

Joint evolution of dispersal and connectivity

Petteri Karisto* & Éva Kisdi[†]

Department of Mathematics and Statistics
University of Helsinki

PO Box 68, FIN-00014 Helsinki, Finland

Author Posting. © The Authors 2019. This is the Author's Version of the work. It is posted here for personal use, not for redistribution. The definitive version was published in *Evolution*, <http://dx.doi.org/10.1111/evo.13866>

Please cite the journal publication.

Keywords: adaptive dynamics, dispersal polymorphism, evolutionary branching, kin competition, patch connectivity

[†] Corresponding author. email: eva.kisdi@helsinki.fi.

* Present address: Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; email: petteri.karisto@ieu.uzh.ch.

Abstract

Functional connectivity, the realized flow of individuals between the suitable sites of a heterogeneous landscape, is a prime determinant of the maintenance and evolution of populations in fragmented habitats. While a large body of literature examines the evolution of dispersal propensity, it is less known how evolution shapes functional connectivity via traits that influence the distribution of the dispersers. Here we use a simple model to demonstrate that, in a heterogeneous environment with clustered and solitary sites (i.e., with variable structural connectivity), the evolutionarily stable population contains strains that are strongly differentiated in their pattern of connectivity (local vs global dispersal), but not necessarily in the fraction of dispersed individuals. Also during evolutionary branching, selection is disruptive predominantly on the pattern of connectivity rather than on dispersal propensity itself. Our model predicts diversification along a hitherto neglected axis of dispersal strategies and highlights the role of the solitary sites – the more isolated and therefore seemingly less important patches of habitat – in maintaining global dispersal that keeps all sites connected.

1. Introduction

For species inhabiting fragmented landscapes, dispersal between the patches of available habitat is of utmost importance (Ronce 2007; Clobert et al. 2012). Dispersal is determined, on the one hand, by the physical properties of the landscape (e.g. the distances between the patches and the presence of geographic barriers) and, on the other hand, by the mobility of the organism. The distinction between the properties of the landscape and the realized movement of individuals is captured in the distinction of structural *vs* functional connectivity (Tischendorf and Fahrig 2000).

As highlighted already by Hanski's classic incidence function model (Hanski 1994; Hanski and Ovaskainen 2000), functional connectivity depends on the traits of the organism, and is therefore subject to evolution. For example, winged insects can disperse more globally, whereas wingless forms disperse only to short distances. In wing-dimorphic insects, functional connectivity depends on the frequency of the winged form, a heritable trait (Zera and Denno 1997). In the butterfly *Melitaea cinxia*, individuals who carry the *Pgi-c* allele are more mobile, so that different genotypes realize different functional connectivities between the same patches (DiLeo et al. 2018). In plants, functional connectivity may depend on the animal vectors that aid seed dispersal, but also on traits of the plant that influence the use of these vectors (Auffret et al. 2017).

The theoretical literature examining the evolution of dispersal has put much emphasis on dispersal propensity (the fraction of individuals who leave their patch), and less on how structural connectivity affects the evolution of functional connectivity through the evolution of other mobility-related traits. In addition, most models of dispersal evolution assume landscapes with simple structural connectivity, such as dispersal via a global dispersal pool (e.g. Hamilton and May 1977; Levins and Hastings 1984; Olivieri et al. 1995; Gandon and Michalakis 1999; Mathias et al.

2001; Gyllenberg and Metz 2001; Parvinen 2002; Kisdi 2004; Parvinen and Metz 2008; Massol et al. 2011; Weigang and Kisdi 2015; Mullon et al. 2018; and many others) or a regular lattice of identical (Harada 1999; Rousset and Gandon 2002) or heterogeneous patches (Bonte et al. 2010; Cenzer and M'Gonigle 2019). In real landscapes, however, structural connectivity is much more variable. For example, the metapopulation of the butterfly *Melitaea cinxia* in the Åland islands inhabits several thousand meadows, which form 125 spatial clusters. Some of the clusters contain well over a hundred meadows, whereas some others have just one meadow (Hanski et al. 2017, see Supplementary Data), showing large differences in the structural connectivity of individual meadows.

Recent models of Henriques-Silva et al. (2015) and Karisto and Kisdi (2017) studied the evolution of dispersal propensity under variable structural connectivity, but assumed that functional connectivity is fully determined by dispersal propensity and the structure of the physical environment. This assumption implies that the patches have fixed relative connectivities; by increasing dispersal propensity, every patch becomes functionally more connected, with no change in which patches are more or less connected relative to others. However, since the traits of an organism influence not only its dispersal propensity but also its movement over the landscape, also relative connectivities can change during evolution, increasing or decreasing the variability of functional connectivity in a given landscape.

In this paper, we show in a simple model that with variable structural connectivity among the habitat patches, natural selection acting on functional connectivity cannot be captured by selection on dispersal propensity alone; and that it is predominantly functional connectivity, not simply dispersal propensity, that diversifies as a response to environmental heterogeneity.

We envisage a population inhabiting clustered and solitary living sites. Dispersing individuals can engage in global (long-range) or local (short-range) dispersal; the latter is enough to move between the sites of the cluster but not to solitary sites. We use adaptive dynamics (Geritz et al. 1998, 2016) to analyse the joint evolution of dispersal propensity and the pattern of functional connectivity, the latter measured by the fraction of long-range dispersal. The evolution of short- vs long-range dispersal means that functional connectivity evolves between the highly asymmetric pattern of only within-cluster dispersal, which leaves the solitary sites isolated, and the symmetric pattern of all sites being connected equally. In particular, we explore (based on Geritz et al. 2016) whether the jointly evolving traits undergo evolutionary branching, and we characterize evolutionarily stable dimorphisms of dispersal strategies where the coexisting strains strongly differ in whether they disperse globally or locally. These coexisting dispersal phenotypes realize contrasting patterns of functional connectivity on the same geographic network of sites.

2. The model

We consider a landscape with a large number of living sites. A site is defined as a habitat fragment small enough to be well-mixed, so that the physical structure of the landscape is captured by the size and structural connectivity of the sites. In general, the sites may vary in size, which can lead to dispersal polymorphisms even if all sites are equally connected (Massol et al. 2011; Kisdi 2016). Since here we want to focus on the effect of variable structural connectivity, we shall assume that the sites have the same size and each site supports one adult individual, as common for plants and other sessile organisms competing for space. Kin competition among the offspring selects for dispersal (Hamilton and May 1977).

Some of the sites may be within easy reach from each other, whereas others may be more isolated by distance or by some barriers. We simplify this variation assuming that a fraction p of the sites are solitary and the rest are part of a single spatial cluster. From each site, a fraction d of the offspring disperse. A fraction q of the dispersed offspring enter a global dispersal pool from where each site of the landscape can be reached with equal probability; the remaining fraction $1-q$ enter a local dispersal pool of the cluster (Figure 1). $q=0$ thus implies that the solitary sites are cut off, whereas $q=1$ means that every site of the landscape is equally well connected to every other site. As traits affecting mobility evolve, the dispersal propensity d as well as the fraction of globally dispersed offspring q evolve.

Global dispersal is more costly than local dispersal; locally dispersed offspring survive dispersal with probability s , but for globally dispersed offspring, the survival probability is only γs (with $\gamma < 1$). For the offspring dispersed from solitary sites, local dispersal is not an option. We thus assume that if they fail to engage in global dispersal (which happens with probability $1-q$), then with probability sb they survive and return to their natal site, otherwise they perish (Figure 1). For completely passive dispersal, $b = 0$; a positive value of b means that the offspring have some control over their movement to seek out a site. (Note that the probability of getting back to the natal site in the cluster is negligible because the cluster comprises many nearby sites where the offspring can settle.) We consider an annual species and assume that each parent has the same number of offspring, B , large enough for within-site demographic stochasticity to be negligible. Table 1 summarizes the notation.

To determine whether a mutant dispersal strategy is able to invade the resident population, we calculate its reproductive output. Let a_{ij} be the number of solitary ($i=1$) and clustered ($i=2$) sites secured by the offspring of a mutant parent who itself occupies a solitary ($j=1$) or a clustered ($j=2$)

site. A mutant in a solitary site has $(1 - d_m)B$ offspring that remain in the natal site plus $d_m(1 - q_m)sbB$ offspring who return after failing to disperse globally. These mutant offspring compete for their natal site with immigrants of the resident strategy. In a fair lottery competition, the probability that the winner is a mutant is given by

$$\frac{[1 - d_m + d_m(1 - q_m)sb]B}{[1 - d_m + d_m(1 - q_m)sb]B + I_G B}$$

where $I_G B$ is the number of immigrants arriving from the global dispersal pool. We factor B , i.e., we write $I_G B$ for the number of immigrants to keep I_G finite when the number of offspring B goes to infinity. In addition, a mutant parent in a solitary site has $d_m q_m s \gamma B$ offspring dispersed globally, each of whom arrives at a solitary site with probability p and wins that site with probability P_S/B . The probability that one particular offspring wins the site becomes infinitesimal as B goes to infinity, and we write P_S/B for this probability so that P_S remains non-vanishing. The expected number of solitary sites obtained by the offspring of a mutant parent in a solitary site is therefore

$$a_{11} = \frac{1 - d_m + d_m(1 - q_m)sb}{1 - d_m + d_m(1 - q_m)sb + I_G} + d_m q_m s \gamma p P_S \quad (1a)$$

Denoting the number of immigrants from the local dispersal pool of the cluster with $I_L B$ and the probability of winning competition in a clustered site with P_C/B , derivations similar to the above yield

$$\begin{aligned}
a_{12} &= d_m q_m s \gamma p P_S \\
a_{21} &= d_m q_m s \gamma (1-p) P_C \\
a_{22} &= \frac{1-d_m}{1-d_m + I_G + I_L} + d_m (1-q_m) s P_C + d_m q_m s \gamma (1-p) P_C
\end{aligned} \tag{1b}$$

The quantities I_G , I_L , P_S and P_C depend on the resident population and are given by straightforward calculations detailed in Section 1 of the Electronic Appendix.

The four numbers in equations (1a,b) are the elements of the mutant's 2x2 projection matrix. The mutant can invade if the leading eigenvalue of its projection matrix is greater than 1 (for the ease of analysis, we use the fitness proxy of Metz and Leimar 2011 instead of the eigenvalue). We envisage evolution by repeated invasions of mutants, and use the techniques of multivariate adaptive dynamics to analyse the long-term evolution of the dispersal traits d and q (Durinx et al. 2008; Leimar 2009; Geritz et al. 2016).

3. Results

(a) Monomorphic resident populations

We first consider the case of a single resident strategy, i.e., a population where (apart from rare mutations) all individuals have the same d and q . A combination of analytical and numerical explorations detailed in Section 2 of the Electronic Appendix shows that two inequalities determine the course of evolution,

- (i) $p > 1 - \gamma$ and
- (ii) $sb > [p - (1 - \gamma)] / p\gamma$.

The first inequality means that there are sufficiently many solitary sites. The second inequality holds when failing global dispersal from a solitary site is not too dangerous (recall that sb is the probability that an offspring who left a solitary site but did not succeed to get in the global dispersal

pool finds its way back to its natal site). Note that if (i) is violated, then the right hand side of (ii) is negative so that (ii) always holds. There are therefore only three cases to distinguish: both inequalities hold; (i) is violated; (i) holds but (ii) is violated.

If both inequalities hold, then there is an interior evolutionary singularity (d^*, q^*) , where the joint evolution of the dispersal traits comes to a halt (Figure 2a). This singularity is nearly always strongly convergence stable, i.e., it attracts the evolutionary dynamics of d and q irrespectively of the covariation between the two traits (Leimar 2009; a narrow parameter region where this is not the case is described in Section 4.1 of the Electronic Appendix). Convergence for any covariation is important because mutations affecting the probability of leaving the natal site d likely also affect the probability of engaging in long-range dispersal q and *vice versa*. To quantify the mutational covariance of d and q , however, we would need more information about the organism's morphological and behavioural traits that influence d and q and the genetic architecture thereof.

Depending on the parameter values, the singularity (d^*, q^*) may or may not be an ESS. Unlike in the case of a single evolving trait, in higher dimensional trait spaces a convergence stable non-ESS singularity is not necessarily an evolutionary branching point (Geritz et al. 2016). For two evolving traits as in our case, Geritz et al. (2016) have shown that the combination of strong convergence stability and lack of evolutionary stability implies evolutionary branching when selection acts on standing genetic variation (i.e., evolution is not mutation-limited). Hence we consider our strongly convergence stable, non-ESS interior singularities to be evolutionary branching points.

Evolutionary branching at an interior singularity occurs when the fraction of solitary sites is intermediate (p is higher than $1 - \gamma$ but not by too much; see Figure 3). This is straightforward to interpret: If both solitary and clustered sites are common, the population splits into two lineages specializing on the two types of sites. An example for evolutionary branching at an interior

singularity is shown in Figure 4a. The evolutionary trajectories shown in this figure have been obtained by numerically solving the canonical equation of adaptive dynamics, see Section 3 of the Electronic Appendix for details.

If inequality (i) is violated, the fraction of solitary sites (p) is too small to maintain global dispersal in a monomorphic population, so that q evolves towards zero and the solitary sites become effectively isolated (Figure 2b). However, when the resident's q is sufficiently close to zero, a mutant with somewhat higher q (and not too different d) can invade and coexist with the resident, whereupon the two strategies experience disruptive selection in the q -direction. Mathematical tools are lacking for this case, but, as explained in Section 4.3 of the Electronic Appendix, we expect that, at least if the genetic correlation between d and q is not strong, evolutionary branching happens near $q = 0$. Notice that branching can happen away from a singularity (Figure 4b). In some instances, one of the two branches goes extinct soon after separation; in this case, the remaining strategy evolves back to $q = 0$ and attempts branching again (see Section 4.3 of the Electronic Appendix).

Finally, if inequality (i) holds but (ii) is violated, then q evolves to 1 (Figure 2c). This is the case when many of the sites are solitary and global dispersal is strongly favoured because offspring dispersed locally from solitary sites die (b is low when inequality (ii) is violated) and/or global dispersal is safe (violation of (ii) is equivalent to $\gamma > (1 - p) / (1 - sbp)$). At $q = 1$, all dispersal is global and therefore all sites are equally connected. The evolution of dispersal propensity is then governed by the Hamilton-May (1977) model with survival probability $s\gamma$, leading to the ESS at $d = 1 / (2 - s\gamma)$. We note that this endpoint is only locally evolutionarily stable; it is easy to verify that a strategy with the same dispersal propensity but only local dispersal ($q = 0$) can invade and coexist with it (Section 4.2 of the Electronic Appendix).

As Figure 3 shows, ESSs with intermediate trait values are uncommon (see also Figure S2 in the Electronic Appendix). Interior branching points occur more frequently, but the two most common outcomes are the ESSs with fully global dispersal ($q = 1$) and disruptive selection near $q = 0$. Only the latter two outcomes are possible if $b = 0$, i.e., if offspring dispersed from a solitary site perish if they fail to disperse globally.

When evolutionary branching happens, disruptive selection acts predominantly in the q -direction. At interior singularities, the direction of strongest disruptive selection is typically within a few degrees from vertical (black trajectory in Figure 4a; Figure S4 in the Electronic Appendix). At the boundary $q = 0$, disruptive selection is always in the q -direction (Figure 4b; Section 4.3 of the Electronic Appendix). We therefore conclude that the diversification of dispersal strategies is mainly due to disruptive selection on the pattern of functional connectivity (determined by q), not dispersal propensity (d) itself. Note, however, that the realized coevolutionary trajectories depend also on the (unknown) mutational covariance of d and q . With a strong positive covariance, the line of divergence turns towards 45 degrees (grey trajectory in Figure 4a).

(b) Dimorphic resident populations

Our numerical experiments showed that coevolution often leads to extremes in the fraction of globally dispersed offspring, i.e., one resident with $q = 0$ and the other with $q = 1$ (as in Figure 4). Therefore, we searched for dimorphic singularities where a strategy with only local dispersal, $(d_1^*, 0)$, coexists with one that disperses only globally, $(d_2^*, 1)$. The values of d_1^* and d_2^* are such that the selection gradient vanishes in the d -direction, whereas selection on q pushes towards the boundary 0 and 1, respectively. For most parameter values, such a dimorphic singularity exists

(see Figure S5 in the Electronic Appendix). For the rest, we found a dimorphic singularity where a locally dispersing strategy $(d_1^*, 0)$ coexists with a strategy that mixes local and global dispersal, but also here the fraction of globally dispersed offspring is high $((d_2^*, q_2^*)$ with $q_2^* > 0.5$). Note that this latter type of singularity occurs only for high values of b and s (Figure S5 in the Electronic Appendix), which may be less common in nature. When b and s are high, offspring who leave a solitary site but fail to engage in global dispersal likely do not perish but return to their natal site. This means less selection for global dispersal in the solitary sites, so that $q_2^* < 1$ can evolve, taking the advantage of local dispersal in the clustered sites.

All the above dimorphic singularities are both strongly convergence stable and evolutionarily stable, irrespective of whether the monomorphic population undergoes evolutionary branching or not. Note that a dimorphic ESS often exists also if a monomorphic population evolves to a local ESS (Geritz et al. 1999); in this case, the dimorphic ESS cannot be reached gradually via evolutionary branching, but it can be reached via the immigration of a second strategy sufficiently different from the monomorphic resident. In our model, a dimorphic ESS always exists, even if almost all sites are solitary or almost all are clustered (the relative abundance of the two residents changes accordingly).

The most prominent property of these dimorphic ESSs is that the two resident strategies have a large difference in q . This implies contrasting patterns of functional connectivity realized in the same landscape by the two residents.

The dispersal propensities of the two residents qualitatively behave as, and are often quantitatively close (but not equal) to the Hamilton-May strategies corresponding to the costs of local and global dispersal, $d_1 = 1/(2-s)$ and $d_2 = 1/(2-\gamma s)$, respectively. In particular, the locally dispersing resident

$(d_1^*, 0)$, which is restricted to the clustered sites, is always the one with higher dispersal propensity ($d_1^* > d_2^*$). Hence in this cluster-specialist strategy, there is selection for high but strictly local dispersal. Conversely, in the globally dispersing resident, selection promotes long-range dispersal but with a more moderate propensity. If dispersal-related traits (such as wings etc.) simultaneously facilitate leaving the natal site and moving longer distances (analogous to the global dispersal pool in our model), then d and q are positively correlated, and this correlation needs to be broken to arrive at the dimorphic ESS. The evolution of multiple dispersal mechanisms (such as ballooning for long-range dispersal vs rappelling over short distances in spiders; Bonte et al. 2008) or more sophisticated behavioural strategies may thus be expected, which can disentangle dispersal propensity from dispersal distance. The problem of opposite selection for dispersal propensity and dispersal distance is most acute when d_1^* is substantially different from d_2^* . This is the case when local dispersal has little cost (high s) but global dispersal is costly (low γ), a situation likely in nature.

4. Discussion

Our main result is that in a heterogeneous landscape of solitary and clustered sites, predominantly traits that influence the pattern of functional connectivity are under disruptive selection. When the population undergoes evolutionary branching, the two emerging lineages experience selection for global vs local dispersal rather than for many vs few offspring dispersed. Many traits may affect both the propensity and the pattern of dispersal, yet their effect on the pattern is more important for diversification. We found evolutionarily stable dimorphisms with one strain dispersing only locally in the cluster and the other dispersing globally among all sites. Due to these contrasting dispersal patterns, the two strains realize very different connectivities among the sites of the same geographic network.

The coexistence of globally and locally dispersing strains is easy to understand in terms of costs and benefits. In our model, dispersal is favoured by competition among the offspring of the same parent (Hamilton and May 1977). For an individual in a clustered site, local dispersal is sufficient to prevent kin competition. This maintains a strain with local dispersal, enjoying the lower dispersal costs but being restricted to the cluster. The unused resources of the solitary sites then favour a second, globally dispersing strain, which is at a disadvantage in the cluster due to its higher dispersal costs, but it is the only exploiter of the solitary sites. We can thus view the locally and globally dispersing strategies as specialists adapted to the clustered and solitary sites, respectively.

The details of the selective forces acting on the dispersal strategy are however less straightforward to explain heuristically (see Gandon and Michalakis 1999; Kisdi 2004; Karisto and Kisdi 2017 for more detailed discussions). Whether somewhat more global or somewhat more local dispersal is more profitable depends on the costs of global and local dispersal and on how many competitors a successfully dispersed offspring will encounter in a solitary or in a clustered site. The propensity of dispersal, i.e., how many of the offspring should be dispersed in the first place and how many should be retained in the natal site, depends on the above and also on how many immigrants arrive at the natal site. The latter is because the more immigrants the offspring have to compete with, the higher is the marginal value of a non-dispersed sibling. Because the number of competitors depends on the prevailing dispersal strategy, it is far from obvious that selection is disruptive on the pattern of dispersal, and, for most parameter values, selection maintains a fully globally dispersing strain and a fully locally dispersing strain in an evolutionarily stable dimorphism.

In homogeneous landscapes with identical habitat patches, dispersal propensity, when it is the only evolving trait, usually attains a monomorphic ESS (Hamilton and May 1977; Levin et al. 1984;

Olivieri et al. 1995; but see Ludwig and Levin 1991). Evolutionary branching can happen if dispersal is constrained by trade-offs with other life history parameters (Fronhofer et al. 2011; Weigang and Kisdi 2015) or co-evolves with other traits such as social behaviour (Mullon et al. 2018). In contrast to homogeneous systems, heterogeneity is expected to promote diversification. Indeed, a number of previous studies have demonstrated evolutionary branching of dispersal propensity in heterogeneous landscapes (Doebeli and Ruxton 1997; Mathias et al. 2001; Parvinen 2002; Parvinen and Metz 2008; Massol et al. 2011; Kisdi 2016; Karisto and Kisdi 2017; etc), but these studies did not allow the pattern of dispersal to evolve.

In spatially explicit models, the pattern of dispersal is given by the distribution of dispersal distance. As expected, in homogeneous landscapes there is an evolutionarily stable distribution (Rousset and Gandon 2002; Fronhofer et al. 2015), and somewhat more surprisingly, North et al. (2011) found only ESSs in a heterogeneous landscape. The simulation study of Bonte et al. (2010) showed local adaptation of dispersal distance on a lattice such that individuals in large patches of suitable habitat exhibited longer dispersal distances. On a homogeneous lattice, Harada (1999) determined the evolutionarily stable fraction of locally (i.e., to nearest neighbour) vs globally dispersed offspring. The recent simulation model of Cenzer and M'Gonigle (2019) assumed that two resources are distributed with different spatial autocorrelation on a lattice, and found that the mean dispersal kernel differs between individuals using different resources even if the resources are fully substitutable. It is however unclear in this model whether selection is disruptive, or mutations maintain variation in the dispersal kernel against stabilizing selection and the result is due to spatial sorting of this variation. Evolutionary branching of dispersal distance due to temporal and spatial heterogeneity is predicted by the model of Snyder (2006), which assumes that the offspring are distributed according to an exponential distribution in continuous space, with the mean of the distribution being the trait under disruptive selection. We found evolutionary branching of the

dispersal pattern under very different ecological assumptions, which highlights the need for a more comprehensive understanding of the selective mechanisms controlling the diversity of dispersal patterns.

The present model shows that disruptive selection can be stronger on the pattern of dispersal than on its propensity, but dispersal propensity can still diversify even if the pattern is fixed. Earlier, we investigated a simpler model (Karisto and Kisdi 2017), where the fraction of global dispersal (q) was a fixed parameter rather than an evolving trait (and we also assumed $b=0$). We found that heterogeneous structural connectivity can lead to the diversification of dispersal propensity (d). Since here we have shown that disruptive selection acts mostly on q , we conclude that the diversification of d we found earlier is driven by residual selection on dispersal propensity that remains after fixing the pattern of dispersal. This is reflected also in the evolutionarily stable dimorphism: whereas in the present model selection always maintains a dimorphism, this is not the case if only the propensity of dispersal can evolve (Karisto and Kisdi 2017).

If we assume no cost to local dispersal ($s=1$) and we fix dispersal propensity at $d=1$, then the clustered sites behave as a large, well-mixed habitat patch. Assuming further that there is no loss when attempting local dispersal from a solitary site ($b=1$) so that forcing $d=1$ does not harm the offspring in solitary sites, we arrive at a model of variable patch size (one large patch and many small sites) with global dispersal given by q . In the context of our present model, this is an extreme and unrealistic choice of parameters, but mathematically it is identical to the model of Kisdi (2016), who showed evolutionary branching (of q , in our present terms) due to the heterogeneity of patch size (see also Massol et al. 2011). For a more realistic model of heterogeneous landscapes, one should vary both the size of the habitat patches and their physical isolation, including multiple clusters that differ in the number of patches and in the cost of dispersal, with the solitary sites as

extremes of this variation. We expect that such a landscape would facilitate further diversification. For example, large well-mixed patches maintain a strain without any dispersal, clustered small sites maintain high but mostly local dispersal, and solitary sites favour global dispersal at a lower propensity.

Our results underscore the role of solitary sites, or more generally, the role of small and remote or otherwise separated patches of habitat. These may not appear to be significant for the population, but in fact the presence of the solitary sites maintains global dispersal. This remains true also when the landscape contains not one but several clusters, which are connected to the solitary sites as well as to each other via global dispersal. In each of the clusters, global dispersal is at a disadvantage, because it is more risky and does not provide more benefit; local dispersal within a cluster removes kin competition, and global dispersal between clusters does not add to this. In contrast, global dispersal is essential for individuals living in solitary sites. If the solitary sites are removed, then the globally dispersing strain goes extinct. Next to the loss of diversity, this increases the risk of global extinction, since in the absence of global dispersal, an extinct cluster cannot be recolonized from the rest of the network. Global dispersal is essential to maintain connectivity between the clusters, and global dispersal is maintained by the solitary sites.

Acknowledgement. This work was financially supported by the Academy of Finland through its Centre of Excellence in Analysis and Dynamics Research. We thank two anonymous reviewers and the Associate Editor, Florence Débarre, for detailed comments.

References

- Auffret A. G., Y. Rico, J. M. Bullock, D. A. P. Hooftman, R. J. Pakeman, M. B. Soons, A. Suarez-Esteban, A. Traveset, H. H. Wagner & S. A. O. Cousins. 2017. Plant functional connectivity – integrating landscape structure and effective dispersal. *J. Ecology* 105: 1648-1656.
- Bonte D., T. Hovestadt & H.-J. Poethke. 2010. Evolution of dispersal polymorphism and local adaptation of dispersal distance in spatially structured landscapes. *Oikos* 119: 560-566.
- Bonte D., J. M. J. Travis, N. De Clercq, I. Zwertvaegher & L. Lens. 2008. Thermal conditions during juvenile development affect adult dispersal in a spider. *Proc. Natl. Acad. Sci.* 105: 17000-17005.
- Cenzer M. & 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, Oxford.
- DiLeo M. F., A. Husby & M. Saastamoinen. 2018. Landscape permeability and individual variation in a dispersal-linked gene jointly determine genetic structure in the Glanville fritillary butterfly. *Evolution Letters* 2: 544-556.
- Doebeli M. & G. D. Ruxton. 1997. Evolution of dispersal rates in metapopulation models: Branching and cyclic dynamics in phenotype space. *Evolution* 51: 1730-1741.

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

Fronhofer E. A., A. Kubisch, T. Hovestadt & H. Poethke. 2011. Assortative mating counteracts the evolution of dispersal polymorphisms. *Evolution* 65: 2461-2469.

Fronhofer E. A., H. J. Poethke & U. Dieckmann. 2015. Evolution of dispersal distance: Maternal investment leads to bimodal dispersal kernels. *J. theor. Biol.* 365: 270-279.

Gandon S. & Y. Michalakis. 1999. Evolutionarily stable dispersal rate in a metapopulation with extinctions and kin competition. *J. theor. Biol.* 199: 275-290.

Geritz, S. A. H., E. Kisdi, G. Meszéna & 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., E. van der Meijden & J. A. J. Metz. 1999. Evolutionary dynamics of seed size and seedling competitive ability. *Theor. Pop. Biol.* 55: 324-343.

Geritz S. A. H., J. A. J. Metz & 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 J. A. J. Metz. 2001. On fitness in structured metapopulations. *J. Math. Biol.* 268: 545-560.

Hamilton W. D. & R. M. May. 1977. Dispersal in stable habitats. *Nature* 269: 578-581.

Hanski I. 1994. A practical model of metapopulation dynamics. *J. Anim. Ecol.* 63: 151-162.

Hanski I. & O. Ovaskainen. 2000. The metapopulation capacity of a fragmented landscape. *Nature* 404: 755-758.

Hanski I., T. Schulz, S. C. Wong, V. Ahola, A. Ruokolainen & S. P. Ojanen. 2017. Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. *Nature Communications* 8: 14504.

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

Henriques-Silva R., F. Boivin, V. Calgano, M. C. Urban & P. R. Peres-Neto. 2015. On the evolution of dispersal via heterogeneity in spatial connectivity. *Proc. R. Soc. B* 282: 20142879.

Karisto P. & E. Kisdi. 2017. Evolution of dispersal under variable connectivity. *J. theor. Biol.* 419: 52-65.

Kisdi E. 2004. Conditional dispersal under kin competition: Extension of the Hamilton-May model to brood size-dependent dispersal. *Theor. Pop. Biol.* 66: 369-380.

Kisdi E. 2016. Dispersal polymorphism in stable habitats. *J. theor. Biol.* 392: 69-82.

Leimar O. 2009. Multidimensional convergence stability. *Evol. Ecol. Res.* 11: 191-208.

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

Ludwig D. & S. A. Levin. 1991. Evolutionary stability of plant communities and the maintenance of multiple dispersal types. *Theor. Pop. Biol.* 40: 285-307.

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

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

Metz J. A. J. & O. Leimar. 2011. A simple fitness proxy for structured populations with continuous traits, with case studies on the evolution of haplo-diploids and genetic dimorphisms. *J. Biol. Dyn.* 5: 163-190.

Mullon C., L. Keller & L. Lehmann. 2018. Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nature Ecology and Evolution* 2: 132-140.

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

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

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

Parvinen K. & J. A. J. Metz. 2008. A novel fitness proxy in structured locally finite metapopulations with diploid genetics, with an application to dispersal evolution. *Theor. Pop. Biol.* 73: 517-528.

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.

Rousset F. & S. Gandon. 2002. Evolution of the distribution of dispersal distance under distance-dependent cost of dispersal. *J. Evol. Biol.* 15: 515-523.

Snyder R. E. 2006. Multiple risk reduction mechanisms: Can dormancy substitute for dispersal? *Ecology Letters* 9: 1106-1114.

Tischendorf L. & L. Fahrig. 2000. On the usage and measurement of landscape connectivity. *Oikos* 90: 7-19.

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

Zera A. J. & R. F. Denno. 1997. Physiology and ecology of dispersal polymorphism in insects.
Annu. rev. Entomol. 42: 207-230.

Table 1. Notation

Symbol	Definition
d (d_m)	resident (mutant) dispersal probability
q (q_m)	resident (mutant) probability that a disperser enters the global pool
p	fraction of solitary sites
s	probability of survival during local dispersal
γs	probability of survival during global dispersal (with $\gamma < 1$)
bs	probability that an offspring who dispersed from a solitary site but failed to enter the global dispersal pool returns to its natal site (with $b < 1$)
B	number of offspring per parent (assumed large)
$I_G B$	number of immigrants from the global pool
$I_L B$	number of immigrants from the local pool (clustered sites only)
P_S/B	probability that a focal individual in a solitary site wins the site
P_C/B	probability that a focal individual in a clustered site wins the site

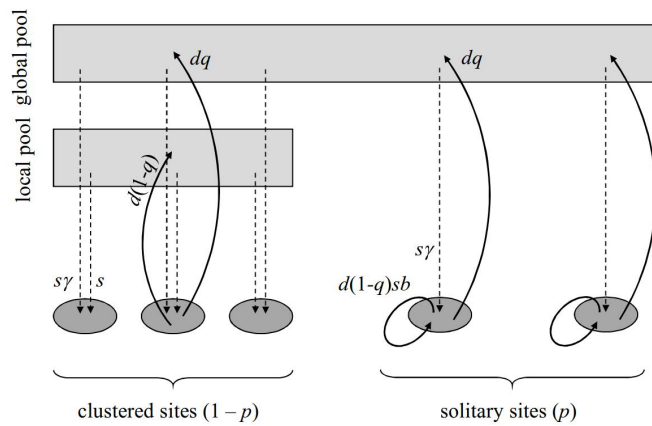


Figure 1. Clustered and solitary sites. From clustered sites, a fraction $d(1-q)$ of the offspring enter the local dispersal pool, and a fraction dq enter the global dispersal pool. Solitary sites have no local dispersal pool, but dispersed offspring who fail to enter the global pool return with probability sb to their natal site. The probability of survival for local dispersal is s , for global dispersal is $s\gamma$.

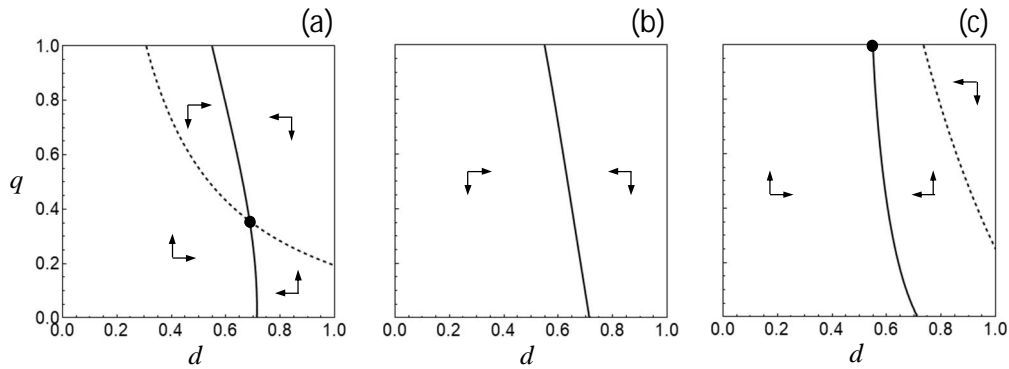


Figure 2. Monomorphic evolutionary dynamics of dispersal propensity (d) and the fraction of long-range dispersal (q). (a) Both inequalities (i) and (ii) hold, interior singularity ($p = 0.75$, $\gamma = 0.3$, $s = 0.6$, $b = 0.9$); (b) inequality (i) is violated, evolution to $q = 0$ (parameters as in (a) except $p = 0.4$); (c) inequality (i) holds but (ii) is violated, boundary ESS (parameters as in (a) except $b = 0.2$).

The arrows mark the direction of selection, which changes across the d - and q -isoclines (continuous and dashed lines, respectively). The dot marks the singularity (d^*, q^*) , which may be an ESS or an evolutionary branching point (evolution does not reach a singularity in panel (b), see the main text and Figure 4b).

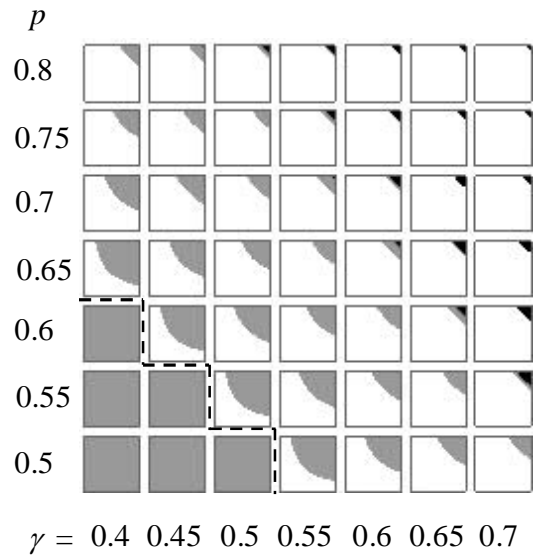


Figure 3. Monomorphic evolutionary singularities. In each panel, the horizontal axis is for s (between 0 and 1) and the vertical axis is for b (also between 0 and 1). Grey: evolution to disruptive selection either at an interior branching point (inequality (i) is satisfied above the dashed line) or near $q = 0$ (below the dashed line); white: ESS with $q = 1$; black: interior ESS. The tendencies seen in this figure continue for values of p and γ not shown here (see Figure S2 in the Electronic Appendix).

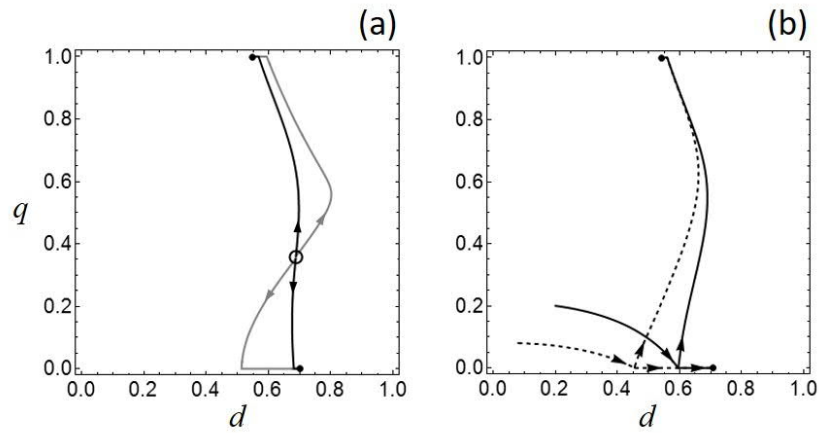


Figure 4. Evolutionary branching and the coevolution of two resident strategies. In (a), evolutionary branching occurs at an interior singularity (circle). The two branches evolve away from each other (arrows) and eventually converge to the dimorphic ESS, where one resident has $q = 0$ and the other has $q = 1$ (black dots). The two trajectories (black and grey) differ in the mutation covariance matrix. Black: d and q mutate with equal variance and zero correlation; grey: equal variance and high positive correlation ($r = 0.8$). In (b), evolutionary branching occurs at $q = 0$. The two trajectories differ only in their initial points; for both, d and q mutate with equal variance and zero correlation. Parameters: (a) $p = 0.75$, $\gamma = 0.3$, $s = 0.6$, $b = 0.9$ (as in Figure 2a); (b) $p = 0.4$, $\gamma = 0.3$, $s = 0.6$, $b = 0.9$ (as in Figure 2b).

Electronic Appendix

This Electronic Appendix contains all mathematical derivations and numerical procedures used for the analysis of the model described in the main text. Unless mentioned otherwise, we performed the analyses with *Mathematica* (Wolfram Research) versions 10.0, 10.2, and 11.3 (the code is available at <https://doi.org/10.5061/dryad.9ghx3ffd0>).

1 Invasion fitness proxy

The invasion fitness of the mutant strategy (d_m, q_m) is the leading eigenvalue of its projection matrix

$$\mathbf{A}_m = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

with the elements

$$\begin{aligned} a_{11} &= \frac{1 - d_m + d_m(1 - q_m)sb}{1 - d_m + d_m(1 - q_m)sb + I_G} + d_m q_m s \gamma p P_S \\ a_{12} &= d_m q_m s \gamma p P_S \\ a_{21} &= d_m q_m s \gamma (1 - p) P_C \\ a_{22} &= \frac{1 - d_m}{1 - d_m + I_G + I_L} + d_m(1 - q_m)s P_C + d_m q_m s \gamma (1 - p) P_C \end{aligned} \tag{A.1}$$

as given in equations (1) of the main text. It remains to obtain I_G , I_L , P_S , and P_C , which depend on the resident population.

Let n_{1i} and n_{2i} denote respectively the number of solitary and clustered sites occupied by the i th resident dispersal strategy (d_i, q_i) , given as the fractions of all sites (i.e., $\sum_i (n_{1i} + n_{2i}) = 1$). If there is only one resident strategy, then $n_{11} = p$ and $n_{21} = 1 - p$; for multiple residents, the equilibrium densities must be determined numerically. The number of immigrants arriving at a site from the global dispersal pool is given by

$$I_G = s \gamma \sum_i d_i q_i (n_{1i} + n_{2i}) \tag{A.2a}$$

From the local dispersal pool of clustered sites, the number of immigrants arriving at a clustered site is

$$I_L = s \sum_i d_i (1 - q_i) n_{2i} / (1 - p) \tag{A.2b}$$

where the division with $(1 - p)$ is because these dispersers are distributed only among the clustered sites.

The probability of winning a site depends on its former occupant. A solitary site is occupied by the i th resident with probability n_{1i}/p (which is 1 if there is only a single resident). The i th resident produces $(1 - d_i)B$ nondispersing and $d_i(1 - q_i)sbB$ returning offspring, so that the probability that a particular offspring wins the site is $1/[1 - d_i + d_i(1 - q_i)sb + I_G]B$. Recall from the main text that P_S is defined such that P_S/B is the probability of winning the site. Summing over all resident types we obtain

$$P_S = \sum_i \frac{n_{1i}}{p} \cdot \frac{1}{1 - d_i + d_i(1 - q_i)sb + I_G} \quad (\text{A.3a})$$

and an analogous calculation yields

$$P_C = \sum_i \frac{n_{2i}}{1 - p} \cdot \frac{1}{1 - d_i + I_G + I_L} \quad (\text{A.3b})$$

We assume $d, q > 0$ for all strategies and $0 < p < 1, s, \gamma > 0$; these assumptions guarantee that the projection matrix is irreducible (and then also primitive). Mathematically, the mutant's invasion fitness is not well defined if its projection matrix is not irreducible. Biologically, $d = 0$ or $q = 0$ is not realistic since some dispersal (both local and global) will occur even if only by accident. Hence we consider $(0, 1] \times (0, 1]$ to be the trait space of dispersal strategies (d, q) .

Let λ denote the leading eigenvalue of \mathbf{A}_m . The mutant can invade if λ is greater than 1; or, equivalently for a 2×2 projection matrix, if one or both of the following conditions hold (Metz and Leimar 2011):

$$\text{(a) } \text{tr } \mathbf{A}_m > 2, \quad \text{(b) } F = \text{tr } \mathbf{A}_m - \det \mathbf{A}_m > 1$$

where tr and \det denote the trace and the determinant, respectively. For a resident strategy, the trace of the projection matrix is always less than 2 (Metz and Leimar 2011), and, by continuity, the same holds for mutants with trait values close to a resident. Hence, assuming all mutations have sufficiently small effects on the dispersal strategy, condition (a) never holds. It then follows that the mutant's leading eigenvalue is greater than 1 if and only if condition (b) holds, i.e., if the mutant's fitness proxy $F = \text{tr } \mathbf{A}_m - \det \mathbf{A}_m$ exceeds 1.

By repeated mutations and invasions, the traits (d_i, q_i) of the i th resident evolve in the trait space guided by the selection gradients. Assuming small mutations, the mutant (d_m, q_m) near the i th resident can invade if $g_d^{(i)}(d_m - d_i) + g_q^{(i)}(q_m - q_i)$ is positive. We define the selection gradients in terms of the fitness proxy,

$$g_d^{(i)} = \left. \frac{\partial F}{\partial d_m} \right|_{d_m=d_i, q_m=q_i}, \quad g_q^{(i)} = \left. \frac{\partial F}{\partial q_m} \right|_{d_m=d_i, q_m=q_i} \quad (\text{A.4})$$

The elements of \mathbf{A}_m , and therefore also the fitness proxy F and the selection gradients $g_d^{(i)}, g_q^{(i)}$ depend on the trait values d_j, q_j and equilibrium densities $n_{1j}, n_{2,j}$ of all resident strategies ($j = 1, \dots, k$) through the variables I_G, I_L, P_S, P_C , but we suppress this in the notation.

The derivatives in (A.4) are easy to relate to the derivatives of the invasion fitness λ (the following derivation is simplified from Metz and Leimar (2011)). Recall that for a 2×2 matrix, the characteristic equation is

$$\lambda^2 - \text{tr} \lambda + \det = 0$$

where tr and \det are respectively the trace and the determinant of \mathbf{A}_m . Differentiating implicitly with respect to d_m and evaluating at $d_m = d_i, q_m = q_i$ (which implies $\lambda = 1$ at the equilibrium of population dynamics), we obtain

$$\left[(2 - \text{tr}) \frac{\partial \lambda}{\partial d_m} - \frac{\partial \text{tr}}{\partial d_m} + \frac{\partial \det}{\partial d_m} \right]_{d_m=d_i, q_m=q_i} = 0$$

i.e.,

$$\frac{\partial \lambda}{\partial d_m} \Big|_{d_m=d_i, q_m=q_i} = \frac{1}{2 - \text{tr}^{(i)}} \frac{\partial F}{\partial d_m} \Big|_{d_m=d_i, q_m=q_i} = \frac{1}{2 - \text{tr}^{(i)}} \cdot g_d^{(i)} \quad (\text{A.5a})$$

where $\text{tr}^{(i)}$ is the trace of \mathbf{A}_m evaluated at $d_m = d_i, q_m = q_i$ and population dynamic equilibrium. Analogously for the selection gradient on q ,

$$\frac{\partial \lambda}{\partial q_m} \Big|_{d_m=d_i, q_m=q_i} = \frac{1}{2 - \text{tr}^{(i)}} \cdot g_q^{(i)} \quad (\text{A.5b})$$

Since $\text{tr}^{(i)} < 2$ for every resident at a population dynamic equilibrium, the selection gradients $g_d^{(i)}$ and $g_q^{(i)}$ are sign-equivalent to the derivatives of the invasion fitness λ .

2 Selection gradients and singularities in monomorphic resident populations

In the monomorphic resident population of dispersal strategy (d, q) , the elements of the mutant's projection matrix are

$$\begin{aligned} a_{11} &= \frac{1 - d_m + d_m(1 - q_m)sb}{1 - d_m + d_m(1 - q_m)sb + s\gamma dq} + \frac{d_m q_m s \gamma p}{1 - d + d(1 - q)sb + s\gamma dq} \\ a_{12} &= \frac{d_m q_m s \gamma p}{1 - d + d(1 - q)sb + s\gamma dq} \\ a_{21} &= \frac{d_m q_m s \gamma (1 - p)}{1 - d + s\gamma dq + sd(1 - q)} \\ a_{22} &= \frac{1 - d_m}{1 - d_m + s\gamma dq + sd(1 - q)} + \frac{d_m(1 - q_m)s + d_m q_m s \gamma (1 - p)}{1 - d + s\gamma dq + sd(1 - q)} \end{aligned}$$

where we have substituted I_G, I_L, P_S, P_C as given in equations (A.2) and (A.3). The fitness proxy F and the selection gradients

$$g_d(d, q) = \frac{\partial F}{\partial d_m} \Big|_{d_m=d, q_m=q}, \quad g_q(d, q) = \frac{\partial F}{\partial q_m} \Big|_{d_m=d, q_m=q} \quad (\text{A.6})$$

are complicated expressions, but we can infer some properties of monomorphic evolution by investigating their properties near the boundaries of the trait space. We shall refer to lines in

the trait space where $g_d(d, q) = 0$ ($g_q(d, q) = 0$) as the d -isocline (q -isocline). A point (d^*, q^*) where the d - and q -isoclines intersect, i.e., where both selection gradients are zero, is an interior evolutionary singularity.

For the selection gradient on q on the boundaries, we obtain the following (algebraic manipulations performed with *Mathematica* 10, Wolfram Research):

(i) On the boundary $q = 1$,

$$g_q(d, 1) = d^2 s^2 \gamma \frac{[p - (1 - \gamma)](1 - d + s\gamma d) - ps\gamma bd}{(1 - d + s\gamma d)^3}$$

and therefore the selection gradient on q is negative for all d if $p < 1 - \gamma$ and positive for all d if $p > (1 - \gamma)/(1 - b)$. If $1 - \gamma < p < (1 - \gamma)/(1 - b)$, then $g_q(d, 1)$ is positive for $d < d_{crit}$ and negative for $d > d_{crit}$, where

$$d_{crit} = \frac{p - (1 - \gamma)}{(p - (1 - \gamma))(1 - s\gamma) + ps\gamma b} \quad (\text{A.7})$$

(ii) On the boundary $d = 1$, the selection gradient on q vanishes at

$$q_{crit} = \frac{b(p - (1 - \gamma))}{b(p - (1 - \gamma) + p\gamma) - \gamma(p - (1 - \gamma))} \quad (\text{A.8})$$

It is easy to check that $0 < d_{crit}, q_{crit} < 1$ if and only if

$$1 - \gamma < p < (1 - \gamma)/(1 - b) \quad (\text{A.9})$$

Care must be taken when $q \rightarrow 0$ or $d \rightarrow 0$, because in the limit, the monomorphic resident projection matrix (\mathbf{A}_m evaluated at $d_m = d, q_m = q$) becomes the identity matrix and therefore the factor $1/(2 - \text{tr})$ in (A.5) explodes to infinity, whereas g_d and g_q go to zero. The fitness proxy is not valid in the limit, and we must consider the derivatives in (A.5), i.e., $g_d/(2 - \text{tr})$ and $g_q/(2 - \text{tr})$.

(iii) As $q \rightarrow 0$, we have

$$\frac{g_q}{2 - \text{tr}} \rightarrow \frac{sd[p - (1 - \gamma)]}{1 - d + sd(1 - p(1 - b))} \quad (\text{A.10})$$

which is positive for $p > 1 - \gamma$ and negative for $p < 1 - \gamma$, independently of d .

(iv) As $d \rightarrow 0$, $g_q/(2 - \text{tr})$ goes to zero, whereas its derivative with respect to d goes to $s[p - (1 - \gamma)]$. Hence for small d , the selection gradient on q is positive (negative) if $p > (1 - \gamma)$ ($p < (1 - \gamma)$).

From the above analysis, we conclude that there is a q -isocline in the trait space that connects to the boundary at the points $(d_{crit}, 1)$ and $(1, q_{crit})$ when (A.9) holds, and there exist no other connections between an interior q -isocline and the boundary.

For the selection gradient on d , we have two limiting cases where our model becomes equivalent to the Hamilton-May (1977) model. With $q = 1$, each site is equally connected and therefore selection on d leads to the Hamilton-May ESS with survival probability $s\gamma$, i.e., to $d_{HM}(s\gamma) = 1/(2 - s\gamma)$. With $q \rightarrow 0$, solitary sites receive no immigrants and therefore selection on d ceases in the solitary sites; in the cluster, selection on d leads to $d_{HM}(s) = 1/(2 - s)$. The latter is shown formally by showing that as $q \rightarrow 0$,

$$\frac{g_d}{2 - \text{tr}} \rightarrow \frac{(1 - p)s[1 - (2 - s)d]}{(1 - d(1 - s))(1 - d(1 - s(1 - p(1 - b))))}$$

which is positive (negative) if d is smaller (greater) than $1/(2 - s)$. Therefore we obtain that

- (v) the points $(d_{HM}(s), 0)$ and $(d_{HM}(s\gamma), 1)$ are on the d -isocline, and the d -isocline does not connect to the boundaries $q = 0$ and $q = 1$ at any other point than these.

Further, we obtain the following properties using computer algebra:

- (vi) As $d \rightarrow 0$, we have

$$\frac{g_d}{2 - \text{tr}} \rightarrow s((1 - p)(1 - q) + \gamma q)$$

so that near the boundary $d = 0$, the selection gradient on d is always positive.

- (vii) The roots of $g_d(1, q) = 0$ are the roots of a cubic polynomial of q with one root at $q = 0$. Since the model reduces to the Hamilton-May model for both $q \rightarrow 0$ and $q = 1$ (see above), $g_d(1, q)$ is negative both at small and at high values of q whenever $s < 1$. Therefore $g_d(1, q)$ has either no root in $(0, 1]$ or it has two roots, \bar{q}_1 and \bar{q}_2 . If there is no root, then $g_d(1, q)$ is negative for all q ; if there are two roots, then $g_d(1, q)$ is positive for $\bar{q}_1 < q < \bar{q}_2$ and negative outside this interval.
- (viii) For $g_d(1, q_{crit})$ to be zero and hence $(1, q_{crit})$ to be singular, $s = [p(1 - b) - (1 - \gamma) + b]/b\gamma$ must hold. For p in the range $1 - \gamma < p < (1 - \gamma)/(1 - b)$ (where q_{crit} is between 0 and 1), the right hand side of this equation evaluates to a number between 1 and $1/\gamma \geq 1$. Hence $(1, q_{crit})$ cannot be singular except in the extreme case of $s = 1$ (no mortality during local dispersal).

Properties (v)-(vii) suggest that the d -isocline either connects the points $(d_{HM}(s), 0)$ and $(d_{HM}(s\gamma), 1)$, or a segment of the d -isocline connects $(d_{HM}(s), 0)$ to $(1, \bar{q}_1)$ and another segment connects $(1, \bar{q}_2)$ to $(d_{HM}(s\gamma), 1)$ (see Figure S1).

Suppose now that (A.9) holds and the q -isocline connects the points $(d_{crit}, 1)$ and $(1, q_{crit})$ (see above). If the d -isocline connects the points $(d_{HM}(s), 0)$ and $(d_{HM}(s\gamma), 1)$, then the q - and d -isoclines must intersect if $d_{crit} < d_{HM}(s\gamma) = 1/(2 - s\gamma)$ (panel (a) in Figure S1). This last inequality is equivalent to

$$sb > \frac{p - (1 - \gamma)}{p\gamma} \tag{A.11}$$

If, on the other hand, segments of the d -isocline connect $(d_{HM}(s), 0)$ to $(1, \bar{q}_1)$ and $(1, \bar{q}_2)$ to $(d_{HM}(s\gamma), 1)$, then the q -isocline must intersect one of these segments when (A.11) holds (panel

(b) in Figure S1). This is because for the intersection point to move off the segment and onto the boundary $d = 1$, it should cross through $(1, q_{crit})$ so that $(1, q_{crit})$ would become singular; but this is excluded by (viii) above.

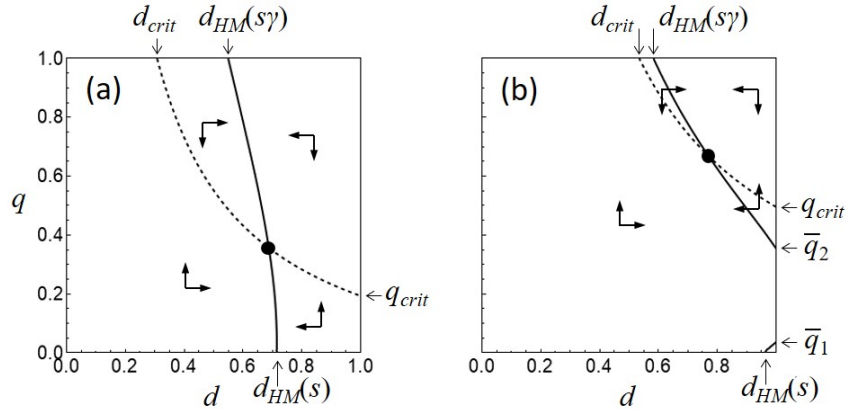


Figure S1: Selection gradients and evolutionary singularities in monomorphic populations. Arrows show the direction of the selection gradient on d (horizontal) and on q (vertical). Continuous line: d -isocline ($g_d = 0$); dashed line: q -isocline ($g_q = 0$); dot: singular strategy. (a) $p = 0.75$, $\gamma = 0.3$, $s = 0.6$, $b = 0.9$ (identical to Figure 2a of the main text); (b) $p = 0.9$, $\gamma = 0.3$, $s = 0.95$, $b = 0.9$.

The inequality in (A.11) is equivalent to $p < (1 - \gamma)/(1 - s\gamma b)$, which implies $p < (1 - \gamma)/(1 - b)$, the second inequality of (A.9). Therefore we need to keep only the first inequality of (A.9) together with (A.11), and we conclude that the isoclines intersect and the model has an interior singularity (d^*, q^*) if

$$p > 1 - \gamma \text{ and } sb > \frac{p - (1 - \gamma)}{p\gamma} \quad (\text{A.12})$$

as given in the main text.

The above derivation is based on the analysis of the selection gradients at the boundaries of the trait space. This analysis cannot exclude the existence of circular isoclines that do not connect to the boundary; hence interior singularities might exist also when the above conditions do not hold. Further, the analysis of the boundaries cannot inform about the uniqueness of the interior singularity, i.e., when the isoclines are shown to intersect, they might intersect more than once.

To exclude these possibilities, we have carried out an exhaustive numerical search for interior singularities using the numerical solver `vpsolve` of *MatLab*. First, we solved the equations defining a singularity, $g_d(d^*, q^*) = 0$ and $g_q(d^*, q^*) = 0$, in the interior of the trait space, $[\epsilon, 1 - \epsilon]^2$ with $\epsilon = 10^{-9}$, for all parameter combinations of $s, \gamma, p, b = 0.04, 0.09, \dots, 0.99$. From these 160 000 parameter combinations, we did not find any that does not fulfil the conditions in (A.12)

and yet yields an interior singularity; and we found an interior singularity for all 8808 parameter combinations that do fulfil (A.12).

Next, we checked the uniqueness of the interior singularities. Note that the second condition in (A.12) holds if $s > \bar{s} = [p - (1 - \gamma)]/p\gamma b$, and $\bar{s} < 1$ if $b > \bar{b} = [p - (1 - \gamma)]/p\gamma$. For every combination of $\gamma, p = 0.04, 0.09, \dots, 0.99$ that satisfies $p > 1 - \gamma$, we took 10 evenly spaced values for b between \bar{b} and 1, and for each value of b , 10 evenly spaced values for s between \bar{s} and 1. For every set of parameters thus obtained, we located an interior singularity (d^*, q^*) as above, and then used `vpsolve` to try to find another solution for $g_d(d, q) = 0$ and $g_q(d, q) = 0$ in the two domains $[10^{-9}, 1] \times [10^{-9}, q^* - 10^{-4}]$ and $[10^{-9}, 1] \times [q^* + 10^{-4}, 1]$ (i.e., excluding the first singularity found). In no case we found a second solution, which shows the interior singularity to be unique.

3 The canonical equation

We use the multivariate canonical equation of adaptive dynamics for structured populations (Durinx et al. 2008) to describe the evolution of dispersal strategies as long as the number of strategies does not change (directional evolution), i.e., before and after evolutionary branching. In the community of k resident strategies $(d_1, q_1), \dots, (d_k, q_k)$, the i th resident trait vector evolves according to

$$\frac{d}{dt} \begin{pmatrix} d_i \\ q_i \end{pmatrix} = \kappa_i \mathbf{C} \begin{pmatrix} g_d^{(i)} \\ g_q^{(i)} \end{pmatrix} \quad (\text{A.13})$$

where κ_i is a positive speed factor, \mathbf{C} is the mutational covariance matrix, and the last vector is the selection gradient vector. Below we describe these quantities in reverse order, applying the definitions of Durinx et al. (2008) to our specific model. For the derivation of (A.13), we refer to Durinx et al. (2008). Note that in this paper, we consider only monomorphic ($k = 1$) and dimorphic ($k = 2$) resident populations.

For the selection gradient vector, we need the derivatives of the leading eigenvalue λ of the mutant projection matrix \mathbf{A}_m . As derived in (A.5), this is given by $\frac{1}{2 - \text{tr}^{(i)}}$ times the selection gradients $g_d^{(i)}, g_q^{(i)}$ defined in terms of the fitness proxy F . In (A.13), we factor the coefficient $\frac{1}{2 - \text{tr}^{(i)}}$ into κ_i .

We assume that the mutational covariance matrix

$$\mathbf{C} = \begin{bmatrix} \sigma_d^2 & r\sigma_d\sigma_q \\ r\sigma_d\sigma_q & \sigma_q^2 \end{bmatrix}$$

is a positive definite matrix, i.e., the variances σ_d^2 and σ_q^2 of the (random) change one mutation causes respectively in d and in q are nonzero and the correlation r between the change in d and in q is less than 1 in absolute value. Since many morphological or behavioural changes that facilitate movement increase both d and q , it is likely that r is positive. For the numerical examples, we assume that \mathbf{C} is constant except when the dynamics have hit the boundary of the trait space; for the analysis of strong convergence stability, it suffices to assume that \mathbf{C}

depends on the trait values smoothly (Leimar 2009). Since q cannot be negative or exceed 1, we substitute σ_q with 0 for the resident that has either $q = 0$ with a selection gradient pointing towards lower q (i.e., $r\sigma_d\sigma_q g_d^{(i)} + \sigma_q^2 g_q^{(i)} < 0$) or $q = 1$ with a selection gradient pointing towards higher q . The value of d is similarly constrained to evolve between 0 and 1.

Finally, the speed factor

$$\kappa_i = \frac{1}{2 - \text{tr}^{(i)}} \cdot [\mu(n_{1i} + n_{2i})] \cdot \frac{1}{V_i}$$

is the product of three quantities, each dependent on the trait values of all resident strategies, but we suppress this in the notation. The first factor, $\frac{1}{2 - \text{tr}^{(i)}}$, corrects for the fact that we define the selection gradient with the fitness proxy rather than the invasion fitness (see above). The bracketed second factor is the probability at which a new mutation appears in the i th strategy per generation. We assume that each individual mutates with probability μ/M (where $M \rightarrow \infty$ is the total number of sites) and $\mu \rightarrow 0$ so that in most generations there is no mutant at all in the entire population. The total number of individuals with the i th strategy $(n_{1i} + n_{2i})M$ times the probability of mutation μ/M gives the probability that there is one mutant individual of the i th strategy. Note that for simplicity, we assume that an established adult (not a newborn) turns into a mutant.

The third factor is the inverse of the variance-like quantity

$$V_i = u_1^{(i)} \text{Var} \left(v_1^{(i)} \xi_{11}^{(i)} + v_2^{(i)} \xi_{21}^{(i)} \right) + u_2^{(i)} \text{Var} \left(v_1^{(i)} \xi_{12}^{(i)} + v_2^{(i)} \xi_{22}^{(i)} \right)$$

that plays a role through demographic stochasticity in the mutant's dynamics, present because the mutant initially exists in a single copy (see Durinx et al. 2008). Here $(v_1^{(i)}, v_2^{(i)})$ and $(u_1^{(i)}, u_2^{(i)})^T$ are respectively the left and right leading eigenvectors of the matrix \mathbf{A}_m , with elements given in (A.1), evaluated at the i th resident, $(d_m, q_m) = (d_i, q_i)$. The eigenvectors must be normalized such that $u_1^{(i)} + u_2^{(i)} = 1$ and $v_1^{(i)} u_1^{(i)} + v_2^{(i)} u_2^{(i)} = 1$. The quantities denoted with the letter ξ are random variables, and V_i is a weighted variance of these. $\xi_{11}^{(i)}$ denotes the random number of solitary sites won by the offspring born to one parent of the i th strategy who lived in a solitary site; and analogously, $\xi_{12}^{(i)}$ is the random number of solitary sites won by the offspring born in a clustered site, $\xi_{21}^{(i)}$ is the random number of clustered sites won by the offspring born in a solitary site, and $\xi_{22}^{(i)}$ is the random number of clustered sites won by the offspring born in a clustered site. The expected values of these random variables are the elements of the projection matrix \mathbf{A}_m evaluated at $(d_m, q_m) = (d_i, q_i)$. If the variances of offspring numbers around the expectations are small, then extinction due to demographic stochasticity is less likely for a mutant whose deterministic dynamics (leading eigenvalue of \mathbf{A}_m) predicts invasion. Small variances imply low V_i and therefore high κ_i ; evolution is faster when advantageous mutants are less likely lost to demographic stochasticity.

In the remainder of this section, we derive the distributions of the ξ 's and calculate V_i . We start with introducing the Bernoulli-distributed random variable

$$\zeta(\alpha) = \begin{cases} 1 & \text{with probability } \alpha \\ 0 & \text{with probability } 1 - \alpha \end{cases}$$

and the Poisson-distributed random variable $\eta(\beta)$,

$$\text{Prob}(\eta(\beta) = k) = \frac{\beta^k}{k!} e^{-\beta}$$

Consider first the random variable $\xi_{11}^{(i)}$. The nondispersing offspring born in a solitary site win either one solitary site (the natal site) or none; hence the number of solitary sites won by the nondispersing offspring is given by a Bernoulli-distributed random variable $\zeta(\alpha_1^{(i)})$. The expectation of this Bernoulli-variable is the probability that one of the nondispersing offspring wins the natal site,

$$\alpha_1^{(i)} = \frac{1 - d_i + d_i(1 - q_i)sb}{1 - d_i + d_i(1 - q_i)sb + I_G}$$

(cf. the first term of a_{11} in (A.1)). The large number of offspring who disperse from a solitary site have each an infinitesimal chance to win a site (since each competes against a large number of offspring), and therefore the number of solitary sites won by them is the Poisson distributed random variable $\eta(\beta_1^{(i)})$ with parameter

$$\beta_1^{(i)} = d_i q_i s \gamma p P_S$$

The total number of solitary sites won by offspring born in one solitary site is

$$\xi_{11}^{(i)} = \zeta(\alpha_1^{(i)}) + \eta(\beta_1^{(i)})$$

Notice that the expectation of $\xi_{11}^{(i)}$ is a_{11} as defined in (A.1) with $d_m = d_i$. Note also that $\zeta(\alpha_1^{(i)})$ and $\eta(\beta_1^{(i)})$ are independent; the fact that the nondispersed offspring won or did not win the natal site does not influence the fate of the dispersed offspring and *vice versa*.

For the other three random variables, by similar arguments we obtain

$$\begin{aligned} \xi_{12}^{(i)} &= \eta(\beta_1^{(i)}) \\ \xi_{21}^{(i)} &= \eta(\beta_2^{(i)}) \\ \xi_{22}^{(i)} &= \zeta(\alpha_2^{(i)}) + \eta(\beta_2^{(i)}) + \eta(\beta_3^{(i)}) \end{aligned}$$

with expectations

$$\alpha_2^{(i)} = \frac{1 - d_i}{1 - d_i + I_G + I_L}, \quad \beta_2^{(i)} = d_i q_i s \gamma (1 - p) P_C, \quad \beta_3^{(i)} = d_i (1 - q_i) s P_C$$

Recall that the variance of a Bernoulli random variable is $\text{Var}(\zeta(\alpha)) = \alpha(1 - \alpha)$ and the variance of the Poisson random variable $\eta(\beta)$ is its parameter β . Hence

$$\begin{aligned} \text{Var}(\xi_{11}^{(i)}) &= \alpha_1^{(i)} (1 - \alpha_1^{(i)}) + \beta_1^{(i)} \\ \text{Var}(\xi_{12}^{(i)}) &= \beta_1^{(i)} \\ \text{Var}(\xi_{21}^{(i)}) &= \beta_2^{(i)} \\ \text{Var}(\xi_{22}^{(i)}) &= \alpha_2^{(i)} (1 - \alpha_2^{(i)}) + \beta_2^{(i)} + \beta_3^{(i)} \end{aligned}$$

Since the ξ 's are independent of each other, we have

$$\begin{aligned} V_i &= u_1^{(i)} \text{Var} \left(v_1^{(i)} \xi_{11}^{(i)} + v_2^{(i)} \xi_{21}^{(i)} \right) + u_2^{(i)} \text{Var} \left(v_1^{(i)} \xi_{12}^{(i)} + v_2^{(i)} \xi_{22}^{(i)} \right) = \\ &= u_1^{(i)} \left(v_1^{(i)} \right)^2 \text{Var}(\xi_{11}^{(i)}) + u_1^{(i)} \left(v_2^{(i)} \right)^2 \text{Var}(\xi_{21}^{(i)}) + u_2^{(i)} \left(v_1^{(i)} \right)^2 \text{Var}(\xi_{12}^{(i)}) + u_2^{(i)} \left(v_2^{(i)} \right)^2 \text{Var}(\xi_{22}^{(i)}) \end{aligned}$$

Substituting the variances of the ξ 's and the eigenvector elements, we arrive at V_i expressed in terms of the model parameters and of the trait values. Note that V_i depends on the trait values of all resident strategies, not only on (d_i, q_i) , via the feedback variables I_G, I_L, P_S, P_L .

To obtain the evolutionary trajectories in Figure 4 of the main text, we numerically integrate the canonical equation in (A.13). For the initial point of coevolution in dimorphic populations when evolutionary branching happens at an interior singularity, we take two strategies near the singularity on either side in the direction of strongest disruptive selection (Figure 4a). For evolutionary branching near $q = 0$, we take two strategies with the same d -coordinate as where the monomorphic dynamics have hit the boundary $q = 0$, and with $q_1 = 0.001$, $q_2 = 0.005$ (Figure 4b).

4 Convergence stability, evolutionary stability, and evolutionary branching

4.1 Interior singularities

Singular strategies are fixed points of the canonical equation. A monomorphic interior singularity (d^*, q^*) is strongly convergence stable if it is a locally asymptotically stable fixed point of the monomorphic canonical equation (equation (A.13) with $k = 1$) for every positive definite covariance matrix \mathbf{C} . This is the case if the symmetric part of the Jacobian of the vector field of the selection gradient $(g_d, g_q)^T$ is negative definite (Leimar 2009), or, equivalently, if the (symmetric) matrix $\Gamma_{00} - \Gamma_{11}$ is negative definite, where

$$\Gamma_{00} = \begin{bmatrix} \frac{\partial^2 F}{\partial d_m^2} & \frac{\partial^2 F}{\partial d_m \partial q_m} \\ \frac{\partial^2 F}{\partial d_m \partial q_m} & \frac{\partial^2 F}{\partial q_m^2} \end{bmatrix}_{\substack{d_m=d=d^* \\ q_m=q=q^*}}, \quad \Gamma_{11} = \begin{bmatrix} \frac{\partial^2 F}{\partial d^2} & \frac{\partial^2 F}{\partial d \partial q} \\ \frac{\partial^2 F}{\partial d \partial q} & \frac{\partial^2 F}{\partial q^2} \end{bmatrix}_{\substack{d_m=d=d^* \\ q_m=q=q^*}}$$

(Geritz et al. 2016). A monomorphic interior singularity is evolutionarily stable if the Hessian matrix Γ_{00} is negative definite. Since we have only two traits evolving, any strongly convergence stable interior monomorphic singularity that is not evolutionarily stable is an evolutionary branching point *sensu* Geritz et al. (2016). Note that Geritz et al. (2016) assume that evolution is not mutation-limited, and therefore the speed of evolution is independent of population size, whereas the canonical equation, which we use to demonstrate evolutionary branching in Figure 4 of the main text, assumes mutation limitation. The latter approach is more justified when the traits are under directional selection, which depletes the standing genetic variation; this happens after the initial phase of evolutionary branching and during evolutionary branching at the boundary $q = 0$ (see below). Genetic variability accumulates more readily when directional selection ceases, i.e., near interior evolutionary singularities.

To obtain the bifurcation diagrams of the monomorphic evolutionary singularities, we determined the position of the interior singularity (d^*, q^*) by solving $g_d(d, q) = 0$, $g_q(d, q) = 0$ numerically, and evaluated Γ_{00} and Γ_{11} accordingly. Figure S2 shows the results for the entire range of p and γ , whereas Figure 3 of the main text shows a richer part at a higher resolution of p and γ . For each combination of p and γ shown in these figures, we varied s and b between 0.02 and 1 with stepsize 0.02 to obtain the individual panels.

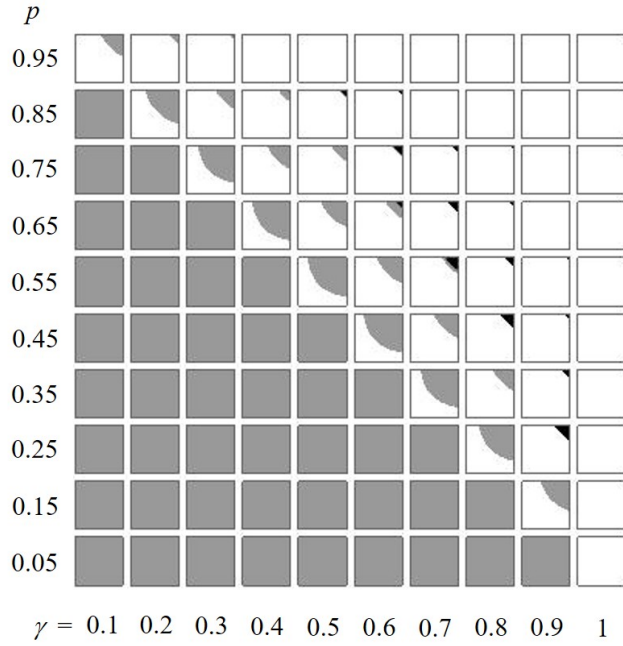


Figure S2: Bifurcation diagrams of the monomorphic evolutionary singularity. In each small panel, s runs from 0 to 1 on the horizontal axis and b from 0 to 1 on the vertical axis. p and γ as shown next to the panels. Grey: disruptive selection / evolutionary branching (including branching near $q = 0$); white: boundary ESS; black: interior ESS. Figure 3 of the main text shows a part of this figure at higher resolution of p and γ .

Each interior singularity found on the grid used for the figures is strongly convergence stable. However, for a narrow range of p between $1 - \gamma$ and $1 - \gamma + \delta p$, the interior singularity is not strongly convergence stable for high s and intermediate b (data not shown). This narrow range falls between the panels of Figures S2 and 3; for $\gamma = 0.4$, δp is less than 0.02, whereas for both $\gamma = 0.1$ and $\gamma = 0.8$, it is less than 0.01. In this range, the singularity as a fixed point of the canonical equation can undergo a Hopf bifurcation when the parameters in the covariance matrix \mathbf{C} are varied, resulting in sustained cycles of the trait values (Red Queen evolution; note that this may need negative correlation between the traits, which is biologically unlikely in our model). Because of the parameter range that yields these not strongly convergence stable interior singularities is narrow, we do not investigate this case further.

4.2 Boundary singularities

As explained in section 2 of this Electronic Appendix, no interior singularity exists if the inequalities in (A.12) do not hold. If $p < 1 - \gamma$, i.e., the first inequality in (A.12) is violated, then the boundary $q = 0$ attracts every orbit of the canonical equation (cf. section 2; the lack of a fixed point excludes a periodic orbit by the Poincaré-Bendixson theorem). We consider this case separately below, and conclude that q is under disruptive selection near the attracting boundary $q = 0$. This corresponds to the all-grey panels in Figure S2 and in Figure 3 of the main text.

The first inequality in (A.12) holds but the second, which is the same as (A.11), is violated in the white areas of Figure S2 and of Figure 3 of the main text. In this case, the boundary singularity $(d_{HM}(s\gamma), 1)$ is an attracting ESS. At the boundary singularity $(d_{HM}(s\gamma), 1)$, the selection gradient $g_q(d_{HM}(s\gamma), 1)$ generically does not vanish (cf. section 2), and evolution near this singularity is dominated by the nonvanishing selection gradient. If (A.11) holds, then $g_q(d_{HM}(s\gamma), 1)$ is negative so that the boundary $q = 1$ is locally repelling. If the opposite of (A.11) holds so that the boundary $q = 1$ is locally attracting, then every site becomes equally connected and evolution on the boundary will lead to the Hamilton-May ESS, $d_{HM}(s\gamma)$. In this case, $(d_{HM}(s\gamma), 1)$ is both strongly convergence stable and locally evolutionarily stable.

Note, however, that $(d_{HM}(s\gamma), 1)$ is not a global ESS, because its monomorphic population can be invaded by $(d_m, q_m) = (d_{HM}(s\gamma), 0)$, a mutant that disperses only locally; and these two strategies coexist by mutual invasibility. To see this, we first evaluate the projection matrix of the mutant $(d, 0)$ in the resident population of $(d, 1)$ to arrive at

$$\mathbf{A}_m = \begin{bmatrix} \frac{1-d+sb d}{1-d+sb d+s\gamma d} & 0 \\ 0 & \frac{1-d+sd}{1-d+s\gamma d} \end{bmatrix}$$

The leading eigenvalue of this matrix is its a_{22} element, which exceeds 1 whenever $\gamma < 1$, and hence the mutant strategy $(d, 0)$ invades the resident $(d, 1)$ for any d (including $d = d_{HM}(s\gamma)$). The reverse case, i.e., that the mutant $(d, 1)$ invades the resident $(d, 0)$, is intuitively obvious because the resident cannot disperse to the solitary sites. For the formal proof, now the mutant projection matrix simplifies to

$$\mathbf{A}_m = \begin{bmatrix} 1 + \frac{ds\gamma p}{1-d+dsb} & \frac{ds\gamma p}{1-d+dsb} \\ \frac{ds\gamma(1-p)}{1-d+sd} & \frac{1-d+ds\gamma(1-p)}{1-d+sd} \end{bmatrix}$$

which gives

$$F = \text{tr } \mathbf{A}_m - \det \mathbf{A}_m = 1 + \frac{ds\gamma p}{1-d+dsb} \cdot \frac{sd}{1-d+sd} > 1$$

i.e., the mutant invades.

For $p = 1 - \gamma$, (A.11) is violated only for $sb = 0$, i.e., on the horizontal and vertical axes in the small panels of Figure S2. With increasing p or γ , an area with a boundary ESS (white) opens up from the axes of the small panels and the area of evolutionary branching (grey) shrinks. At the boundary of the white areas in Figure S2, where (A.11) holds with equality, an interior

singularity appears through a transcritical bifurcation of the boundary singularity $(d_{HM}(s\gamma), 1)$. At this bifurcation point, the Hessian matrix simplifies to

$$\Gamma_{00} = 2s^2 \begin{bmatrix} -\gamma^2(2-s\gamma)^2 & -\gamma(p-(1-\gamma))(2-s\gamma) \\ -\gamma(p-(1-\gamma))(2-s\gamma) & (1-\gamma^2)-p \end{bmatrix}$$

The determinant of this Hessian is positive when

$$\gamma^2(2-s\gamma)^2 [p-(1-\gamma^2)-(p-(1-\gamma))^2] > 0$$

which is equivalent to $p > 2(1-\gamma)$. We thus conclude that the Hessian is negative definite when $p > 2(1-\gamma)$ and indefinite otherwise. Therefore the interior singularity, when it appears through the transcritical bifurcation, is an evolutionary branching point if

$$1-\gamma < p < 2(1-\gamma) \tag{A.14}$$

holds (recall that for $p < 1-\gamma$ there is no interior singularity). The interior singularity can turn into an ESS for higher values of s and b as seen in panels of Figure 3 in the main text, where a grey band of evolutionary branching points separates white areas with no interior singularity and black areas with interior ESSs. The interior singularity appears as an ESS, and our numerical results in Figure S2 and Figure 3 of the main text show that evolutionary branching does not occur if $p > 2(1-\gamma)$. A similar derivation shows that near the transcritical bifurcation of the interior singularity, it is always strongly convergence stable.

4.3 Evolutionary branching near $q = 0$

For $p < 1-\gamma$, the boundary $q = 0$ is attracting at every d . The orbits of the canonical equation hit the boundary at a point that depends on the initial condition (see Figure 4b in the main text) and at a nonvanishing speed (cf. (A.10) in section 2). Recall that at $q = 0$, the selection gradient based on the fitness proxy F is invalid (section 2). Moreover, the limit $q \rightarrow 0$ does not commute with the derivatives taken for the selection gradients; and at $q = 0$, the eigenvalues cross so that the leading eigenvalue λ is not differentiable with respect to d_m at $d_m = d$ and with respect to q_m at $q_m = 0$ (see Figure S3b below). Since, by assumption, strategies with $q = 0$ are not admitted, henceforth we assume $q > 0$ even though we let q evolve arbitrarily close to 0.

To explore evolution near the boundary $q = 0$, we first show that two strategies with small but nonzero q can coexist. Since F is continuous in q , the limit

$$\lim_{q \rightarrow 0} F(d, q_m, d, q) = 1 + \frac{q_m^2 d^2 p s^2 \gamma}{(1-d(1-s))(1-d(1-sb))} > 1$$

shows that also for small but nonzero q , there are mutants $q_m = q + \delta q$ with small positive δq that can invade; more precisely, for every $\delta q > 0$ there exists a $q > 0$ such that the mutant $(d_m, q_m) = (d, q + \delta q)$ can invade the resident (d, q) . Hence the pairwise invasibility plot of q (with $d_m = d$ fixed) has a “+” area above the diagonal $q_m = q$, which comes arbitrarily close to the diagonal near $q = 0$ (Figure S3a).

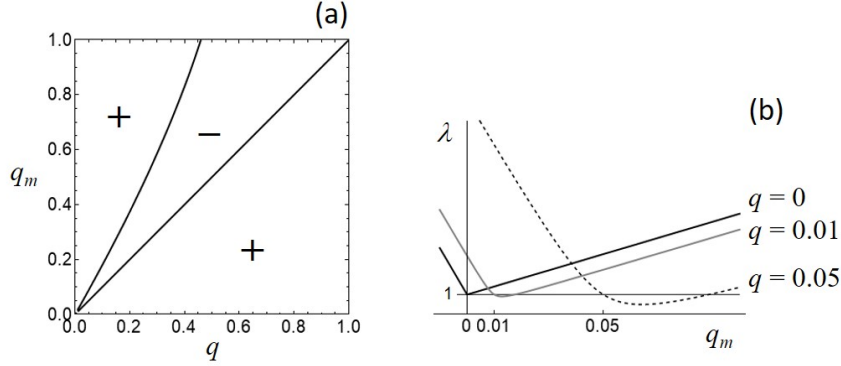


Figure S3: (a) Pairwise invasibility plot (PIP) for q when the boundary $q = 0$ is attracting. “+” areas: $F > 1$, the mutant can invade; “-” area: $F < 1$, the mutant dies out. (b) The leading eigenvalue of the mutant projection matrix as a function of q_m for $q = 0$ (black), $q = 0.01$ (grey) and $q = 0.05$ (dashed). Negative values of q_m are included for clarity. Parameters as in Figure 2b of the main text ($p = 0.4$, $\gamma = 0.3$, $s = 0.6$, $b = 0.9$) and $d_m = d = 0.4$.

Mutants with $d_m = d$, $q_m < q$ can always invade when the resident has a small positive q . To see this, first recall that since we have $p < 1 - \gamma$, (A.10) shows that g_q is negative for small q ; i.e., immediately below the diagonal there is a “+” area in the pairwise invasibility plot (see Figure S3a). On the other hand, the limit

$$\lim_{q_m \rightarrow 0} F(d, q_m, d, q) = 1 + \frac{q^2 d^2 s^2 \gamma (1 - \gamma)}{(1 - d + ds(b + q\gamma))(1 - d + ds(1 - q + q\gamma))} > 1$$

shows that there is a “+” area also near the horizontal axis of the pairwise invasibility plot. Therefore to have a “-” area below the diagonal, $F - 1$, as a function of q_m , should have two zeros below q . However, it is readily shown that the roots of $F - 1$ are the roots of a cubic polynomial. Since $q_m = q$ is one root and the boundary of the “+” area above the diagonal is a second root, there cannot be two more roots below the diagonal. It follows that the configuration of the pairwise invasibility plot near its origin is always as shown in Figure S3a. Two strategies (d, q) and (d, q_m) with $d, q, q_m > 0$ and (q, q_m) in the “+” area above the diagonal can coexist (Geritz et al. 1998), and this coexistence set can be reached by a small mutation δq when the resident q is near zero. By continuity, coexistence is also possible if the two strategies have somewhat different d , i.e., there is an open set of mutants that can form a dimorphism with the former resident. Once the population is in the neighbourhood of the boundary $q = 0$, it will stay there as long as it is monomorphic, and it will almost surely become dimorphic.

Consider now a dimorphic population where both strategies still have small q . As long as the two strategies are similar to each other, the invasion fitness proxy, as a function of the mutant strategy, is qualitatively similar to what it was in a monomorphic population. It follows from the configuration of the pairwise invasibility plot (Figure S3a) near $q = 0$ that $F(d, q_m, d, q)$, as a function of q_m , has a minimum; with increasing q_m , it is first greater than 1 (“+” area below the diagonal), then less than 1 (“-” area), then greater than 1 again (“+” area above the diagonal). Numerical examples are shown in Figure S3b. A minimum of the invasion fitness proxy

implies that there is disruptive selection on q ; the two resident strategies will evolve increasingly different values of q as long as d is fixed, i.e., q will undergo evolutionary branching (Geritz et al. 1998). Since there are no mutants with nonpositive q , the branch with the lower value of q remains near $q = 0$, but the other branch quickly evolves higher values of q , resulting in asymmetric divergence (see Figure 4b in the main text). Note that the model of Kisdi (2016) is a special case of our model with only q evolving, $d = 1$ fixed, and $s = 1$, $b = 1$. Accordingly, Kisdi (2016) found the same nondifferentiability and asymmetric branching at $q = 0$, but without the confounding effect of evolving d .

Concomitantly with evolutionary branching in q , d is under directional selection unless branching happens near $(d_{HM}(s), 0)$. Directional selection on d would prevent branching if the selection gradient on q were vanishing (cf. evolutionary branching generically happens at a singularity, not at any point of an isocline). At $q = 0$, however, the invasion fitness λ is nondifferentiable, so that there is a very sharp turn at its minimum also for small positive q (Figure S3b). This results in strong selection on q (comparable to that on d) already at a small separation between the branches, which makes branching possible also away from $(d_{HM}(s), 0)$ (Figure 4b of the main text).

With multiple traits evolving jointly, there is no guarantee that the two branches will remain in the coexistence set even for the initial stage of branching (Geritz et al. 2016; although for 2-dimensional trait spaces coexistence during initial separation is ensured at regular branching points such as the interior branching points of our model, this does not apply to branching near $q = 0$). Indeed, our numerical experiments showed examples where, after initial separation in q , one branch lagged behind with evolving its dispersal propensity d and went extinct (one such example, with initial values $d_1 = d_2 = 0.38$, $q_1 = 0.001$ and $q_2 = 0.005$ and with parameters $p = 0.4$, $\gamma = 0.3$, $s = 0.6$, $b = 0.9$ and \mathbf{C} the identity matrix, is included in the *Mathematica* notebook of the canonical equation available at <https://doi.org/10.5061/dryad.9ghx3ffd0>). Should one of the branches go extinct, the remaining monomorphic population would evolve back to the boundary $q = 0$, and would attempt branching again; the population thus may alternate between monomorphic and dimorphic states. If d evolves to $d_{HM}(s)$ in the meanwhile, then directional selection on d vanishes and branching in the q -direction succeeds.

It is however difficult to ascertain that repeated bouts of failed evolutionary branching would bring the population closer to $(d_{HM}(s), 0)$, for two reasons. First, if d and q are correlated, then the branch that evolves higher q during asymmetric branching may also evolve d away from $d_{HM}(s)$ as a correlated response (with a positive correlation between d and q , this happens if the population reaches near $q = 0$ with $d > d_{HM}(s)$). If subsequently the other branch goes extinct, the remaining monomorphic population has d further away from $d_{HM}(s)$ than it was before branching; but it reverts towards $d_{HM}(s)$ as the monomorphic population evolves back to the vicinity of $q = 0$.

Second, during the time the population is dimorphic, the selection gradients on d are complicated by spatial sorting of the two residents between the solitary and clustered sites. Below we argue that, by continuity, the selection gradient on d in dimorphic populations is similar to

that of in monomorphic populations and therefore points towards $d_{HM}(s)$, as long as both q_1 and q_2 are near zero and d_1 and d_2 are close to each other but not yet close to $d_{HM}(s)$. (Note that two strategies with significantly different d cannot coexist unless they are separated also in q , i.e., unless evolutionary branching has succeeded; this follows from the fact that there is no coexistence in the Hamilton-May model, see Appendix E of Kisdi 2016.)

Suppose q_m, q_1 and q_2 are of order $\epsilon \ll 1$. The mutant projection matrix is then of the form

$$\mathbf{A}_m = \begin{bmatrix} 1 + O(\epsilon) & O(\epsilon) \\ O(\epsilon) & \phi(d_m, d_1, d_2) + O(\epsilon) \end{bmatrix}$$

with

$$\phi(d_m, d_1, d_2) = \frac{1 - d_m}{1 - d_m + I_L} + d_m s P_C, \quad I_L = s \sum_{i=1}^2 \frac{n_{2i}}{1 - p} d_i, \quad P_C = \sum_{i=1}^2 \frac{n_{2i}}{1 - p} \cdot \frac{1}{1 - d_i + I_L}$$

(cf. equations (A.1)-(A.3)). With $\epsilon = 0$, the leading eigenvalue of \mathbf{A}_m is $\max(1, \phi(d_m, d_1, d_2))$, and the two eigenvalues coincide when $\phi = 1$. With $\epsilon > 0$, \mathbf{A}_m is primitive and therefore, by the Perron-Frobenius theorem, its eigenvalues never coincide; further, the leading eigenvalue depends smoothly on d_m . The leading eigenvalue is therefore $\phi(d_m, d_1, d_2) + O(\epsilon)$. If d_1 and d_2 are close to each other such that both can be written as $d + O(\epsilon)$, then $I_L = sd + O(\epsilon)$ and $P_C = \frac{1}{1 - d + sd} + O(\epsilon)$, so that for the leading eigenvalue

$$\frac{1 - d_m}{1 - d_m + sd} + \frac{d_m s}{1 - d + sd} + O(\epsilon)$$

we recover the invasion fitness of the Hamilton-May (1977) model when $\epsilon \rightarrow 0$. It follows that mutants with dispersal propensity closer to $d_{HM}(s)$ are at an advantage.

4.4 Direction of evolutionary branching

The direction of strongest disruptive selection at interior branching points is given by the eigenvector of the Hessian matrix Γ_{00} that belongs to the greatest positive eigenvalue. Note that Γ_{00} is the Hessian of the invasion fitness proxy F , not directly of the invasion fitness given by the leading eigenvalue λ of the mutant projection matrix. A straightforward derivation analogous to the one in section 1, presented in a generalized form by Metz and Leimar (2011), shows that the two Hessians differ in only the positive factor $1/(2 - \text{tr}^*)$, where tr^* is the trace of the projection matrix \mathbf{A}_m evaluated at $d_m = d = d^*$, $q_m = q = q^*$ and it is always less than 2 as mentioned in section 1. Hence the eigenvectors of the two Hessians coincide, and we can use the Hessian of the invasion fitness proxy F .

To sample the parameter space, we varied p , γ , s and b between 0.05 and 0.95 with increments of 0.05, and took all combinations of these values where p and γ obey (A.14) (otherwise no interior branching point exists). These combinations of parameter values yield 4852 interior evolutionary branching points, each with one positive and one negative eigenvalue of the Hessian. Figure S4 shows that the eigenvectors corresponding to the positive eigenvalues are

typically just a few degrees from the q -direction (i.e., vertical in Figure S1 and Figures 2 and 4 of the main text). Hence disruptive selection acts predominantly on q , the trait which quantifies the pattern of connectivity in our model.

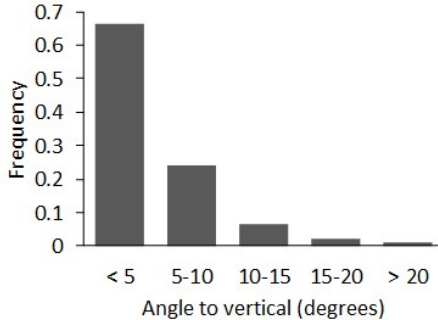


Figure S4: The distribution of the angle between the q -direction (vertical) and the direction of the leading eigenvector of the Hessian matrix Γ_{00} over a systematic sample of 4852 evolutionary branching points.

When evolutionary branching occurs near the boundary $q = 0$, then generically the two branches evolve d in the same direction while they separate in the q -direction (Figure 4b in the main text). Thus branching does not happen along a straight line in trait space, i.e., it does not have a “direction”. Divergence is however exclusively due to disruptive selection in the q -direction.

The direction where the nascent branches diverge from each other depends also on the co-variation between the evolving traits d and q (cf. (A.13)). Although disruptive selection acts nearly or entirely in the q -direction, evolutionary branching can happen much closer to a 45° line in the trait space when the two traits have strong positive correlation (grey trajectory in Figure 4a of the main text).

5 Dimorphic singularities

In order to investigate possible final outcomes of evolution once the population has become dimorphic, we first search for dimorphic singularities of the form $((d_1^*, 0), (d_2^*, 1))$, i.e., boundary singularities such that one resident strategy never disperses globally ($q_1 = 0$) whereas the other disperses only globally ($q_2 = 1$). Note that in dimorphic populations, $\text{tr}^{(1)} < 2$ even if $q_1 = 0$, and therefore the invasion fitness proxy can be used. Let $F(d_m, q_m; d_1, q_1, d_2, q_2)$ denote the fitness proxy of a mutant (d_m, q_m) in the dimorphic resident population of (d_1, q_1) and (d_2, q_2) , and let $g_d^{(i)}$ and $g_q^{(i)}$ be respectively the selection gradients on d and on q of the i th resident as given in (A.4). For a given set of parameters (p, γ, s, b) , we determine the singular trait values

by numerically solving the equations

$$F(d_2^*, 1, d_1^*, 0, d_2^*, 1) = 1, \quad g_d^{(1)} = 0, \quad g_d^{(2)} = 0$$

for the unknowns n_{21} , d_1^* and d_2^* . The first of these equations is the condition for population dynamical equilibrium. Even though suppressed in the notation, the mutant projection matrix and therefore F depend on the equilibrium densities n_{11} and n_{21} (see equations (A.1)-(A.3); in monomorphic populations the equilibrium densities are trivial, a fraction p of the resident occupies solitary sites and the remainder is in the clustered sites). With $q_1 = 0$, the first resident is fully outcompeted from the solitary sites and thus $n_{11} = 0$; recall also that $n_{12} = p - n_{11}$ and $n_{22} = 1 - p - n_{21}$, which leaves only n_{21} as unknown. The second and third equations are the conditions for a singular dimorphism, where the selection gradients $g_d^{(i)}$ depend on the resident trait values as well as on the equilibrium densities.

Once we have located a boundary singularity $((d_1^*, 0), (d_2^*, 1))$, we verify that the boundaries are attracting, i.e., check whether $g_q^{(1)} < 0$ and $g_q^{(2)} > 0$ hold at the singularity (recall that at a boundary singularity, the selection gradients perpendicular to the boundaries are nonvanishing). If this is so, then the singularity is evolutionarily stable if $[\partial^2 F / \partial d_m^2]$ is negative at both residents, $d_m = d_1^*, q_m = 0$ and $d_m = d_2^*, q_m = 1$. Since the nonvanishing selection gradients push the dynamics onto the boundary $q_1 = 0, q_2 = 1$, convergence stability can be evaluated from the two-dimensional dynamics

$$\begin{aligned} \frac{d}{dt} d_1 &= \kappa_1 \sigma_d^2 g_d^{(1)} \\ \frac{d}{dt} d_2 &= \kappa_2 \sigma_d^2 g_d^{(2)} \end{aligned}$$

(cf. (A.13)). The fixed point $((d_1^*, 0), (d_2^*, 1))$ of this system is asymptotically stable for any positive speed factors κ_1, κ_2 and positive variance σ_d^2 if and only if $[\partial g_d^{(i)} / \partial d_i] < 0$ for $i = 1, 2$ and $[\partial g_d^{(1)} / \partial d_1][\partial g_d^{(2)} / \partial d_2] > [\partial g_d^{(1)} / \partial d_2][\partial g_d^{(2)} / \partial d_1]$ at the singularity (note that the partial derivative is taken such that the equilibrium density is substituted into the selection gradients as a function of resident trait values before differentiation).

We found that a boundary dimorphic singularity $((d_1^*, 0), (d_2^*, 1))$ exists for most parameter combinations (white areas in Figure S5). All singularities of this form that we have located were both strongly convergence stable and evolutionarily stable. For high s, b, γ and low p , however, the selection gradient $g_q^{(2)}$ switches to be negative, whereby a singularity $((d_1^*, 0), (d_2^*, q_2^*))$ (with $q_2^* < 1$) appears through a transcritical bifurcation. To check the evolutionary stability of this singularity, we need to replace the condition $[\partial^2 F / \partial d_m^2 < 0]$ with the 2×2 Hessian matrix of the second resident being negative definite. For strong convergence stability, we verified that the symmetric part of the Jacobian of the three-dimensional vector field $g_d^{(1)}(d_1, d_2, q_2)$, $g_d^{(2)}(d_1, d_2, q_2)$, $g_q^{(2)}(d_1, d_2, q_2)$ is negative definite (see section 4). This condition is sufficient though not necessary, because it allows for arbitrary correlations between mutational changes in the trait values, but mutations affecting d_1 are in fact independent from those affecting (d_2, q_2) . All singularities we found of the form $((d_1^*, 0), (d_2^*, q_2^*))$ were evolutionarily stable and satisfied the sufficient condition for strong convergence stability (black areas in Figure S5).

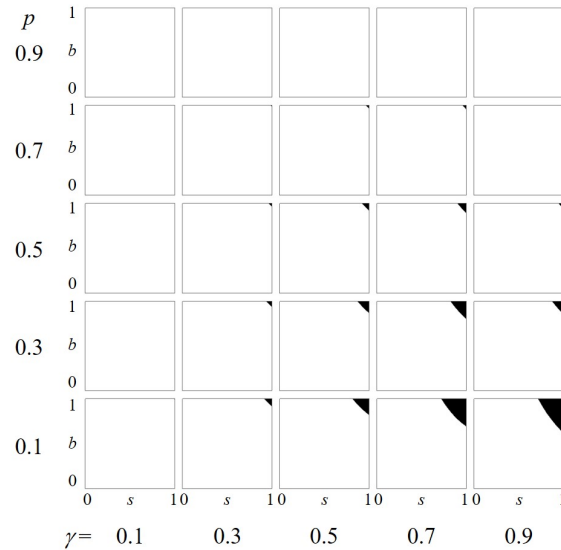


Figure S5: Dimorphic evolutionary singularities. There is a strongly convergence stable as well as evolutionarily stable dimorphic singularity of the type $((d_1^*, 0), (d_2^*, 1))$ in the white areas and of the type $((d_1^*, 0), (d_2^*, q_2^*))$ with $q_2^* < 1$ in the black areas. The results are based on evaluating the singularities on a grid of s and b varying between 0.04 and 0.99 by steps of 0.05, and drawing the boundary between the white and black areas as the zero contour line of $g_q^{(2)}$ using the ListContourPlot procedure of *Mathematica*.

The value of parameter b , which determines the probability of returning to the natal site for a disperser from a solitary site who fails to engage in global dispersal, is irrelevant for strategies with $q = 0$ (because they are never present in a solitary site at equilibrium) and also for strategies with $q = 1$ (because they never fail to disperse globally). Accordingly, for the boundary singularities $((d_1^*, 0), (d_2^*, 1))$, the ESS dispersal propensities d_1^* and d_2^* are independent of b . For the singularities $((d_1^*, 0), (d_2^*, q_2^*))$, also p has very little effect on d_1^* and d_2^* , so that the dispersal propensities of these boundary ESSs are mostly determined by s and γ . Figure S6 shows the effect of the latter two parameters for $p = 0.5$, the results are similar for other values of p as well. As expected, both d_1^* and d_2^* increases with s , and d_2^* , the dispersal propensity of the resident which occupies the solitary sites, also significantly increases with γ . Both d_1^* and d_2^* exceed 0.5 (recall that also in the Hamilton-May model, dispersal always exceeds 0.5), and $d_1^* > d_2^*$, i.e., the resident strategy restricted to the clustered sites has a higher probability of dispersal than the resident with only global dispersal. As seen in Figure S6, the difference between d_1^* and d_2^* is large when there is a large contrast between the costs of local and global dispersal, i.e., when s is high and γ is low.

At singularities of the type $((d_1^*, 0), (d_2^*, q_2^*))$, the first resident is still restricted to the clustered sites (due to $q_1^* = 0$). According to our numerical results (see the data file at <https://doi.org/10.5061/dryad.9ghx3ffd0>), the second resident has a high probability of dispersing globally ($q_2^* > 0.5$ at every dimorphic singularity we have located). Hence the two residents

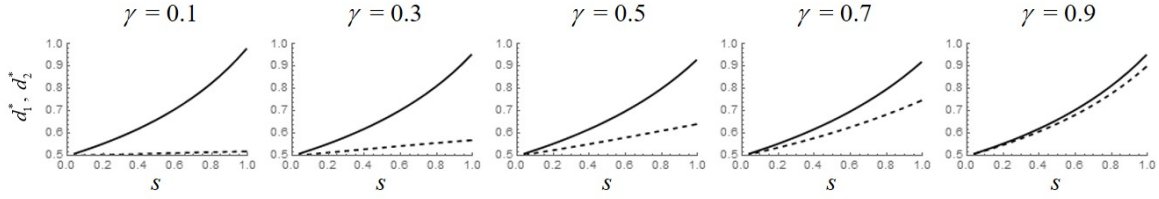


Figure S6: Dispersal propensities in evolutionarily stable dimorphisms $((d_1^*, 0), (d_2^*, 1))$. Black lines: d_1^* , dashed lines: d_2^* . $p = 0.5$. The value of b is irrelevant (see text), but the results are valid only for evolutionarily stable dimorphisms of the form $((d_1^*, 0), (d_2^*, 1))$, i.e., in the white areas of Figure S5.

are always strongly differentiated in their pattern of connectivity, measured by q . q_2^* is the lowest (i.e., closest to 0.5) when b and s are high (this is consistent with the fact that a dimorphic singularity with $q_2^* < 1$ appears via a transcritical bifurcation through $q_2^* = 1$ as b or s increases, cf. Figure S5) and when p is low. $d_1^* > d_2^* > 0.5$ holds also for these singularities. d_1^* and d_2^* increase with b (the possibility of returning to a solitary site after failing to disperse globally affects the second resident directly, but also the first resident responds to changing b because it coevolves with the second). d_1^* and d_2^* also increase with s , but decrease with p . Unexpectedly, changing γ may have a non-monotonic effect on a singularity of the type $((d_1^*, 0), (d_2^*, q_2^*))$.