_Y \psi(y) n(y) \rho(y, \theta, \theta, \psi(y) D)\, dy = \alpha \int_Y \psi(y) n(y)\, dy + \nu.$$

At the ESS immigration strategy in equation (6) of the main text, $\psi(y) = 0$ when $y \in Y_0$, $\rho(y, \theta, \theta, \psi(y) D) = 1$ when $y \in Y_f$, and $\psi(y) = 1$ when $y \in Y_1$, so that the above equation simplifies to

$$\Phi(D) = \alpha \int_{Y_1} n(y)(\rho(y, \theta, \theta, D) - 1)\, dy - \nu = 0. \tag{S.6.1}$$

To interpret this result heuristically, recall that only those individuals who immigrate into a patch with y in the set Y_1 produce more than one descendants who return to the dispersal pool; and this surplus must balance the mortality in the dispersal pool. On the other hand, individuals following the ESS immigration strategy never settle in patches where they would produce less than one disperser, i.e., the net balance of immigration does not decrease the size of the dispersal pool. The advantage of using equation (S.6.1) instead of $R_m(\psi, \theta, \psi, \theta) = 1$ is that we can determine the steady-state value of D without calculating $\psi(y)$ for $y \in Y_f$ (i.e., where it is nontrivial).

Since $\rho(y, \theta, \theta, D)$ decreases in D and therefore also the set Y_1 becomes a narrower interval as D increases, $\Phi(D)$ is a decreasing function of D, so that the solution of equation (S.6.1) is unique.

We obtain the joint evolutionary singularity (ψ^*, θ^*) by the following procedure:

1. Pick a local adaptation trait value θ and determine the corresponding set Y_0 of non-worthwhile patches using condition (S.4.6) in section S.4 above (for large τ_{max}, the simpler formula in condition (S.4.8) is a good approximation).

2. Choose an arbitrary value for D. For the given values θ and D, evaluate $\rho(y, \theta, \theta, D)$ to find the set of patch types Y_1, where $\rho(y, \theta, \theta, D) > 1$. Compute $\Phi(D)$ in equation (S.6.1).

3. Solve $\Phi(D) = 0$ numerically by repeating step 2 for several values of D and interpolating the function Φ.

4. Using the value of D found in step 3 and the matching set Y_1, solve $\rho(y, \theta, \theta, f(y)D) = 1$ for $f(y)$ at points y not in Y_0 or Y_1 (in practice, obtain $f(y)$ on a mesh and interpolate for all points in Y_f). With this, we have obtained the ESS immigration strategy ψ in equation (6) of the main text, with the corresponding size of the dispersal pool D, but for the arbitrarily chosen value of the local adaptation trait θ. Following the notation of section S.4, we denote this immigration strategy with ψ^θ.

5. With ψ^θ known for the chosen θ, compute the selection gradient

$$\left.\frac{\partial R_m(\psi^\theta, \theta_m, \psi^\theta, \theta)}{\partial \theta_m}\right|_{\theta_m=\theta} = \frac{\alpha}{\alpha\bar{\psi}^\theta + \nu} \int_Y \psi^\theta(y) n(y) \left.\frac{\partial \rho(y, \theta_m, \theta, \psi^\theta(y) D)}{\partial \theta_m}\right|_{\theta_m=\theta} dy.$$

   Vary the value of θ and repeat steps 1-5 to find θ^* such that the selection gradient vanishes, i.e., equation (7) of the main text holds. The joint singularity is then $(\psi^*, \theta^*) = (\psi^{\theta^*}, \theta^*)$.

Note that in the examples of Figure 1 in the main text, symmetry implies that $\theta^* = 0$ is singular.

S.7 Asymmetric patch type distribution

In the examples of Figure 1 in the main text, we assumed a symmetric distribution of patch types. The evolutionarily stable immigration strategies are similar also if the patch types follow an asymmetric distribution. As an example, we take a modified lognormal distribution truncated to the interval $Y = [-3, 3]$,

$$n(y) = \begin{cases} Ce^{-(\ln(y+3)-1)^2/0.245}/(y+3) & \text{if } -3 < y < 3 \\ 0 \quad \text{otherwise} \end{cases} \tag{S.7.1}$$

with the normalization factor C chosen such that the distribution integrates to 1. The local adaptation trait θ^* of the monomorphic singularity is shifted towards the more common patch types, but the evolutionarily stable immigration strategy remains symmetric about θ^* (Figure S1). This is because the decision whether or not to immigrate into a given patch does not depend on whether the patch is of a common or rare type. The only way the patch type distribution affects this decision is via the size of the dispersal pool, which influences the sizes of the local populations. Since the local population dynamics depend on the squared difference $(\theta - y)^2$ (cf. equation (2) of the main text), patches with y deviating from the resident θ by the same amount in either direction are equally good, and therefore are accepted with the same probability by the evolutionarily stable immigration strategy. In other words, the symmetry of the evolutionarily stable immigration strategy is due to the symmetry of the mortality rate $c(\theta - y)^2$, and is independent of the symmetry of the patch type distribution.

The strength of selection towards the ESS immigration strategy does depend on the patch size distribution. In the example of Figure S1, patches with y less than ca -2 are very rare, and therefore the immigration strategy $\psi(y)$ towards these patches is rarely expressed and subject to selection. One can thus expect that even if the immigration strategy evolves fast, $\psi(y)$ can deviate from the ESS for these rare patch types. This however has little effect on diversification;

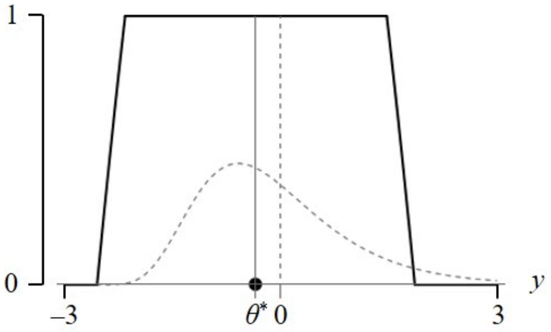

Figure S1: Evolutionarily stable immigration strategy for an asymmetric patch type distribution. The local adaptation trait $\theta^* = -0.35$ (black dot) is convergence stable in the sense of section S.5, but not evolutionarily stable. Black line: $\psi(y)$; dashed grey line: the distribution of patch types, $n(y)$, as given in equation (S.7.1). Parameter values: $r = 5$, $k = 80$, $c = 1$, $\gamma = 2$, $\alpha = 1$, $\nu = 1$; $\mu = 0.1$, $\tau_{max} = 200$ (as in Figure 1 in the main text); $D = 49.48$.

what happens in rare patches is unimportant compared to what happens in more common ones, so that the partitioning of the more common patch types of the peripheral niche (in Figure S1, the peripheral niche with $y > 0$) drives diversification as explained in the main text.

S.8 Evolution in dimorphic populations

If the metapopulation harbours two resident strategies (ψ_1, θ_1) and (ψ_2, θ_2) at steady-state, then the basic reproduction number of a mutant strategy (ψ_m, θ_m) is given by

$$R_m(\psi_m, \theta_m, \psi_1, \theta_1, \psi_2, \theta_2) = \frac{\alpha \int_Y \psi_m(y) n(y) \rho(y, \theta_m, \psi_1(y) D_1, \psi_2(y) D_2) dy}{\alpha \bar{\psi}_m + \nu} \tag{S.8.1}$$

where ρ is calculated analogously to the monomorphic case (equations (S.3.2) and (S.3.3)) but with $N = N_1 + N_2$ and the within-patch dynamics given in equation (4) of the main text. For readability, in this section we suppress the resident traits θ_1, θ_2 in the arguments of ρ, and list only the local adaptation trait of the focal immigrant.

The first resident produces more dispersers than the second resident in a patch of type y if it is better adapted to this patch type, i.e.,

$$\rho(y,\theta_1,\psi_1(y)D_1,\psi_2(y)D_2) > \rho(y,\theta_2,\psi_1(y)D_1,\psi_2(y)D_2) \iff |y-\theta_1| < |y-\theta_2|. \tag{S.8.2}$$

Since equality almost surely does not hold (it holds only at the single point $y = (\theta_1+\theta_2)/2$, which has measure zero for any patch type distribution with a probability density function), it follows that for no values of $\psi_1(y)D_1$ and $\psi_2(y)D_2$ are both ρ values in condition (S.8.2) equal to 1. This implies that at the ESS, $\psi_1(y)$ and $\psi_2(y)$ cannot simultaneously be within the open interval $(0,1)$ (see Proposition 1 in section S.4 above). Further, if θ_1 is better adapted to the patch and $\psi_1(y) < 1$, then $\psi_2(y)$ must be zero, whereas if $\psi_2(y) > 0$, then $\psi_1(y)$ must be 1. For any given D_1, D_2, the set Y of all patch types is hence partitioned in the following subsets (we abbreviate the condition $|y-\theta_1| < |y-\theta_2|$ with $\theta_1 \succ \theta_2$, "θ_1 is better adapted"):

$$\begin{aligned}
Y_{00} &= \{y : \rho(y,\theta_1,0,0) < 1,\ \rho(y,\theta_2,0,0) < 1\} \\
Y_{f0} &= \{y : \theta_1 \succ \theta_2,\ \rho(y,\theta_1,0,0) \geqslant 1,\ \rho(y,\theta_1,D_1,0) \leqslant 1\} \\
Y_{10} &= \{y : \theta_1 \succ \theta_2,\ \rho(y,\theta_1,D_1,0) > 1,\ \rho(y,\theta_2,D_1,0) < 1\} \\
Y_{1f} &= \{y : \theta_1 \succ \theta_2,\ \rho(y,\theta_2,D_1,0) \geqslant 1,\ \rho(y,\theta_2,D_1,D_2) \leqslant 1\} \\
Y_{11} &= \{y : \rho(y,\theta_1,D_1,D_2) > 1,\ \rho(y,\theta_2,D_1,D_2) > 1\} \\
Y_{f1} &= \{y : \theta_1 \prec \theta_2,\ \rho(y,\theta_1,0,D_2) \geqslant 1,\ \rho(y,\theta_1,D_1,D_2) \leqslant 1\} \\
Y_{01} &= \{y : \theta_1 \prec \theta_2,\ \rho(y,\theta_1,0,D_2) < 1,\ \rho(y,\theta_2,0,D_2) > 1\} \\
Y_{0f} &= \{y : \theta_1 \prec \theta_2,\ \rho(y,\theta_2,0,0) \geqslant 1,\ \rho(y,\theta_2,0,D_2) \leqslant 1\}.
\end{aligned} \tag{S.8.3}$$

These subsets are verbally described in the main text. By a straightforward generalization of the monomorphic case, the dimorphic ESS immigration strategies for fixed θ_1, θ_2 are given by

$$\psi_1^*(y) = \begin{cases} 0 & \text{if } y \in Y_{00} \cup Y_{0f} \cup Y_{01} \\ f_{m,1}(y) & \text{if } y \in Y_{f0} \\ f_{d,1}(y) & \text{if } y \in Y_{f1} \\ 1 & \text{if } y \in Y_{10} \cup Y_{1f} \cup Y_{11} \end{cases} \tag{S.8.4a}$$

and

$$\psi_2^*(y) = \begin{cases} 0 & \text{if } y \in Y_{00} \cup Y_{f0} \cup Y_{10} \\ f_{m,2}(y) & \text{if } y \in Y_{0f} \\ f_{d,2}(y) & \text{if } y \in Y_{1f} \\ 1 & \text{if } y \in Y_{01} \cup Y_{f1} \cup Y_{11} \end{cases} \tag{S.8.4b}$$

where the functions $f_{m,1}, f_{d,1}, f_{m,2}, f_{d,2}$ ('m' for monomorphic local populations in patches where the other resident does not immigrate into and 'd' for dimorphic) satisfy

$$\begin{aligned} &\rho(y, \theta_1, f_{m,1}(y)D_1, 0) = 1, \quad \rho(y, \theta_1, f_{d,1}(y)D_1, D_2) = 1 \\ &\rho(y, \theta_2, 0, f_{m,2}(y)D_2) = 1, \quad \rho(y, \theta_2, D_1, f_{d,2}(y)D_2) = 1 \end{aligned} \tag{S.8.4c}$$

and D_1, D_2 are such that

$$R_m(\psi_1^*, \theta_1, \psi_1^*, \theta_1, \psi_2^*, \theta_2) = 1, \quad R_m(\psi_2^*, \theta_2, \psi_1^*, \theta_1, \psi_2^*, \theta_2) = 1. \tag{S.8.4d}$$

To find the ESS immigration strategies in practice, the simplification that led to equation (S.6.1) in the monomorphic case can unfortunately not be used for dimorphic populations. The analogous equation for the first resident would be $\alpha \int_{Y_{10} \cup Y_{1f} \cup Y_{11}} n(y)(\rho(y, \theta_1, D_1, \psi_2(y)D_2) - 1)dy - \nu = 0$, but this depends on ψ_2, which has nontrivial values in Y_{1f}; and *vice versa* for the second resident. This means that one has to resort to determining D_1, D_2 in the hard way. The numerical procedure of obtaining the ESS immigration strategies for fixed values of the local adaptation traits is as follows:

1. Given the values θ_1 and θ_2, the set Y_{00} can be determined directly.
2. Choose arbitrary values for D_1 and D_2. Determine all sets in (S.8.3) and solve equations (S.8.4c). Construct the immigration strategies $\psi_1^{(D_1,D_2)}$ and $\psi_2^{(D_1,D_2)}$ as in equations (S.8.4a,b); note that these are not the ESS because D_1 and D_2 are arbitrary.

3. Calculate, for $i = 1, 2$, the number of emigrants produced,

$$\mathcal{E}_i(D_1, D_2) = \int_Y \int_0^{\tau_{max}} \gamma N_i(\tau, y) q(\tau) n(y) \, d\tau \, dy$$

   where $N_i(\tau, y)$ is from equation (4) of the main text with $I_i(y) = \alpha \psi_i^{(D_1, D_2)} D_i$.

4. Solve the two nonlinear equations

$$\frac{\mathcal{E}_1(D_1, D_2)}{\alpha \int_Y n(y) \psi_1^{(D_1, D_2)} dy + \nu} = D_1, \quad \frac{\mathcal{E}_2(D_1, D_2)}{\alpha \int_Y n(y) \psi_2^{(D_1, D_2)} dy + \nu} = D_2 \tag{S.8.5}$$

   for D_1 and D_2 numerically, by repeating steps 2-3 for various values of D_1, D_2. The immigration strategies $\psi_1^{(D_1, D_2)}$ and $\psi_2^{(D_1, D_2)}$ obtained with the solution of equation (S.8.5) are the ESS strategies ψ_1 and ψ_2.

Once the ESS immigration strategies ψ_1 and ψ_2 are found for given θ_1 and θ_2, we can evaluate the selection gradients on the local adaptation traits, and find the singular dimorphism that satisfies

$$\left. \frac{\partial R_m(\psi_i^*, \theta_m, \psi_1^*, \theta_1^*, \psi_2^*, \theta_2^*)}{\partial \theta_m} \right|_{\theta_m = \theta_i^*} = 0 \quad \text{for } i = 1, 2 \tag{S.8.6}$$

when ψ_1^* and ψ_2^* are the ESS immigration strategies for θ_1^* and θ_2^*. The singular dimorphism is evolutionarily stable if

$$\left. \frac{\partial^2 R_m(\psi_i^*, \theta_m, \psi_1^*, \theta_1^*, \psi_2^*, \theta_2^*)}{\partial \theta_m^2} \right|_{\theta_m = \theta_i^*} < 0 \quad \text{for } i = 1, 2. \tag{S.8.7}$$

In our examples, we build on symmetry around $y = 0$ and restrict the analysis to $\theta_2 = -\theta_1$. In principle, even a symmetric model can have (symmetric pairs of) asymmetric singularities (see Kisdi and Geritz 1999 for an example), but in the present model we do not expect that asymmetric singularities would exist. The restriction implies $D_1 = D_2$ such that equations (S.8.5) reduce to a single equation, which greatly facilitates the numerical analysis.

S.9 Mutual invasibility

Suppose that the patch type distribution $n(y)$ is symmetric about $y = 0$. In this section, we show that near the monomorphic singularity (ψ^*, θ^*) with $\theta^* = 0$ and ψ^* symmetric about $y = 0$, there always exist pairs of strategies that differ only in their local adaptation trait and can invade each other's resident population, i.e., they coexist in a protected dimorphism.

Let ψ^* be fixed. Coexistence by mutual invasibility is possible near a scalar singularity if

$$\left.\frac{\partial^2 R_m(\theta_m, \theta)}{\partial\theta_m \partial\theta}\right|_{\theta_m=\theta=\theta^*} < 0$$

(Geritz et al. 1998), where we suppressed the dependence of R_m on the fixed immigration strategy. From the definition of R_m in equation (5) of the main text and equation (S.3.3), it is clear that the above inequality is satisfied if

$$\left.\frac{\partial^2 F(T, y, \theta_m, \theta, \psi^*(y)D)}{\partial\theta_m \partial\theta}\right|_{\theta_m=\theta=\theta^*} < 0$$

holds for every y and every T. Using equation (2) of the main text and (S.3.2), this second derivative evaluates to

$$\left.\frac{\partial^2 F}{\partial\theta_m \partial\theta}\right|_{\theta_m=\theta=0} = -\frac{2c\gamma r}{k}\, y \int_T^{\tau_{max}} \chi(\tau) \int_T^{\tau} \left.\frac{dN(\zeta, y)}{d\theta}\right|_{\theta=0} d\zeta\, d\tau \tag{S.9.1}$$

where $\chi(\tau) = (\tau - T)e^{-\mu(\tau-T)+\int_T^\tau r(1-N(\zeta,y)/k)-cy^2-\gamma d\zeta}$ is a positive factor.

$N(\zeta, y)$, the local population size at patch age ζ in a patch of type y depends on the resident strategy θ via two effects (cf. equations (1) and (2) of the main text). First, by changing the size of the dispersal pool D, a change of the resident θ changes the immigration rate $I(y) = \alpha\psi(y)D$. At the symmetric singularity at $\theta^* = 0$, however, the first order effect $dD/d\theta$ vanishes. This is because, by symmetry, the dispersal pool size of the resident θ is the same as that of the resident

$-\theta$; and therefore D must attain an extremum at $\theta = 0$. Second, a change in θ changes the term $-c(\theta - y)^2$ in the *per capita* growth rate in equation (2) of the main text. Heuristically, it is clear that at any fixed ζ, $N(\zeta, y)$ must increase with decreasing death rate, i.e., with decreasing $(\theta - y)^2$ (a formal proof is given at the end of this section). Therefore, for positive y, $\frac{dN(\zeta,y)}{d\theta}$ at $\theta = 0$ is positive for all ζ, so that the mixed derivative in equation (S.9.1) is indeed negative. For negative y, $\frac{dN(\zeta,y)}{d\theta}$ at $\theta = 0$ is negative for all ζ; this combines with the negative factor y before the integral so that the mixed derivative is again negative. Hence we conclude that with the immigration strategy fixed at the singular, symmetric strategy ψ^*, two strategies with different local adaptation traits near the singular value $\theta^* = 0$ can coexist.

As the metapopulation evolves towards (ψ^*, θ^*), however, the immigration strategy will have not reached ψ^* yet when coexistence becomes an issue. Suppose that the current resident local adaptation trait θ is close to, but not equal to, $\theta^* = 0$; with the immigration strategy evolving fast, ψ will be close to but not identical to ψ^*. Since ψ is not exactly symmetric, D does not attain an extremum at $\theta^* = 0$; but, by continuity, the nonzero slope of D as the function of θ is of the same order as $(\theta - \theta^*)$. The quantity on the right hand side of equation (S.9.1) is nonvanishing, so that the contribution from $dD/d\theta$ will not change its sign when the resident θ is sufficiently close to the singularity.

A similar argument shows that the above result extends also to asymmetric patch size distributions, provided that the deviation from a symmetric distribution is sufficiently small.

In the remainder of this section, we prove that $N(\tau, y)$ indeed increases with decreasing $(\theta - y)^2$ at any patch age τ. Using the abbreviation $z = (\theta - y)^2$ and considering $I = I(y)$ to be fixed, we can rewrite equation (S.2.9) in the form

$$\frac{2I}{N(\tau, y)} = \frac{2}{\tau}\,\phi(\tfrac{\tau}{2}u) - (r - \gamma - cz) \tag{S.9.2}$$

with $u = \sqrt{4Ir/k + (r - \gamma - cz)^2}$ and $\phi(u) = u \coth(u)$, where coth is the hyperbolic cotangent. Differentiating equation (S.9.2), we obtain

$$\frac{d}{dz}\frac{2I}{N(\tau, y)} = \phi'(\tfrac{\tau}{2}u)\frac{du}{dz} + c.$$

It is straightforward to show that $0 < \phi'(u) < 1$ for every $u > 0$. Since

$$\frac{du}{dz} = -c\frac{r - \gamma - cz}{\sqrt{4Ir/k + (r - \gamma - cz)^2}}$$

i.e., $-c$ times a factor less than 1 in absolute value, we conclude that $\frac{2I}{N(\tau,y)}$ increases with z and therefore $N(\tau, y)$ increases with decreasing $z = (\theta - y)^2$.

S.10 Adaptive dynamics in a minimal model

In this section, we consider a minimal model to investigate the dynamics of diversification via the evolution of contrasting immigration strategies in the peripheral niche, using the canonical equation of adaptive dynamics (Dieckmann and Law 1996; Durinx et al. 2008; Metz et al. 2016). For reasons described at the end of this section (see "Open problems"), here we assume a finite number of patch types (only three in our numerical examples) and we assume that there are no catastrophes (i.e., the catastrophe rate is $\mu = 0$ and the maximum age of a patch is $\tau_{max} \to \infty$). This is therefore a rather special case of our main model, which we analyse here independently of the main article. An independent analysis is necessary because the methods used in the main model do not apply; here the basic reproduction number R_m is undefined because the dispersal generations are infinitely long, and mutants immigrating into patches where they are locally favoured produce infinitely many emigrants who return to the dispersal pool.

As we point out in the Discussion of the main article, when there are only a finite number of patch types, it can happen that none of them is in the peripheral niche. Here we choose the parameters of the model such that we have patches in the peripheral niche. This choice is generic in the sense that small changes in the parameters would not change the behaviour of the model qualitatively, but with sufficiently different values of the parameters, which imply no patches in the peripheral niche, the results would differ.

Further, our main model assumes infinitely many patches with a continuous distribution of patch type as well as of patch age. With no catastrophes (hence no age distribution) and only finitely many patch types, we can accommodate also metapopulations with only a limited number of patches. This is because without catastrophes, patches of the same type are indistiguishable; hence it does not matter whether we have a large number of separate patches or whether these patches are united into one (or a few) large patch(es). For simplicity, we continue using the terminology of classical metapopulations with infinitely many patches and the frequencies of patch types; these could however be replaced with one (or a few) large patch(es) for each type, the frequency of the type replaced with the area of the the large patch(es).

Assume that the patches are of ω different types, with y_j ($j = 1, ..., \omega$) denoting the within-patch optimal phenotypes and n_j the frequencies of patch types ($\sum_{j=1}^{\omega} n_j = 1$). As in our main model, let θ_i denote the local adaptation trait of resident i ($i = 1, ..., S$, where S is the number of resident strategies present; in our examples, $S = 1$ or 2) and let $\psi_{i,j}$ be the probability that resident i immigrates into a patch of type j upon encounter. The ecological dynamics of a mutant with local adaptation trait θ_m and immigration strategy $\psi_{m,j}$ is given by

$$\begin{aligned} \frac{dN_{m,j}}{dt} &= g(\hat{N}_j, y_j, \theta_m)N_{m,j} - \gamma N_{m,j} + \alpha\psi_{m,j}D_m \quad \text{for } j = 1, ..., \omega \\ \frac{dD_m}{dt} &= \sum_{j=1}^{\omega} \gamma n_j N_{m,j} - (\alpha\bar{\psi}_m + \nu)D_m \end{aligned} \tag{S.10.1}$$

where $\bar{\psi}_m = \sum_{j=1}^{\omega} n_j \psi_{m,j}$ and

$$g(\hat{N}_j, y_j, \theta_m) = b - (d_0 + d_1 \hat{N}_j) - c(\theta_m - y_j)^2$$

is the *per capita* growth rate of the mutant in a patch of type j, with $\hat{N}_j = \sum_{i=1}^{S} \hat{N}_{i,j}$ being the total resident density in such a patch at equilibrium. Because there are no catastrophes, every patch is at an equilibrium with resident densities $\hat{N}_{i,j}$, which can be found from the resident ecological dynamics analogous to equations (S.10.1) with $\theta_m = \theta_i$, $\psi_{m,j} = \psi_{i,j}$, $i = 1, ..., S$. Notice that here we replace the logistic formula $r\left(1 - \frac{N}{k}\right)$, used in the main model, with the equivalent formula $b - (d_0 + d_1 N)$, which contains the birth rate b and the density-dependent death rate $d_0 + d_1 N$ explicitly; as detailed below, this is necessary to derive the canonical equation.

For structured populations, the canonical equation for the ith resident ($i = 1, ..., S$) can be written as

$$\frac{d\mathbf{X}_i}{dt} = \kappa_i \, \mathbf{C} \, \frac{\partial r}{\partial \mathbf{X}_m}\bigg|_{\mathbf{X}_m = \mathbf{X}_i} \tag{S.10.2}$$

(Durinx et al. 2008). Here $\mathbf{X}_i = [\theta_i, \psi_{i,1}, ..., \psi_{i,\omega}]^T$ is the trait vector of resident i, κ_i is a scalar speed factor, $\mathbf{C}$ is the mutational variance-covariance matrix, and $\frac{\partial r}{\partial \mathbf{X}_m}\Big|_{\mathbf{X}_m = \mathbf{X}_i}$ is the selection gradient vector, i.e., the derivative of the mutant invasion fitness r with respect to the mutant trait values. For the covariance matrix, we assume that the traits are uncorrelated and the mutational variance is w_θ for the local adaptation trait and w_ψ for the components of the immigration strategy, i.e., $\mathbf{C}$ is diagonal with elements $C_{11} = w_\theta$ and $C_{jj} = w_\psi$ for $2 \leq j \leq \omega + 1$. In the next two subsections, we detail the calculation of the selection gradient and of the speed factor; for the results, readers may want to skip to the subsection "Numerical experiments".

The selection gradient

The mutant metapopulation is a structured population with D_m individuals in the dispersal pool and $n_j N_{m,j}$ individuals in patches of type j for each patch of the metapopulation. The population vector of the mutant is therefore

$$\begin{bmatrix} D_m \\ n_1 N_{m,1} \\ \vdots \\ n_\omega N_{m,\omega} \end{bmatrix}$$

and the Jacobian matrix of its linearized dynamics is

$$\mathbf{J}_m = \begin{bmatrix} -(\alpha\bar{\psi}_m + \nu) & \gamma & \gamma & \dots & \gamma \\ \alpha n_1 \psi_{m,1} & g(\hat{N}_1, y_1, \theta_m) - \gamma & 0 & \dots & 0 \\ \alpha n_2 \psi_{m,2} & 0 & g(\hat{N}_2, y_2, \theta_m) - \gamma & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha n_\omega \psi_{m,\omega} & 0 & 0 & \dots & g(\hat{N}_\omega, y_\omega, \theta_m) - \gamma \end{bmatrix}$$

The invasion fitness r of the mutant is the leading eigenvalue of its Jacobian. Since each resident has zero invasion fitness and the mutant strategy is close to one of the residents, we can approximate r to first order using the eigenvalue sensitivity formula $\Delta r = \mathbf{v}^T \Delta \mathbf{J} \mathbf{u}$ (Caswell 1989; for use in evolution, see Wickman et al. 2017 and references therein), where

$$\Delta\mathbf{J} = \begin{bmatrix} -\alpha\Delta\bar{\psi} & \gamma & \gamma & \dots & \gamma \\ \alpha n_1 \Delta\psi_1 & \frac{\partial g}{\partial \theta_m}|_1 \,\Delta\theta & 0 & \dots & 0 \\ \alpha n_2 \Delta\psi_2 & 0 & \frac{\partial g}{\partial \theta_m}|_2 \,\Delta\theta & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha n_\omega \Delta\psi_\omega & 0 & 0 & \dots & \frac{\partial g}{\partial \theta_m}|_\omega \,\Delta\theta \end{bmatrix}$$

Assuming the mutant is close to the ith resident, $\Delta\psi_j = \psi_{m,j} - \psi_{i,j}$, $\Delta\theta = \theta_m - \theta_i$, and $\frac{\partial g}{\partial \theta_m}|_j$ is the derivative of $g(\hat{N}_j, y_j, \theta_m)$ evaluated at $\theta_m = \theta_i$. The eigenvectors $\mathbf{v}$ and $\mathbf{u}$ are the left and right eigenvectors of the resident Jacobian scaled such that $\mathbf{v}^T\mathbf{u} = 1$, i.e.,

$$\mathbf{u}^{(i)} = \begin{bmatrix} \hat{D}_i \\ n_1\hat{N}_{i,1} \\ \vdots \\ n_\omega \hat{N}_{i,\omega} \end{bmatrix}, \quad \mathbf{v}^{(i)} = v_0^{(i)} \begin{bmatrix} 1 \\ \gamma/(\gamma - g(\hat{N}_1, y_1, \theta_i)) \\ \vdots \\ \gamma/(\gamma - g(\hat{N}_\omega, y_\omega, \theta_i)) \end{bmatrix} \text{ with } v_0^{(i)} = \frac{1}{\hat{D}_i + \sum_{j=1}^{\omega} \frac{\gamma n_j \hat{N}_{i,j}}{\gamma - g(\hat{N}_j, y_j, \theta_i)}}.$$

Notice that since the local dynamics is at equilibrium so that $[g(\hat{N}_j, y_j, \theta_i) - \gamma]\hat{N}_{i,j} + \alpha\psi_{i,j}\hat{D}_i = 0$, the difference $\gamma - g(\hat{N}_j, y_j, \theta_i)$ is positive for each patch type j where $\psi_{i,j} > 0$. Substituting the eigenvectors into $\Delta r = \mathbf{v}^T \Delta\mathbf{J}\mathbf{u}$, we arrive at

$$\Delta r = v_0^{(i)} \left[\alpha\hat{D}_i \sum_{j=1}^{\omega} n_j \frac{g(\hat{N}_j, y_j, \theta_i)}{\gamma - g(\hat{N}_j, y_j, \theta_i)} \Delta\psi_j + \sum_{j=1}^{\omega} n_j \frac{\gamma \hat{N}_{i,j}}{\gamma - g(\hat{N}_j, y_j, \theta_i)} \left. \frac{\partial g}{\partial \theta_m} \right|_j \Delta\theta \right].$$

The selection gradient vector of the ith resident therefore contains the elements

$$\left. \frac{\partial r}{\partial \theta_m} \right|_{\theta_m = \theta_i} = v_0^{(i)} \sum_{j=1}^{\omega} n_j \frac{\gamma \hat{N}_{i,j}}{\gamma - g(\hat{N}_j, y_j, \theta_i)} \left. \frac{\partial g(\hat{N}_j, y_j, \theta_m)}{\partial \theta_m} \right|_{\theta_m = \theta_i} \tag{S.10.3a}$$

and

$$\left. \frac{\partial r}{\partial \psi_{m,j}} \right|_{\psi_{m,j} = \psi_{i,j}} = v_0^{(i)} \alpha \hat{D}_i n_j \frac{g(\hat{N}_j, y_j, \theta_i)}{\gamma - g(\hat{N}_j, y_j, \theta_i)}, \quad j = 1, ..., \omega. \tag{S.10.3b}$$

Analogously to the main model, in a monomorphic population with fixed local adaptation trait θ, the immigration strategy

$$\psi_j = \begin{cases} 0 & \text{if } g(0, y_j, \theta) < 0 \\ f_j & \text{if } g(\hat{N}_j, y_j, \theta) = 0 \\ 1 & \text{if } g(\hat{N}_j, y_j, \theta) > 0 \end{cases} \tag{S.10.4}$$

with f_j such that $\alpha f_j \hat{D} = \gamma \hat{N}_j$, i.e., with $f_j = \frac{\gamma}{\alpha\hat{D}} \frac{b - d_0}{d_1} \left[1 - \frac{c}{b - d_0} (\theta - y_j)^2 \right]$, is a weak ESS.

The speed factors

To calculate the speed factors κ_i in the canonical equation (S.10.2), we follow Durinx et al. (2008). This calculation is somewhat involved. Since we are interested in the orbit of the canonical equation and in particular its attractor, we can re-scale time in a nonlinear fashion such that only the ratios of the speed factors matter, their absolute values do not. Hence $\kappa = 1$ can be used for monomorphic resident populations and for dimorphic populations when symmetry ensures $\kappa_1 = \kappa_2$.

Let $\xi_{j,l}^{(i)}$ denote the random number of offspring that an individual with the ith resident strategy will produce in patches of type j, provided the individual itself was born in a patch of type l. The expectation of $\xi_{j,l}^{(i)}$, which we denote with $L_{jl}^{(i)}$, are the elements of the $\omega \times \omega$ matrix $\mathbf{L}^{(i)}$, which is the next generation matrix of the ith resident. We shall need the expectations and the covariances of $\xi_{j,l}^{(i)}$; we start with the expectations.

To ease the notation, we introduce the shorthands $\Gamma_j^{(i)} = d_0 + d_1\hat{N}_j + c(\theta_i - y_j)^2 + \gamma$ for the rate at which residents with the $i$the strategy are removed from a patch of type j (either due to death or emigration); $\Pi_j^{(i)}$ for the probability of getting into a patch of type j (either directly or after visiting patches of other types) for an individual currently in the dispersal pool; and $\Lambda_j^{(i)}$ for the expected number of offspring produced in a patch of type j during its remaining lifetime for an individual currently in the dispersal pool. By first step analysis, we have

$$\Pi_j^{(i)} = \frac{\alpha n_j \psi_{i,j}}{\alpha\bar{\psi}_i + \nu} + \sum_{l \neq j} \frac{\alpha n_l \psi_{i,l}}{\alpha\bar{\psi}_i + \nu} \frac{\gamma}{\Gamma_l^{(i)}} \Pi_j^{(i)} \iff \Pi_j^{(i)} = \frac{\alpha n_j \psi_{i,j}}{\alpha\bar{\psi}_i + \nu - \sum_{l \neq j} \alpha n_l \psi_{i,l} \frac{\gamma}{\Gamma_l^{(i)}}}$$

and

$$\Lambda_j^{(i)} = \Pi_j^{(i)} \left(\frac{b}{\Gamma_j^{(i)}} + \frac{\gamma}{\Gamma_j^{(i)}} \Lambda_j^{(i)} \right) \iff \Lambda_j^{(i)} = \frac{\Pi_j^{(i)} b}{\Gamma_j^{(i)} - \gamma \Pi_j^{(i)}}.$$

The elements of the next generation matrix $\mathbf{L}^{(i)}$ are

$$L_{jl}^{(i)} = \begin{cases} \dfrac{\gamma}{\Gamma_l^{(i)}} \Lambda_j^{(i)} & \text{if } l \neq j \\ \\ \dfrac{b}{\Gamma_j^{(i)}} + \dfrac{\gamma}{\Gamma_j^{(i)}} \Lambda_j^{(i)} & \text{if } l = j \end{cases}$$

The expression in the first row is the probability that an individual born in a patch of type l makes it to the dispersal pool, times the number of offspring it can expect to have in patches of type j once it is in the dispersal pool. The expression in the second row adds also the offspring an individual produces before emigrating from the patch where it was born.

Next, we turn to the covariances between the random variables $\xi_{j,l}^{(i)}$. As an example, consider the covariance between $\xi_{1,3}^{(i)}$ and $\xi_{2,3}^{(i)}$, the number of offspring produced in patches of type 1 and in patches of type 2 by an individual born in a patch of type 3. To have any of $\xi_{1,3}^{(i)}$ and $\xi_{2,3}^{(i)}$ nonzero, the focal individual has to leave its birth patch. If it first immigrates into a patch of type 1, produces offspring there, and dies there, then it will have no offspring produced in a patch of type 2; and conversely, if it first immigrates into a patch of type 2, it may die there and have no offspring in a patch of type 1. $\xi_{1,3}^{(i)}$ and $\xi_{2,3}^{(i)}$ are therefore not independent but negatively correlated. However, what happens while an individual stays in a patch, i.e., how many offspring it produces there and with what probability it returns to the dispersal pool, is independent of how many offspring it produced during its earlier visits to other patches. $\xi_{1,3}^{(i)}$ and $\xi_{2,3}^{(i)}$ are therefore *conditionally independent* given the number of visits to patches of type 1 and type 2.

We therefore start with the probability $P^{(i)}(k_1, ..., k_\omega)$ that an individual with the ith resident strategy, who is now in the dispersal pool, visits patches of type $1, ..., \omega$ exactly $k_1, ..., k_\omega$ times during its remaining life. Importantly, we define a "visit" as immigration, stay and emigration (not death); we shall take care of stays ending in deaths later, and such a stay does not count towards k_j. $P^{(i)}(k_1,...,k_\omega)$ is given by a negative multinomial distribution,

$$P^{(i)}(k_1,...,k_\omega) = \frac{(k_1+...+k_\omega)!}{k_1!...k_\omega!} p_0 p_1^{k_1}...p_\omega^{k_\omega}$$

with success probabilities

$$p_j = p_j^{(i)} = \frac{\alpha n_j \psi_{i,j}}{\alpha \bar{\psi}_i + \nu} \frac{\gamma}{\Gamma_j^{(i)}}$$

and failure probability $p_0 = p_0^{(i)} = 1 - \sum_{j=1}^{\omega} p_j$.

To consider a given pair of random variables, say $\xi_{j_1,l}^{(i)}$ and $\xi_{j_2,l}^{(i)}$, we can marginalise $P^{(i)}(k_1,...,k_\omega)$ considering only three types of successes: visit to a patch of type j_1, visit to a patch of type j_2, and visit to a patch of any other type. Let $MP_{j_1,j_2}(k_1,k_2)$ denote the probability of k_1 visits to patches of type j_1 and k_2 visits to patches of type j_2 (note the change in indexing k). These marginalised probabilities are

$$MP_{j_1,j_2}^{(i)}(k_1,k_2) = \sum_{k_3=0}^{\infty} \frac{(k_1+k_2+k_3)!}{k_1!k_2!k_3!} p_0 \pi_1^{k_1} \pi_2^{k_2} \pi_3^{k_3}$$

with success probabilities $\pi_1 = p_{j_1}$, $\pi_2 = p_{j_2}$, and $\pi_3 = \sum_{j \neq j_1,j_2} p_j$.

During each visit to a patch of type j, an individual of strategy i produces a geometrically distributed number of offspring with success probability $b/(b+\Gamma_j^{(i)})$. Recall that the sum of geometrically distributed random variables follows a negative binomial distribution; therefore if this individual has k visits to patches of type j, it produces a total of z offspring during these visits with probability

$$NB_j^{(i)}(z;k) = \frac{(z+k-1)!}{z!(k-1)!} \left(\frac{b}{b+\Gamma_j^{(i)}} \right)^z \left(\frac{\Gamma_j^{(i)}}{b+\Gamma_j^{(i)}} \right)^k$$

Therefore the probability that, starting from the dispersal pool, an individual visits patches of type j_1 and patches of type j_2 respectively k_1 and k_2 times and during these visits produces z_1

offspring in patches of type j_1 plus z_2 offspring in patches of type j_2 is

$$MP^{(i)}_{j_1,j_2}(k_1,k_2) \cdot NB^{(i)}_{j_1}(z_1;k_1) \cdot NB^{(i)}_{j_2}(z_2;k_2).$$

As a last step in probability calculus, we marginalise over k_1 and k_2 and also include a possible last stay in a patch that ends with death rather than emigration. The probability, for an individual of the ith resident strategy now in the dispersal pool, to produce z_1 offspring in patch type j_1 and z_2 offspring in patch type j_2 during its remaining lifetime is

$$\begin{aligned}
\Psi^{(i)}_{j_1,j_2}(z_1,z_2) \quad = \quad & \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\left[\frac{MP^{(i)}_{j_1,j_2}(k_1,k_2)}{p_0^{(i)}}\left(\sum_{j\neq j_1,j_2}\frac{\alpha n_j \psi_{i,j}}{\alpha\bar{\psi}_i+\nu}\left[1-\frac{\gamma}{\Gamma_j^{(i)}}\right]+\frac{\nu}{\alpha\bar{\psi}_i+\nu}\right)+\right.\\
& +\frac{MP^{(i)}_{j_1,j_2}(k_1-1,k_2)}{p_0^{(i)}}\left(\frac{\alpha n_{j_1}\psi_{i,j_1}}{\alpha\bar{\psi}_i+\nu}\left[1-\frac{\gamma}{\Gamma_{j_1}^{(i)}}\right]\right)+\\
& \left.+\frac{MP^{(i)}_{j_1,j_2}(k_1,k_2-1)}{p_0^{(i)}}\left(\frac{\alpha n_{j_2}\psi_{i,j_2}}{\alpha\bar{\psi}_i+\nu}\left[1-\frac{\gamma}{\Gamma_{j_2}^{(i)}}\right]\right)\right] NB^{(i)}_{j_1}(z_1;k_1)\, NB^{(i)}_{j_2}(z_2;k_2)
\end{aligned}$$

for $z_1, z_2 > 0$ (which means $k_1, k_2 > 0$). In the first line of this expression, the factor in the parentheses is the probability that the individual immigrates into and dies in a patch of type other than types j_1 or j_2, or dies in the dispersal pool. In the second line, the parentheses contain the probability that the individual immigrates into and dies in a patch of type j_1; and in the third line, the same for j_2.

Now we can calculate the covariance between $\xi^{(i)}_{j_1,l}$ and $\xi^{(i)}_{j_2,l}$. Consider the case $j_1 \neq j_2$, and suppose first that the birth patch is of a third type (i.e., $l \neq j_1, j_2$). In this case, both $\xi^{(i)}_{j_1,l}$ and $\xi^{(i)}_{j_2,l}$ are zero unless the individual emigrates from the birth patch and enters the dispersal pool. The covariance is therefore

$$COV(\xi^{(i)}_{j_1,l}, \xi^{(i)}_{j_2,l}) = \frac{\gamma}{\Gamma_l^{(i)}}\sum_{z_1=1}^{\infty}\sum_{z_2=1}^{\infty}\Psi^{(i)}_{j_1,j_2}(z_1,z_2)z_1z_2 - L^{(i)}_{j_1l}L^{(i)}_{j_2l} \quad \text{for } j_1 \neq j_2$$

If the individual is born in a patch of type j_1 i.e., $(l = j_1)$, then it produces a random number of offspring in a patch of type j_1 before it first emigrates or dies in its natal patch. This is however independent of $\xi^{(i)}_{j_2,l}$, and therefore the covariance is the same as above. The same holds for $l = j_2$. Finally, if $j_1 = j_2 = j$, then the covariance is the variance of $\xi^{(i)}_{j,l}$. To obtain this variance, marginalise $P^{(i)}(k_1, ..., k_\omega)$ to only two types of successes (visit to a patch of type j and visit to a patch of any other type) and calculate $\tilde{\Psi}^{(i)}_j(z)$, the probability of getting z offspring in patches of type j, analogously to the above. Then

$$COV(\xi^{(i)}_{j,l}, \xi^{(i)}_{j,l}) = \begin{cases} \dfrac{\gamma}{\Gamma^{(i)}_l} \displaystyle\sum_{z=1}^{\infty} \tilde{\Psi}^{(i)}_j(z) z^2 - \left(L^{(i)}_{jl}\right)^2 & \text{for } j \neq l \\ \dfrac{\gamma}{\Gamma^{(i)}_l} \displaystyle\sum_{z=1}^{\infty} \tilde{\Psi}^{(i)}_l(z) z^2 - \left(L^{(i)}_{ll} - \dfrac{b}{\Gamma^{(i)}_l}\right)^2 + \dfrac{b}{\Gamma^{(i)}_l}\left(\dfrac{b}{\Gamma^{(i)}_l} + 1\right) & \text{for } j = l \end{cases}$$

The exception in the second line is made for the case when j is the type of the birth patch. In this case, one has to add the variance of the number of offspring produced in the birth patch before emigration or death.

The covariances between the offspring numbers enter the speed factor of the canonical equation through the quantity

$$\sigma_i^2 = \sum_{l=1}^{\omega} U^{(i)}_l \sum_{j_1=1}^{\omega} \sum_{j_2=1}^{\omega} V^{(i)}_{j_1} V^{(i)}_{j_2} COV(\xi^{(i)}_{j_1,l}, \xi^{(i)}_{j_2,l}) \tag{S.10.5}$$

where $[U^{(i)}_1, ..., U^{(i)}_\omega]^T$ and $[V^{(i)}_1, ..., V^{(i)}_\omega]^T$ are respectively the leading right and left eigenvectors of the next generation matrix of the ith resident, $\mathbf{L}^{(i)}$, normalised according to $\sum_{l=1}^{\omega} U_l = 1$ and $\sum_{l=1}^{\omega} U_l V_l = 1$ (Durinx et al. 2008).

Another quantity we need for the speed factor is the generation time of the ith resident, $T^{(i)}_f$. The generation time "switches" between the selection gradient written in terms of the invasion fitness, as in (S.10.3), and in terms of the leading eigenvalue R_0 of the mutant's next generation matrix (the latter is not to be mixed up with the basic reproduction number of dispersal generations, R_m, used in our main model). More precisely, for a mutant close to the ith resident, $\log R_0 = T_f^{(i)} r$ holds to first order (see equation (19) of Durinx et al. 2008). The mutant next generation matrix is obtained analogously to that of the residents,

$$L_{jl}^{(m)} = \begin{cases} \dfrac{\gamma}{\Gamma_l^{(m)}} \Lambda_j^{(m)} & \text{if } l \neq j \\ \\ \dfrac{b}{\Gamma_j^{(m)}} + \dfrac{\gamma}{\Gamma_j^{(m)}} \Lambda_j^{(m)} & \text{if } l = j \end{cases}$$

with

$$\begin{aligned} \Gamma_j^{(m)} &= d_0 + d_1 \hat{N}_j + c(\theta_m - y_j)^2 + \gamma, \\ \Pi_j^{(m)} &= \frac{\alpha n_j \psi_{m,j}}{\alpha \bar{\psi}_m + \nu - \sum_{l \neq j} \alpha n_l \psi_{m,l} \frac{\gamma}{\Gamma_l^{(m)}}}, \\ \Lambda_j^{(m)} &= \frac{\Pi_j^{(m)} b}{\Gamma_j^{(m)} - \gamma \Pi_j^{(m)}}. \end{aligned}$$

We determined R_0 using the eigenvalue sensitivity formula $R_0 = 1 + \mathbf{V}^{(i)T} \Delta \mathbf{L} \mathbf{U}^{(i)}$, where $\mathbf{U}^{(i)} = [U_1^{(i)}, ..., U_\omega^{(i)}]^T$ and $\mathbf{V}^{(i)} = [V_1^{(i)}, ..., V_\omega^{(i)}]^T$ are the leading right and left eigenvectors of the next generation matrix of the ith resident normalised as above. Taking $\Delta \mathbf{L}$ to be $\Delta \mathbf{L} = [\Delta L_{jl}] = \left[\frac{\partial L_{jl}^{(m)}}{\partial \theta_m}\right](\theta_m - \theta_i)$ (here the choice that the mutant differs from the resident only in θ is arbitrary), we obtain, to first order, $\log R_0 = \mathbf{V}^{(i)T} \left[\frac{\partial L_{jl}^{(m)}}{\partial \theta_m}\right] \mathbf{U}^{(i)} (\theta_m - \theta_i) = T_f^{(i)} \frac{\partial r}{\partial \theta_m} (\theta_m - \theta_i)$, and therefore

$$T_f^{(i)} = \frac{\mathbf{V}^{(i)T} \left[\frac{\partial L_{jl}^{(m)}}{\partial \theta_m}\right] \mathbf{U}^{(i)}}{\frac{\partial r}{\partial \theta_m}} \tag{S.10.6}$$

where all derivatives are evaluated at the ith resident ($\theta_m = \theta_i$, $\psi_{m,j} = \psi_{i,j}, j = 1, ..., \omega$).

As derived by Durinx et al. (2008), the speed factors of the canonical equation are

$$\kappa_i = \frac{\varepsilon B_i T_f^{(i)}}{\sigma_i^2} \tag{S.10.7}$$

where ε is the mutation rate, $B_i = b\Omega \sum_{j=1}^{\omega} n_j \hat{N}_{i,j}$ is the total rate of birth of the ith resident (where Ω is the (large) number of patches in the metapopulation), and σ_i^2 and $T_f^{(i)}$ are as given in (S.10.5) and (S.10.6), respectively. Without loss of generality, we fix $\varepsilon\Omega = 1$. Since B_i, $T_f^{(i)}$ and σ_i^2 depend on the trait values, κ_i changes as the residents evolve.

Numerical experiments

We investigated the dynamics of evolution by numerically solving the canonical equation for uncorrelated traits,

$$\begin{aligned} \frac{d\theta_i}{dt} &= \kappa_i w_\theta \left.\frac{\partial r}{\partial \theta_m}\right|_{\theta_m=\theta_i,\, \boldsymbol{\psi}_m=\boldsymbol{\psi}_i} \\ \frac{d\psi_{i,j}}{dt} &= \kappa_i w_\psi \left.\frac{\partial r}{\partial \psi_{m,j}}\right|_{\theta_m=\theta_i,\, \boldsymbol{\psi}_m=\boldsymbol{\psi}_i} \qquad \text{for } j = 1, ..., \omega \end{aligned} \tag{S.10.8}$$

jointly for residents $i = 1, ..., S$ (a system of $S \times (\omega + 1)$ differential equations). Because we are interested in orbits rather than in the explicit time dependence of the trait values, in each step of the numerical integration we normalised the vector of the right hand sides of the canonical equation, and forwarded the orbit in the $S \times (\omega + 1)$ dimensional trait space in the direction of the normalised vector by a small amount Δ (in the figures, $\Delta = 0.01$). This normalisation is a simple method of adaptive stepsize, and speeds up computations when the selection gradients are small (such as near singularities). Due to the trait space being high dimensional, we must plot the trait values against time steps to visualize the results, but one should keep in mind that these time steps correspond to unequal periods of real evolutionary time.

For our experiments, we fixed the model parameters as $b = 5.1$, $d_0 = 0.1$, $d_1 = 5/80$, $c = 1$, $\gamma = 5$, $\alpha = 1$, $\nu = 0.1$ (the same as in Figures 2 and 3 of the main text, but recall that here we

	w_ψ/w_θ	(θ_1, θ_2)	$\boldsymbol{\psi}_{1,2}$	diversification
1	1	$(-0.03,\ 0.03)$	(0.792, 1, 0.792)	yes* (Fig. S2)
2	1	$(-0.03,\ -0.015)$	(0.852, 1, 0.722)	no
	1	$(-0.03,\ 0.01)$	(0.818, 1, 0.766)	yes
3[†]	10	$(-0.03,\ 0.01)$	(0.823, 1, 0.761)	yes (Fig. S3)
	10	$(-0.03,\ -0.0014)$	(0.823, 1, 0.761)	no (yes)**
	10	$(-0.03,\ -0.0013)$	(0.823, 1, 0.761)	yes
	10	$(-0.03,\ 0.0298)$	(0.823, 1, 0.761)	yes
	10	$(-0.03,\ 0.03)$	(0.823, 1, 0.761)	no (yes)**
4	1	$(-0.03,\ -0.03)$	(0.852, 1, 0.722) (0.815, 1, 0.790)	yes (Fig. S4)

Table S1: Initial dimorphisms in the numerical experiments. The choice of initial dimorphisms and the outcomes are discussed in the text, the bold numbers in the first column refer to the sets of experiments. *Diversification occurs for any relative speed, tested also for $w_\psi/w_\theta = 0.01, 0.1, 10, 100$. **Loss of diversification is due to low population size of one of the residents. Diversification succeeds if mutation limitation is removed, so that the speed of evolution is not proportional to population size (see the subsection "Removing mutation limitation"). [†]See text for a brief summary of repeating this set of experiments with $w_\psi/w_\theta = 3$ and $w_\psi/w_\theta = 2$.

have no catastrophes) and assumed three patch types ($\omega = 3$) with

$$\begin{aligned} n_1 &= 0.25, & n_2 &= 0.5, & n_3 &= 0.25 \\ y_1 &= -1.5, & y_2 &= 0, & y_3 &= 1.5 \end{aligned}$$

It is easy to confirm that $\theta = 0$, $\boldsymbol{\psi} = (0.792, 1, 0.792)$ is a monomorphic ESS, where the local adaptation trait is under stabilising selection and $\boldsymbol{\psi}$ satisfies the weak ESS condition in (S.10.4). Table S1 lists the numerical experiments we detail below.

First, we initiated the dimorphic adaptive dynamics with symmetric residents near the monomorphic ESS, $\theta_{1,2} = \pm 0.03$, both with the ESS immigration strategy $\boldsymbol{\psi}_{1,2} = (0.792, 1, 0.792)$. As expected from our main model, the two residents evolve different immigration strategies to partition the peripheral niche; the resident with negative θ evolves higher immigration into patches of type 1 but lower into patches of type 3, whereas the resident with positive θ evolves the opposite

(Figure S2a,b). The local adaptation traits of the two residents initially converge towards each other (Figure S2c), as they would converge towards the ESS at $\theta = 0$ if their immigration strategies were fixed. However, as the residents evolve different immigration strategies and therefore start to use different patch types, their local adaptation traits diverge (Figure S2d).

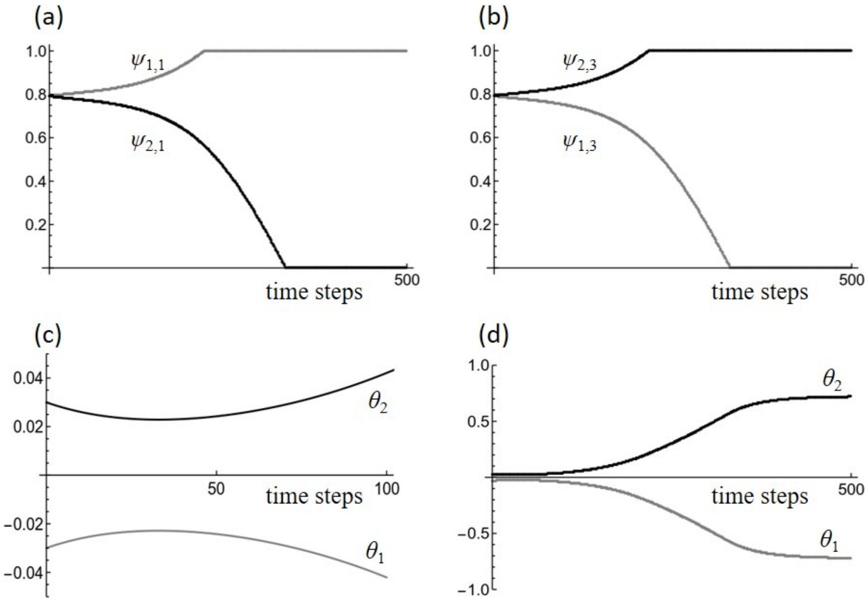

Figure S2: Dimorphic adaptive dynamics of the local adaptation trait and the immigration strategy with symmetric initial trait values $\theta_1 = -0.03$, $\theta_2 = 0.03$, $\boldsymbol{\psi}_{1,2} = (0.792, 1, 0.792)$, and $w_\theta = 1$, $w_\psi = 1$. Grey and black lines show respectively the first and the second resident. Panels (a) and (b), the evolution of immigration into patches of type 1 and of type 3, respectively; immigration into the middle patch type remains constant at $\psi_{1,2} = \psi_{2,2} = 1$. Panels (c) and (d), the evolution of the local adaptation trait; panel (c) is an enlargement from panel (d) for the critical initial period of the dynamics.

The relative speed of evolution of the local adaptation trait and of the immigration strategy can be set via the ratio of the mutational variances, w_θ and w_ψ. The results in Figure S2 were obtained assuming that these are of the same order of magnitude ($w_\theta = 1$, $w_\psi = 1$; note that since w_θ is in the squared units of the local adaptation trait but w_ψ is dimensionless, their values cannot be compared directly, i.e., without reference to the parameters that also depend on the unit of the local adaptation trait, c and $y_1, ..., y_\omega$). However, the results are qualitatively the same for any choice of w_θ and w_ψ (verified for $w_\theta / w_\psi = 0.01, 0.1, 10, 100$; data not shown). If the immigration strategy has low mutational variance compared to the local adaptation trait (and thus violates the assumption we make in the main model), then the resident local adaptation traits converge for a longer time and therefore get closer to $\theta = 0$. Since the selection gradient of the local adaptation trait vanishes at $\theta = 0$, in the vicinity the speed of its evolution scales with $\theta_{1,2}^2$. At the same time, selection on the immigration strategy scales with $\theta_{1,2}$. Once $\theta_{1,2}$ are sufficiently near zero, the immigration strategy evolves faster even if its mutational variance is small, which leads to the partitioning of the peripheral niche and diversification as described in the main text.

The symmetric initial point assumed above is however not likely to be the starting point in reality. For the remainder, we consider situations where a monomorphic resident evolves towards the singularity and the population becomes dimorphic because a mutant coexists with the original resident before reaching the singularity, which means an asymmtric starting point. For a second set of numerical experiments, we assumed $w_\theta = 1$, $w_\psi = 1$, and started with a monomorphic population $\theta = -0.428$, $\boldsymbol{\psi} = (1, 1, 0.001)$. This choice may be considered as arbitrary, but it is also motivated by the fact that $\theta = -0.428$, $\boldsymbol{\psi} = (1, 1)$ is the ESS in a metapopulation with only two patch types, y_1 and y_2, present in proportions $1 : 2$ (note the lack of the peripheral niche). We can thus imagine that the initial monomorphic population starts evolving because new patches of the third type y_3 become available, with proportions $1 : 2 : 1$ as assumed above. We evolved the initial monomorphic population until its local adaptation trait reached $\theta = -0.03$; by this time, its immigration strategy evolved to $\boldsymbol{\psi} = (0.852, 1, 0.722)$. This is then the point where we assume the population becomes dimorphic by the invasion of a mutant with different θ.

Due to the asymmetry of the immigration strategy $\boldsymbol{\psi} = (0.852, 1, 0.722)$ (immigration into the first type of patches is more likely than immigration into the third type), the local adaptation traits of mutants that can coexist with the resident $\theta = -0.03$ are also asymmetrically located, and in fact no coexistence is possible with a mutant $\theta > 0$. Therefore we chose the initial dimorphism to be $\theta_1 = -0.03$ and $\theta_2 = -0.015$, both sharing $\boldsymbol{\psi} = (0.852, 1, 0.722)$. These two strategies coexist and have comparable population sizes, but their evolution quickly leads to the loss of dimorphism as the first resident goes extinct (with other choices of θ_2, it can be the second resident which goes extinct).

This failure of diversification appears to be the consequence of the monomorphic immigration strategy "lagging" behind what would be the ESS for the current local adaptation trait. For a monomorphic population with $\theta = -0.03$, the ESS immigration strategy is $\boldsymbol{\psi} = (0.818, 1, 0.766)$, which is closer to symmetric. Again, this affects which mutants can coexist with the resident ($\theta_2 = -0.015$ cannot); starting the dimorphism with $\theta_1 = -0.03$ and $\theta_2 = 0.01$, both residents sharing $\boldsymbol{\psi} = (0.818, 1, 0.766)$, we get diversification (data not shown).

Speeding up the evolution of the immigration strategy reduces its lag. To confirm that faster evolution of the immigration strategy facilitates diversification, we conducted a third set of experiments assuming $w_\theta = 1$ and $w_\psi = 10$. Once again, we started a monomorphic population with $\theta = -0.428$ and $\boldsymbol{\psi} = (1, 1, 0.001)$ and evolved it until it reached $\theta = -0.03$; this time its immigration strategy evolved to $\boldsymbol{\psi} = (0.823, 1, 0.761)$, much closer to the ESS than before. At this point, we introduced a mutant as above to initialise the dimorphic adaptive dynamics with $\theta_1 = -0.03$, $\theta_2 = 0.01$, sharing $\boldsymbol{\psi} = (0.823, 1, 0.761)$. This dimorphic population undergoes diversification as predicted by our main model (Figure S3). The comparison of this result with our earlier experiments suggests that fast evolution of immigration is important to ensure a sufficiently symmetric inital dimorphism.

The choice of $\theta_2 = 0.01$ in Figure S3 is to some extent arbitrary. Given $\theta_1 = -0.03$ and the shared immigration strategy $\boldsymbol{\psi}_{1,2} = (0.823, 1, 0.761)$, coexistence is possible for $\theta_2 \in (-0.0015, 0.034)$. We found that initial dimorphisms close to the edge of the coexistence interval can fail to diver-

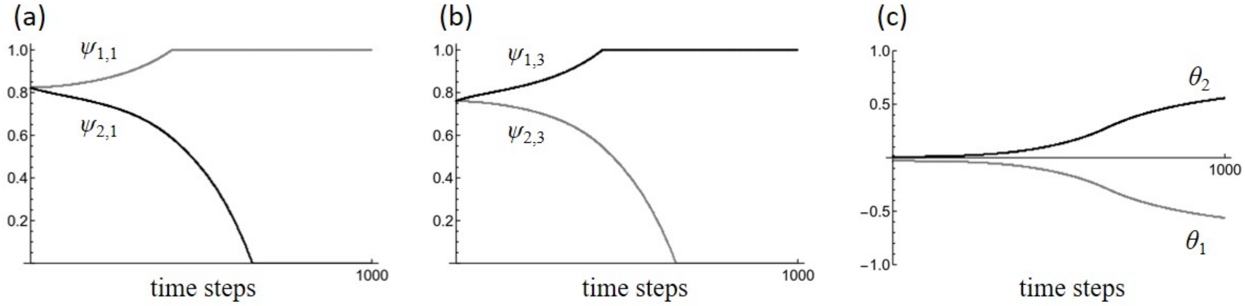

Figure S3: Dimorphic adaptive dynamics of the local adaptation trait and the immigration strategy with asymmetric initial trait values $\theta_1 = -0.03$, $\theta_2 = 0.01$, $\boldsymbol{\psi}_{1,2} = (0.823, 1, 0.761)$, and $w_\theta = 1$, $w_\psi = 10$. Grey and black lines show respectively the first and the second resident. Panels (a) and (b), the evolution of immigration into patches of type 1 and of type 3, respectively; immigration into the middle patch type remains constant at $\psi_{1,2} = \psi_{2,2} = 1$. Panel (c), the evolution of the local adaptation trait.

sify, and the reason for extinction is slow evolution due to small population size in one of the residents. Specifically, evolution with initial $\theta_2 \in (0.03, 0.034)$ ends in the extinction of the second resident; but failure occurs in a much narrower range at the lower end of the coexistence interval, an initial dimorphism with $\theta_2 = -0.0013$ succeeds to diversify. This asymmetry may be due to the fact that the initial, shared immigration strategy is less suitable for a resident with higher θ_2, and since its population size is small, it receives few mutations and cannot evolve fast enough to compensate. We note that most initial dimorphisms that fail to diversify are unlikely to arise in the first place, because $\theta_2 > 0.03$ represent the largest and therefore least likely mutation steps from $\theta_1 = -0.03$, and also mutations close to the upper end of the coexistence interval have low invasion fitness and therefore a low probability of establishment in face of demographic stochasticity.

For a brief exploration of how much speed difference is needed for diversification, we repeated the third set of experiments with lower values of w_ψ / w_θ. For $w_\psi / w_\theta = 3$, we found that diversification occurs with $\theta_2 \in (-0.003, 0.011)$, which is about half of the coexistence interval $(-0.007, 0.0206)$; again, diversification fails towards the high end of the coexistence interval. For

$w_\psi / w_\theta = 2$, diversification succeeds from only a narrow interval $\theta_2 \in (-0.001, 0.002)$ (compare with the coexistence interval $(-0.011, 0.012)$; the coexistence interval changes with the relative speed of evolution because speed affects the immigration strategy evolved during the monomorphic phase, which is the immigration strategy in the initial dimorphism).

Diversification may start also with the coexistence of two strategies differing in immigration, not in the local adaptation trait. We add a brief fourth experiment to demostrate this possibility in Figure S4 (but note that not all initial dimorphisms of this kind succeed to diversify). Speeding up the evolution of the immigration strategy may be detrimental to diversification in this scenario, because the two residents will evolve their immigration strategy to the same ESS before their local adaptation trait could diverge. Should this happen, however, then the resulting monomorphic population will be in a favourable position to diversify via a dimorphism of the local adaptation trait, as shown in our third set of experiments.

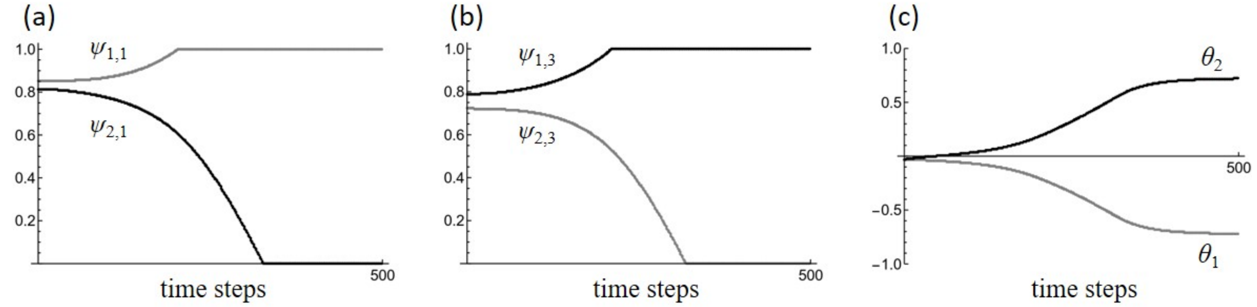

Figure S4: Diversification starting with two residents that differ in their immigration strategy, assuming $w_\theta = 1$, $w_\psi = 1$. The initial trait values are $\theta_{1,2} = -0.03$, $\boldsymbol{\psi}_1 = (0.852, 1, 0.722)$ (which is where monomorphic evolution ends, see text), and $\boldsymbol{\psi}_2 = (0.815, 1, 0.79)$. Notation as in Figure S3; immigration into the middle patch type remains constant at $\psi_{1,2} = \psi_{2,2} = 1$.

Removing mutation limitation

In the neighbourhood of evolutionary singularities, selection becomes slow and therefore mutation limitation is likely to break down. Removing mutation limitation, and instead assuming that selection operates on standing genetic variation, is equivalent to setting $\kappa_1 = \kappa_2 = 1$ in the canonical equation (S.10.8) (Abrams et al. 1993; Geritz et al. 2016). We have re-run all numerical experiments in Table S1 without mutation limitation. We found no difference in the results, except that diversification occurred also when one of the initial residents had a small population size. Under mutation limitation, this resident went extinct because it had not enough new mutations to fuel its evolution fast enough relative to the other resident, but without mutation limitation, population size is irrelevant and diversification succeeds.

Conclusions

The numerical experiments described above support diversification via the mechanism described in our main article, partitioning of the peripheral niche. We find that the evolution of the immigration strategy needs to be fast enough, but a strict timescale separation ($w_\psi / w_\theta \to \infty$) is not necessary. Fast evolution of the immigration strategy is important to ensure a sufficiently symmetric configuration of the initial dimorphism. Because behavioural strategies are commonly thought to evolve faster than morphological traits, the required speed difference seems realistic.

Our conclusions are drawn from a limited number of numerical experiments. We do not embark on a comprehensive analysis of the model described in this section, because this model contains an intrinsic inconsistency; it assumes costly dispersal with a positive emigration rate γ in a metapopulation where nothing selects for emigration. In our main model, catastrophes, an extreme kind of environmental stochasticity, create selection in favour of emigration.

Open problems

There are two open problems, one mathematical and one biological, which prevent us from deriving the canonical equation of adaptive dynamics in our main metapopulation model. Firstly, the canonical equation has not been derived for structured metapopulations with catastrophes. The canonical equation we used above assumes a point attractor of the ecological dynamics (Durinx et al. 2008). With catastrophes, the local populations are not at equilibrium, and therefore the density-dependent vital rate(s) change with the age of the patch. This seriously complicates the computation of the speed factors.

The second problem is that in our main model, the immigration strategy is a function-valued trait. The canonical equation has been derived for function-valued traits (Parvinen et al. 2013; Metz et al. 2016), so that this presents no mathematical problem, even if obtaining an approximate solution for the corresponding infinite-dimensional system would be computationally demanding. However, we would need to specify the covariance kernel analogous to the covariance matrix **C** above. For simplicity, in the above we assumed that **C** is a diagonal matrix, which means that the immigration probabilities $\psi_1, ..., \psi_\omega$ evolve independently. While this may be acceptable for a few distinct patch types, for a function-valued trait it would be unrealistic to assume that mutations affect $\psi(y)$ independently for each value of y, no matter how close by. We should therefore assume a nontrivial covariance kernel, but its correct shape is hard to establish.

References

All references are given at the end of the main article (see also the section "References cited only in the online enhancements" there).