Asymmetric patch size distribution leads to disruptive selection on dispersal

François Massol, Anne Duputié, Patrice David, Philippe Jarne

Appendix S1: Complete mathematical analysis of the model

Model basics

We consider a metapopulation consisting of an infinite number of patches, with a specified distribution of carrying capacities. The patch carrying capacity K represents a number of sites within the patch, each of which can be occupied by a single individual. The model obeys the following rules: (i) all individuals have the same mortality rate (m); (ii) a dead individual is immediately replaced by either a resident or immigrant offspring; (iii) all individuals have the same fecundity; (iv) individuals of strategy d send a proportion d of their offspring to the propagule pool ($1-d$ remain in their natal patch); (v) a proportion c of propagules dies before reaching a randomly chosen destination patch.

Formalization of the two-type model

Computing the fitness of a rare mutant type relies on the probability that the local abundance of a dispersal type increases when a dead individual is replaced. This probability, noted ν_i^K, depends on the initial abundance (i) of type d and the carrying capacity of the habitat patch (K). Type d increases in abundance when (i) an individual of another type dies (probability $(K-i)/K$) and (ii) the vacant site is colonized by type d offspring. We assume that all offspring present in the patch have an equal probability of colonizing, *i.e.* each type colonizes with a probability equal to its local frequency. Type d will, on the other hand, decrease in abundance when a focal type individual dies and is replaced by another type (probability μ_i^K).

To tackle the problem of dispersal evolution, we consider two types (1 and 2) competing for microsites in an ideal metapopulation consisting of an infinity of patches. The proportion of dispersed offspring in type s is noted d_s ($0 \le d_s \le 1$). In this context, ν_i^K and μ_i^K can be rewritten as follows, to describe the probabilities of increase and decrease of type 1 abundance in a patch that has carrying capacity K:

$$\nu_i^K = \frac{K-i}{K}\frac{(1-d_1)i+(1-c)d_1\bar{i}}{(1-d_1)i+(1-c)d_1\bar{i}+(1-d_2)(K-1-i)+(1-c)d_2(\overline{K}-\bar{i})} \quad (1)$$

$$\mu_i^K = \frac{i}{K}\frac{(1-d_2)(K-i)+(1-c)d_2(\overline{K}-\bar{i})}{(1-d_1)(i-1)+(1-c)d_1\bar{i}+(1-d_2)(K-i)+(1-c)d_2(\overline{K}-\bar{i})} \quad (2)$$

where ν_i^K (resp. μ_i^K) is the probability that the next replacement (death-recruitment event) increases (resp. decreases) the number of type 1 individuals in the patch, c is the dispersal cost, $\bar{i} = \mathrm{E}[i]$ is the average number of type 1 individuals per patch, and $\overline{K} = \mathrm{E}[K]$ is the average number of microsites per patch. The master equation that determines the dynamics of the probability p_k^K that a patch with carrying capacity K contains exactly k individuals of type 1 is as follows:

$$\frac{dp_k^K}{dt} = mK\left(\nu_{k-1}^K p_{k-1}^K + \mu_{k+1}^K p_{k+1}^K - (\nu_k^K + \mu_k^K)\, p_k^K\right) \quad (3)$$

where m is the *per capita* mortality rate. The K factor (right-hand side of equation [3]) reflects the fact that deaths occur more often in large patches (*i.e.* the mortality rate is constant *per capita*, not per patch). The steady state of the master equation can be computed as:

$$p_k^K = p_0^K \prod_{i=0}^{k-1} \left(\frac{\nu_i^K}{\mu_{i+1}^K} \right) \tag{4}$$

Two consistency relationships enter the picture. First, the probabilities that a patch with carrying capacity K contains a certain number of individuals of type 1 should sum to 1, *i.e.*:

$$\forall K, \sum_{i=0}^{K} p_i^K = 1 \tag{5}$$

Second, the average abundance of type 1, $\bar{i}$, is computed as:

$$\bar{i} = \sum_K \sum_{i=0}^{K} i p_i^K \pi_K \tag{6}$$

with π_K being the probability that a patch has a carrying capacity of K.

Let $\mathbf{P}$ be the vector of p_k^K sorted as $p_0^1, p_1^1, p_0^2, p_1^2, ...$ We can re-write the master equation in a more compact way:

$$\frac{d\mathbf{P}}{dt} = \mathbf{G}(\bar{i}, \overline{K}).\mathbf{P} \tag{7}$$

where $\mathbf{G}(\bar{i}, \overline{K})$ is the matrix incorporating the effects of natality and mortality. This matrix is block-diagonal because the probability of having k individuals from species 1 in a patch with carrying capacity K only depends on the probabilities of having $k-1$ or $k+1$ individuals in a similar patch, not in patches with different carrying capacities. This allows us to do all calculations with a particular value of K and then to sum all of its "parts" weighted by the π_K 's. Essentially, this property can be formalized as:

$$\frac{d\mathbf{P}_K}{dt} = \mathbf{G}_K(\bar{i}, \overline{K}).\mathbf{P}_K \tag{8}$$

where $\mathbf{P}_K$ is the vector $(p_0^K, p_1^K, ..., p_K^K)$ of population state in a patch with carrying capacity K and $\mathbf{G}_K$ is the corresponding matrix block in $\mathbf{G}$.

Invasibility criterion and singular strategies

A necessary and sufficient criterion for the invasion of type 1 (the mutant) in a metacommunity entirely occupied by type 2 (the resident) is given by (Chesson 1984; Metz and Gyllenberg 2001):

$$R_m(d_1, d_2) > 1 \tag{9}$$

where $R_m(d_1, d_2)$ is defined as the expected number of dispersers produced by a type 1 colony between its foundation and its eventual demise in the absence of type 1 immigrants (Chesson 1984; Metz and Gyllenberg 2001). This expectation is the scalar product of two vectors, $\mathbf{A}$ and $\mathbf{Z}$, defined respectively as the vector of disperser

production and the vector of quasi-equilibrium state probabilities when type 1 is rare in the metapopulation (Metz and Gyllenberg 2001).

As the dynamics of patches with different carrying capacities can be decoupled (equation [8]), the computation of the R_m criterion can be similarly simplified by first calculating the "partial" R_m^K quantity corresponding to patches with carrying capacity K and then summing these partial R_m^K weighted by the corresponding probability of finding a patch with carrying capacity K, π_K:

$$R_m = \sum_K R_m^K \pi_K = \sum_K (\mathbf{A}_K . \mathbf{Z}_K)\pi_K \tag{10}$$

where the vectors $\mathbf{A}_K$ and $\mathbf{Z}_K$ are defined as the disperser production vector and quasi-equilibrium state probability vector for patches with carrying capacity K. Both vectors have K dimensions (one for each possible number of type 1 individuals present in a patch containing at least one individual of type 1). The value a_i^K of the ith component of vector $\mathbf{A}_K$ is equal to id_1 (*i.e.* a patch that contains i type 1 individuals produce dispersers at a rate proportional to i). The vector $\mathbf{Z}_K$ is given by (Metz and Gyllenberg 2001):

$$\mathbf{Z}_K = -\widetilde{\mathbf{G}}_K^{-1} . \mathbf{Y}_K \tag{11}$$

where $\mathbf{Y}_K = [1, 0, 0, ...0]$ is the vector of initial state probabilities for a patch with carrying capacity K just colonized by a type 1 immigrant, and matrix $\widetilde{\mathbf{G}}_K$ is equal to $\mathbf{G}_K(0, \overline{K})$ without its first row and first column. The expression for the kth component of $\mathbf{Z}_K$, z_k^K, is obtained from the value of z_1^K (following the analogue to equation [4] at quasi-equilibrium):

$$z_k^K = z_1^K \prod_{i=1}^{k-1} \left(\frac{\nu_i^K}{\mu_{i+1}^K} \right) \tag{12}$$

However, when type 1 is rare, $\bar{i} = 0$, and thus:

$$\nu_i^K = \frac{K-i}{K} \frac{(1-d_1)i}{(1-d_1)i + (1-d_2)(K-i-1) + (1-c)d_2\overline{K}} \tag{13}$$

$$\mu_i^K = \frac{i}{K} \frac{(1-d_2)(K-i) + (1-c)d_2\overline{K}}{(1-d_1)(i-1) + (1-d_2)(K-i) + (1-c)d_2\overline{K}} \tag{14}$$

so that:

$$z_k^K = z_1^K \prod_{i=1}^{k-1} \left(\frac{K-i}{i+1} \right) \left(\frac{(1-d_1)i}{(1-d_2)(K-i-1) + (1-c)d_2\overline{K}} \right) \tag{15}$$

which yields, after some algebra and using the Γ function (extension of the factorial function):

$$z_k^K = z_1^K \frac{(K-1)!}{k(K-k)!} \left(\frac{1-d_1}{1-d_2} \right)^{k-1} \frac{\Gamma\left(K - k + \frac{(1-c)d_2\overline{K}}{1-d_2} \right)}{\Gamma\left(K - 1 + \frac{(1-c)d_2\overline{K}}{1-d_2} \right)} \tag{16}$$

The expression for z_1^K is obtained simply from the fact that (i) a patch with carrying capacity K offers empty microsites with rate mK, (ii) a type 1 disperser wins the competition for an empty microsite with probability $\frac{1-c}{(1-d_2)(K-1)+(1-c)d_2\overline{K}}$, and (iii) a unique type 1 individual dies with rate m (*i.e.* its life span is $1/m$):

$$z_1^K = \frac{(1-c)K}{(1-d_2)(K-1)+(1-c)d_2\overline{K}} \tag{17}$$

The computation of R_m follows:

$$R_m(d_1,d_2) = \sum_K \frac{(1-c)d_1K_2F_1\left[1,1-K;2-K-\frac{(1-c)d_2\overline{K}}{1-d_2};\frac{1-d_1}{1-d_2}\right]\pi_K}{(1-d_2)(K-1)+(1-c)d_2\overline{K}} \tag{18}$$

where ${}_2F_1[a,b;c;z]$ is the hypergeometric series, defined as:

$${}_2F_1[a,b;c;z] = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)\Gamma(c)z^k}{\Gamma(a)\Gamma(b)\Gamma(c+k)k!}$$

Singular strategies of dispersal are found by nullifying the selection gradient, *i.e.* d is a singular strategy when (Dieckmann and Law 1996; Geritz et al. 1998):

$$\left[\frac{\partial R_m}{\partial d_1}\right]_{d_1=d_2=d} = 0 \tag{19}$$

The selection gradient is computed from equation (18):

$$\begin{aligned}\left[\frac{\partial R_m}{\partial d_1}\right]_{d_1=d_2=d} &= \sum_K \frac{(1-dK+d(1-c)\overline{K})K}{d\overline{K}(1-d+d(1-c)\overline{K})}\pi_K \\ &= \frac{\overline{K}(1+d(1-c)\overline{K})-d\overline{K^2}}{d\overline{K}(1-d+d(1-c)\overline{K})}\end{aligned} \tag{20}$$

which yields the unique singular strategy, d^* given by:

$$d^* = \mathrm{Min}\left[\frac{1}{(c+\gamma_2)\overline{K}},1\right] \tag{21}$$

where $\gamma_2 = \frac{\overline{K^2}-\overline{K}^2}{\overline{K}^2}$ is the squared coefficient of variation of carrying capacities. The constraint $d \le 1$ implies the final form of equation (21). However, the selection gradient does not vanish at $d = d^* = 1$ because this is not a "true" singular strategy (*i.e.* selection on d drives it towards larger and larger values, but the selective pressure does not vanish once d reaches its maximum).

NB: When $\overline{K} \to \infty$, $d^*\overline{K}$ still equals $1/(c+\gamma_2)$, *i.e.* the denominator in equation (20) has a finite limit near d^*, and thus the problem of finding d^* admits no true singularity (the numerator in equation [20] equals 0 when $d = d^*$, whatever the value of $\overline{K}$).

How is this result connected with classical theory on dispersal evolution?

In now classical studies on dispersal evolution (Frank 1986; Gandon and Michalakis 2001; Ajar 2003; Jansen and Vitalis 2007), dispersal is accounted for as an altruistic behavior, *i.e.* higher dispersal is favored by a higher relatedness among inhabitants of the same patch. Classical results are obtained through the concept of "inclusive fitness" (Ajar 2003), which is a measure of fitness that accounts for the direct and indirect effects of a strategy/gene on its bearer's fitness, and is proportional to the selection gradient. Evolutionary biologists attuned to inclusive fitness formulae usually expect the expression of the inclusive fitness (or selection gradient) to be of the following kind:

$$W_{IF} \propto b_0 + b_1 F - c_0 \tag{22}$$

where W_{IF} is the inclusive fitness, b_0 measures direct benefits, b_1 measures indirect benefits linked to the relatedness coefficient, F, among individuals within a patch, and c_0 measures direct costs. Let's recast the selection gradient in a patch with carrying capacity K using the quantity $I = d(1-c)\overline{K}/(1-d)$:

$$\begin{aligned}\left[\frac{\partial R_m^K}{\partial d_1}\right]_{d_1=d_2=d} &= \frac{(1-dK+d(1-c)\overline{K})K}{d\overline{K}(1-d+d(1-c)\overline{K})} \\ &= \frac{K}{d\overline{K}}\left(1-\frac{(K-1)d/(1-d)}{1+I}\right) \\ &= \left(\frac{K}{\overline{K}}\right)\left(\frac{1}{d}\right)\left[1-\left(\frac{1}{1-c}\right)\left(\frac{K-1}{\overline{K}}\right)\left(\frac{I}{1+I}\right)\right]\end{aligned} \tag{23}$$

In this equation, we can recognize the relatedness coefficient, F_K, equal to the probability that two random individuals in the same K-patch share the same ancestor in this patch. Indeed, the probability that two random individuals are the same is $1/K$. When this is not the case, we can trace back the single ancestor of each individual at each successive death-replacement event. Whenever one of the two ancestors is born, one of three things can happen: (i) its direct parent is the ancestor of the other individual (probability equal to $x_K = (1-d)/[(1-d)(K-1)+d(1-c)\overline{K}]$); (ii) its direct parent is an immigrant (probability equal to $y_K = d(1-c)\overline{K}/[(1-d)(K-1)+d(1-c)\overline{K}]$); (iii) its direct parent is neither an immigrant, nor the ancestor of the other individual (probability $z_K = (1-d)(K-2)/[(1-d)(K-1)+d(1-c)\overline{K}]$). The probability that the two individuals have a common ancestry in the patch is thus given by:

$$\begin{aligned}F_K &= \frac{1}{K}+\left(\frac{K-1}{K}\right)\left(x_K\sum_{p=0}^{\infty} z_K^p\right) \\ &= \frac{1}{K}+\left(\frac{K-1}{K}\right)\left(\frac{1}{1+I}\right)\end{aligned} \tag{24}$$

Injected in equation (23), this yields:

$$\left[\frac{\partial R_m^K}{\partial d_1}\right]_{d_1=d_2=d} = \left(\frac{K}{d(1-c)\overline{K}}\right)\left[\frac{K}{\overline{K}}F_K-\left(\frac{K}{\overline{K}}-1\right)-c\right] \tag{25}$$

This equation stresses the existence of indirect benefits of not competing with related individuals, proportional to $KF_K / \overline{K}$, and costs associated with dispersal, proportional to $\left[\left(K / \overline{K}\right)-1\right]+c$, *i.e.* the sum of the direct dispersal cost, proportional to c, and the cost associated to environmental heterogeneity, proportional to $\left(K / \overline{K}\right)-1$. This last term represents a true cost for patches larger than average, and a benefit for smaller patches. In Ajar's (2003) formulation, equation (25) can be rewritten as:

$$\left[\frac{\partial R_m^K}{\partial d_1}\right]_{d_1=d_2=d} = \left(\frac{K}{d\overline{K}}\right) W_{IF}(K) \tag{26}$$

where $W_{IF}(K) = \frac{1}{1-c}\left[\frac{K}{\overline{K}} F_K - \left(\frac{K}{\overline{K}} - 1\right) - c\right]$ is the per capita inclusive fitness component for individuals in K-patches. If we note $\mathrm{E}_{exp}\left[X\right] = \sum X_K \left(\frac{K}{\overline{K}}\right)\pi_K$ the average of variable X computed over all individuals from all patch types, $W_{IF}(K)$ can be rewritten as:

$$W_{IF}(K) = \frac{\mathrm{E}_{exp}\left[K\right] - K}{\mathrm{E}_{exp}\left[K-1\right]} \tag{27}$$

Comparison of equation (21) with the dispersal ESS found in other models

Hamilton & May (1977), Comins *et al.* (1980), Frank (1986), and Taylor & Frank (1996) already tackled the issue of finding the dispersal ESS in finite populations in the absence of carrying capacity heterogeneity. In the absence of perturbations, this ESS is $d^* = \left[1 + 2cK - \sqrt{1 + 4c^2 K(K-1)}\right] / 2cK(1+c)$. This formula is clearly different from equation (21) in the absence of carrying capacity heterogeneity ($d^* = 1/cK$), the reason being that classical models have been studied in discrete time whereas ours is in continuous time. This small difference in initial model assumptions creates a small discrepancy in the ESS level for dispersal because of differences in relatedness among patch mates:

(i) the relatedness in non-overlapping generation models with self-replacement is (Taylor and Frank 1996; Ajar 2003):

$$f_K = \frac{\left[1 - d + d(1-c)\right]^2}{\left[1 - d + d(1-c)K\right]^2 - d^2(1-c)^2 K(K-1)} \tag{28}$$

(ii) the relatedness obtained in our model (*i.e.* with overlapping generations) is:

$$F_K = \frac{1 - d + d(1-c)}{1 - d + d(1-c)K} \tag{29}$$

A little algebra can prove that $F_K > f_K$, *i.e.* individuals living in the same patch are more related to each other in continuous time than in discrete time. The reason for this is that the parent individual does not cease to live at its offspring's birth, and hence taking two random individuals in a patch may result in taking a parent-offspring pair in the continuous-time model (but not in the discrete-time model). This effect can also be

modeled in a discrete-time context, provided adult survival is accounted for (Irwin and Taylor 2000). Provided that dispersal is an altruistic behavior, higher relatedness implies a higher ESS dispersal probability, which explains qualitatively the apparent discrepancy between equation (21) and earlier findings.

Convergence stability

Convergence stability of the singular strategy is obtained when (Geritz *et al.* 1998):

$$\frac{d}{dd^*}\left(\left[\frac{\partial R_m}{\partial d_1}\right]_{d_1=d_2=d^*}\right)<0 \tag{30}$$

This condition can be computed from equation (20):

$$(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]d^{*2}-2\overline{K}\left[(1-c)\overline{K}-1\right]d^*-\overline{K}<0 \tag{31}$$

As long as $\overline{K}>1/(1-c)$, the product of the d-roots of $(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]d^2-2\overline{K}\left[(1-c)\overline{K}-1\right]d-\overline{K}$ are negative, so that convergence stability of the singular strategy d^* is obtained when:

$$d^*<\frac{\overline{K}\left[(1-c)\overline{K}-1\right]+\sqrt{\overline{K}^2\left[(1-c)\overline{K}-1\right]^2+(c+\gamma_2)\overline{K}^3\left[(1-c)\overline{K}-1\right]}}{(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]} \tag{32}$$

i.e. after some manipulations:

$$0<\frac{\sqrt{\overline{K}^2\left[(1-c)\overline{K}-1\right]^2+(c+\gamma_2)\overline{K}^3\left[(1-c)\overline{K}-1\right]}}{(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]} \tag{33}$$

which is true when $\overline{K}>1/(1-c)$. When $\overline{K}<1/(1-c)$, the discriminant of $(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]d^2-2\overline{K}\left[(1-c)\overline{K}-1\right]d-\overline{K}$ is always negative since:

$$\overline{K}^2\left[(1-c)\overline{K}-1\right]^2+(c+\gamma_2)\overline{K}^3\left[(1-c)\overline{K}-1\right]=\overline{K}\left[(1-c)\overline{K}-1\right](\overline{K^2}-\overline{K})<0 \tag{34}$$

Thus, the roots of $(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]d^2-2\overline{K}\left[(1-c)\overline{K}-1\right]d-\overline{K}$ are complex and $(c+\gamma_2)\overline{K}^2\left[(1-c)\overline{K}-1\right]d^2-2\overline{K}\left[(1-c)\overline{K}-1\right]d-\overline{K}<0$ is true. The convergence stability of d^* need not be checked when $(c+\gamma_2)\overline{K}<1$ (*i.e.* when d^* is forced to be 1) given that the selection gradient insures convergence stability.

Evolutionary stability

Evolutionary stability of the singular strategy is obtained when (Geritz *et al.* 1998):

$$\left[\frac{\partial^2 R_m}{\partial d_1^2}\right]_{d_1=d_2=d}<0 \tag{35}$$

Following equation (10), this is equivalent to:

$$\sum_K \pi_K\left[\frac{\partial^2 R_m^K}{\partial d_1^2}\right]_{d_1=d_2=d}<0 \tag{36}$$

Computation of $\left[\frac{\partial^2 R_m^K}{\partial d_1^2}\right]_{d_1=d_2=d}$ is tedious, but results in:

$$\left[\frac{\partial^2 R_m^K}{\partial d_1^2}\right]_d = \frac{2K(K-1)\left[d(K-(1-c)\overline{K})-2\right]}{d\overline{K}\left[1-d+d(1-c)\overline{K}\right]\left[2(1-d)+d(1-c)\overline{K}\right]} \tag{37}$$

The condition of evolutionary stability is then:

$$\sum_K \pi_K \frac{2K(K-1)\left[d(K-(1-c)\overline{K})-2\right]}{d\overline{K}\left[1-d+d(1-c)\overline{K}\right]\left[2(1-d)+d(1-c)\overline{K}\right]} < 0 \tag{38}$$

Using the standardized skewness $\gamma_3 = \frac{\overline{K^3}-3\overline{K^2}\,\overline{K}+2\overline{K}^3}{\left(\overline{K^2}-\overline{K}^2\right)^{3/2}}$, and squared coefficient of variation $\gamma_2 = \frac{\overline{K^2}-\overline{K}^2}{\overline{K}^2}$, this is equivalent to:

$$\gamma_3 < 2\gamma_2^{1/2} + \frac{(c\overline{K}-1)}{\overline{K}}\gamma_2^{-1/2} + \frac{c(\overline{K}-1)}{\overline{K}}\gamma_2^{-3/2} \tag{39}$$

when $(c+\gamma_2)\overline{K} > 1$ (cf. equation [21]). When $(c+\gamma_2)\overline{K} < 1$, evolutionary stability is always achieved because $d^* = 1$ is not an analytical extremum of the fitness function, but rather due to trait constraint.

If we now look for conditions of disruptive selection, we can combine the opposite inequality of equation (39) with the condition $(c+\gamma_2)\overline{K} > 1$. Disruptive selection then occurs when

$$\frac{1}{\overline{K}} - \gamma_2 < c < \frac{\gamma_3 - 2\gamma_2^{1/2} + \gamma_2^{-1/2}/\overline{K}}{\gamma_2^{-1/2} + \frac{\overline{K}-1}{\overline{K}}\gamma_2^{-3/2}} \tag{40}$$

This means that a moderate cost of dispersal is required for disruptive selection. When $\overline{K}$ is sufficiently large, this equation practically means that c must obey an upper bound inequality:

$$c < \frac{\left(\gamma_3 - 2\gamma_2^{1/2}\right)\gamma_2^{3/2}}{1+\gamma_2} \tag{41}$$

From this last inequality, we deduce a sufficient criterion for disruptive selection at high $\overline{K}$:

$$c_{max} = \frac{\left(\gamma_3 - 2\gamma_2^{1/2}\right)\gamma_2^{3/2}}{1+\gamma_2} > 1 \tag{42}$$

Criterion (42) is very practical since we can compute it without knowing the particular density of microsites per unit area. Indeed, if we only have information on the distribution of patch areas, we can use criterion (40) to check whether disruptive

selection on dispersal can occur due to patch size asymmetry, since γ_2 and γ_3 are dimensionless.

A lower bound on skewness and some implications

Consider a discrete positive stochastic variable X, with mean value $\overline{X}$. The application of Cauchy-Schwarz's inequality states that:

$$\overline{X}^2 \leq \overline{X^3}.\overline{X} \tag{43}$$

with equality when variable X takes only one value. Taking $X = K$ and remembering that $\gamma_2 = \frac{\overline{K^2}-\overline{K}^2}{\overline{K}^2}$ and $\gamma_3 = \frac{\overline{K^3}-3\overline{K^2}\,\overline{K}+2\overline{K}^3}{\left(\overline{K^2}-\overline{K}^2\right)^{3/2}}$, we obtain a lower bound on γ_3 given by:

$$\gamma_2^{1/2} - \gamma_2^{-1/2} \leq \gamma_3 \tag{44}$$

The difference D_3 between the upper and lower bounds for γ_3 leading to the existence of an ESS singular strategy when $\gamma_2 > (1-c\overline{K})/\overline{K}$ (equations [39] and [44]) is given by:

$$D_3 = \gamma_2^{1/2} + \frac{((1+c)\overline{K}-1)}{\overline{K}}\gamma_2^{-1/2} + \frac{c(\overline{K}-1)}{\overline{K}}\gamma_2^{-3/2} \tag{45}$$

When $\overline{K} > 1/c$, the condition $\gamma_2 > (1-c\overline{K})/\overline{K}$ is always true. In that case, the asymptotic behavior of D_3 when γ_2 is near 0 depends on the value of $\overline{K}-1$ (equation [45]), which is positive, *i.e.* D_3 becomes infinitely positive (and thus there is an ESS at that limit). When $1 < \overline{K} < 1/c$, the condition $\gamma_2 < (1-c\overline{K})/\overline{K}$ is verified for very low values of γ_2. At this point, the singular value of d has a non-vanishing selection gradient, and thus there is an ESS at that limit. When γ_2 tends towards infinity, D_3 is of order $\gamma_2^{1/2}$, *i.e.* of the same order as the lower bound on γ_3 imposed by Cauchy-Schwarz's inequality (equation [44]). A distribution with a large γ_2 and $\gamma_3 > 2\gamma_2^{1/2}$ always implies the existence of an evolutionary branching point.

Interpretation of inequality (39) in terms of inclusive fitness

After some algebraic manipulations, inequality (39) can be rewritten as:

$$\left[\gamma_3\gamma_2^{3/2} + \gamma_2\left(1-\gamma_2\right)\right]\overline{K}^2 < \left(c+\gamma_2\right)\left(1+\gamma_2-\frac{1}{\overline{K}}\right)\overline{K}^2 \tag{46}$$

The left-hand side term can be identified as:

$$\left[\gamma_3\gamma_2^{3/2} + \gamma_2\left(1-\gamma_2\right)\right]\overline{K}^2 = \mathrm{Var}_{exp}\left[K\right] \tag{47}$$

whereas the right-hand side terms are given by:

$$\left(c+\gamma_2\right)\left(1+\gamma_2-\frac{1}{\overline{K}}\right)\overline{K}^2 = \frac{\mathrm{E}_{exp}\left[K-1\right]}{d^*} \tag{48}$$

Injecting equations (47) and (48) into inequality (46) yields the following equivalent to inequality (39):

$$d^* \frac{\mathrm{Var}_{exp}\left[K\right]}{\mathrm{E}_{exp}\left[K-1\right]} < 1 \tag{49}$$

We can go one step further by noting that (from equation [27]):

$$\mathrm{Var}_{exp}\left[W_{IF}\left(K\right)\right] = \frac{\mathrm{Var}_{exp}\left[K\right]}{\mathrm{E}_{exp}\left[K-1\right]^2} \tag{50}$$

We now have the final form of inequality (39)'s equivalent in terms of inclusive fitness:

$$d^* \mathrm{E}_{exp}\left[K-1\right] \mathrm{Var}_{exp}\left[W_{IF}\left(K\right)\right] < 1 \tag{51}$$

or, in other words, when selection is not directional (*i.e.* $(c+\gamma_2)\overline{K} > 1$), selection is stabilizing if, and only if, the grand average number of intra-patch neighbors ($\mathrm{E}_{exp}\left[K-1\right]$) times the dispersal rate at the singular strategy, times the variance in individual inclusive fitness, does not exceed unity.

References

Ajar, E. 2003. Analysis of disruptive selection in subdivided populations. BMC Evol. Biol. 3

Chesson, P. L. 1984. Persistence of a markovian population in a patchy environment. Z Wahrscheinlichkeit. 66:97-107.

Comins, H. N., W. D. Hamilton, and R. M. May. 1980. Evolutionarily stable dispersal strategies. J. Theor. Biol. 82:205-230.

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

Frank, S. A. 1986. Dispersal polymorphisms in subdivided populations. J. Theor. Biol. 122:303-309.

Gandon, S., and Y. Michalakis. 2001. Multiple causes for the evolution of dispersal. Pp. 155-167 *in* J. Clobert, E. Danchin, A. A. Dhondt and J. D. Nichols, eds. Dispersal. Oxford University Press, New York.

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

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

Irwin, A. J., and P. D. Taylor. 2000. Evolution of dispersal in a stepping-stone population with overlapping generations. Theor. Popul. Biol. 58:321-328.

Jansen, V. A. A., and R. Vitalis. 2007. The evolution of dispersal in a Levins' type metapopulation model. Evolution 61:2386-2397.

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

Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. J. Theor. Biol. 180:27-37.

Appendix S2: Application of the model to geometric distributions

We consider a geometric distribution of carrying capacities defined by parameter p ($0 < p < 1$):

$$\pi_K = p(1-p)^{K-1} \tag{52}$$

The moments of this distribution can be computed:

$$\overline{K} = 1/p \tag{53}$$

$$\gamma_2 = 1-p \tag{54}$$

$$\gamma_3 = \frac{2-p}{\sqrt{1-p}} \tag{55}$$

These moments allow the computation of the ESS criterion (equation [39]) when $\gamma_2 > (1-c\overline{K})/\overline{K}$, *i.e.* $2p < 1+c$:

$$\frac{2-p}{\sqrt{1-p}} < 2\sqrt{1-p} + \frac{c-p}{\sqrt{1-p}} + \frac{c}{\sqrt{1-p}} \tag{56}$$

i.e.

$$p < c \tag{57}$$

If $p < c$, then there is an ESS. When $c < p < (1+c)/2$, there is disruptive selection on dispersal. For $p > (1+c)/2$, the singular dispersal value becomes 1 and is an ESS. For most values of parameter p, disruptive selection occurs rather when c is low than when it is high ($c = 0$ maximizes the width of the p-window for disruptive selection). Conversely, $p = 1/2$ maximizes the width of the c-window that allows for disruptive selection to take place.

Appendix S3: Accounting for demographic stochasticity

Our model is explicitly based on the assumption that each death event is immediately followed by a replacement event. Thus, local populations are immune to demographic stochasticity since their population sizes are always equal to their carrying capacities. Here, we describe a more general formulation of the model, which encompasses the model described in the main text and in Appendix S1 and a whole class of recruitment-limited lottery models.

Let us begin with the description of a monomorphic metapopulation with dispersal rate *d*. We note *m* and *b* the *per capita* mortality and birth rates, respectively, and $p_{k,K}$ is the probability that a patch (with carrying capacity *K*) contains exactly *k* individuals. The master equation describing the dynamics of $p_{k,K}$ is given by:

$$\begin{aligned}\frac{dp_{k,K}}{dt} &= b\left(1-\frac{k-1}{K}\right)\left[(1-d)(k-1)+d(1-c)\bar{k}\right]p_{k-1,K}+m(k+1)\,p_{k+1,K}\\ &\quad -\left\{b\left(1-\frac{k}{K}\right)\left[(1-d)k+d(1-c)\bar{k}\right]+mk\right\}p_{k,K}\end{aligned} \tag{58}$$

where $\bar{k}$ is the average number of individuals per patch over the whole metapopulation.

Equation (58) is based on the following rules:

1. offspring are produced at a rate *b*;
2. offspring disperse with probability *d*;
3. dispersing offspring disperse to any patch (global dispersal);
4. dispersing offspring die with probability *c*;
5. any offspring trying to settle in a patch with carrying capacity *K* and current population *k* has a probability $1-\frac{k}{K}$ of succeeding;
6. any unsuccessful offspring dies;
7. settled individuals die at rate *m*.

Note that "a process *X* happens at rate *x*" means that the process *X* is a random Poisson process with rate *x*, *i.e.* that the probability of *X* happening during an infinitesimal time interval *dt* is *x.dt*.

When $b/m \to \infty$, the process described by equation (58) is equivalent to the model described in the main text and in Appendix S1. When b/m is finite, demographic stochasticity can play a role in the dynamics of the metapopulation. Contrary to the model presented in the main text, the adaptive dynamics of *d* associated with equation (58) cannot be written in a compact way, mostly because the resident phenotype's abundance distribution has to be found together with the value of $\bar{k}$ and also because the transition matrix *G* becomes much bigger than in the previous model.

To assess the robustness of our conclusions to the existence a small dose of demographic stochasticity, we ran simulations in a landscape of patches following a truncated geometric distribution and setting $b/m = 100$ (see main text). As explained in

the Results section, the main differences between simulations with and without demographic stochasticity are seen under disruptive selection regimes and concern drifter types. Essentially, demographic stochasticity disfavours drifters because drifters are less efficient than dwellers in terms of birth rate (drifters lose a significant fraction of their offspring due to dispersal cost). In the presence of local demographic stochasticity, this features can cause total extinction of drifter lineages (see e.g. on Fig. S3a). This in turn drives a transient directional selection regime on dwellers and afterwards triggers another evolutionary branching and the reformation of an equivalent drifter type. This effect does not change the distribution of dispersal types when observed over a long time scale (Fig S3b).

Appendix S4: Computing the plastic ESS

In this appendix, we assume that dispersal is plastic, *i.e.* that dispersal cannot be described through a single parameter, d, but through a function, $d(K)$, giving the optimal dispersal rate for offspring born in a patch with carrying capacity K. The formula given in Discussion for the ESS of plastic dispersal strategy results from the following reasoning.

For an individual living in a *K*-patch, the selection gradient on $d(K)$ is (equations [20] and [23] in Appendix S1):

$$\left[\frac{\partial R_m^K}{\partial d_1(K)}\right]_{d_1(K)=d_2(K)=d(K)} = \frac{(1-d(K)K+(1-c)\overline{d(K)K})K}{\overline{d(K)K}(1-d(K)+(1-c)\overline{d(K)K})} \tag{59}$$

The rationale behind this equation is that the average immigrant pressure (dispersal cost notwithstanding) is changing from $d\overline{K}$ (in the non-plastic model) to $\overline{d(K)K}$ (in the condition-dependent dispersal model).

Solving equation (59) at the singular strategy (*i.e.* left-hand side equal to 0), we obtain

$$d(K)K = 1+(1-c)\overline{d(K)K} \tag{60}$$

and thus, averaging over patch types:

$$\overline{d(K)K} = 1/c \tag{61}$$

Plugging equation (61) into equation (60), we finally get:

$$d(K) = \frac{1}{cK} \tag{62}$$

in accordance with the formula given in Discussion.

It is reassuring to note that this equation does not go against some simple intuitions on plastic dispersal:

(i) individuals in large patches should disperse less than individuals in small patches (because intra-patch relatedness is lower in larger patches);

(ii) having some clues as to the distribution of patch sizes removes the selective pressure against dispersal due to risk aversion (*i.e.* the term proportional to γ_2 in main text equation [3]).

Interestingly, there are more migrants circulating among patches in the model that includes plasticity than in the model developed in the paper (compare equation [61] vs. $d^*\overline{K} = 1/(c+\gamma_2)$ when inequality [4] is satisfied).

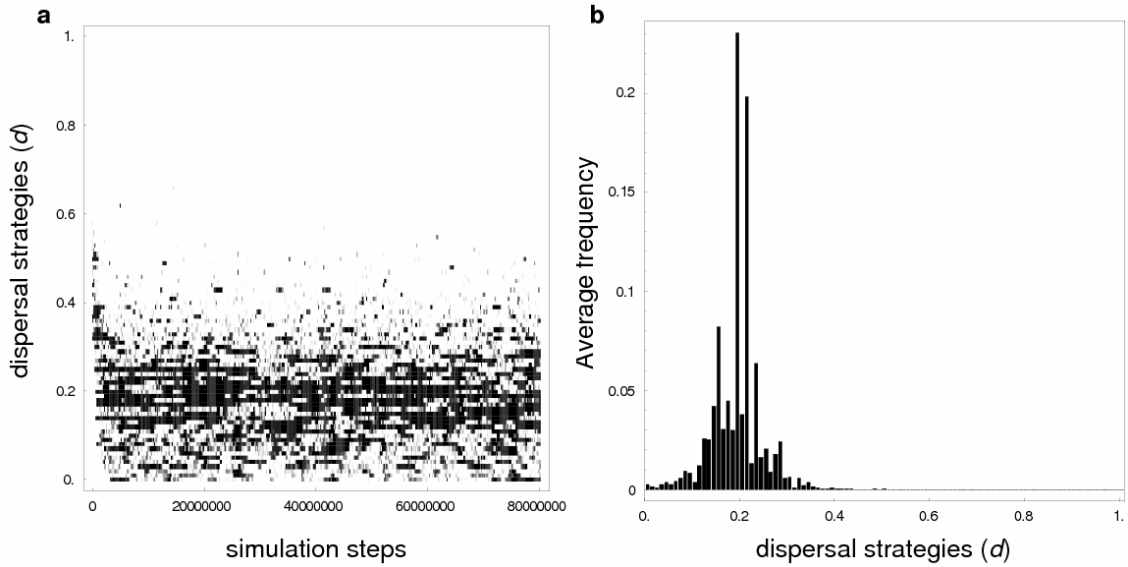

Fig. S1. Simulation results for $p = 0.2$ (example of stabilizing selection on dispersal). (*a*) Distribution of dispersal levels (*d*, ordinates) at different times during the simulation (steps, abscissas). Darker shades indicate a higher frequency of the corresponding dispersal level. (*b*) Frequency distribution of dispersal levels (*d*) when averaged over the last 200 records (between the 40,000,001st and 80,000,000th steps).

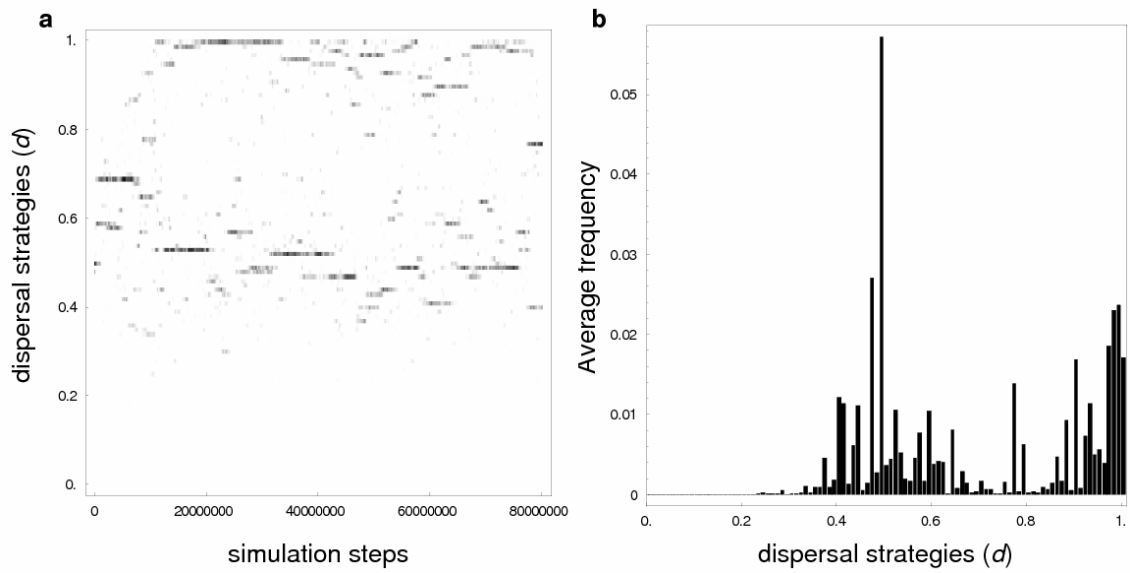

Fig. S2. Simulation results for $p = 0.5$ (example of disruptive selection on dispersal). (*a*) Distribution of dispersal levels (*d*, ordinates) at different times during the simulation (steps, abscissas). Darker shades indicate a higher frequency of the corresponding dispersal level. (*b*) Frequency distribution of dispersal levels (*d*) when averaged over the last 200 records (between the 40,000,001st and 80,000,000th steps).

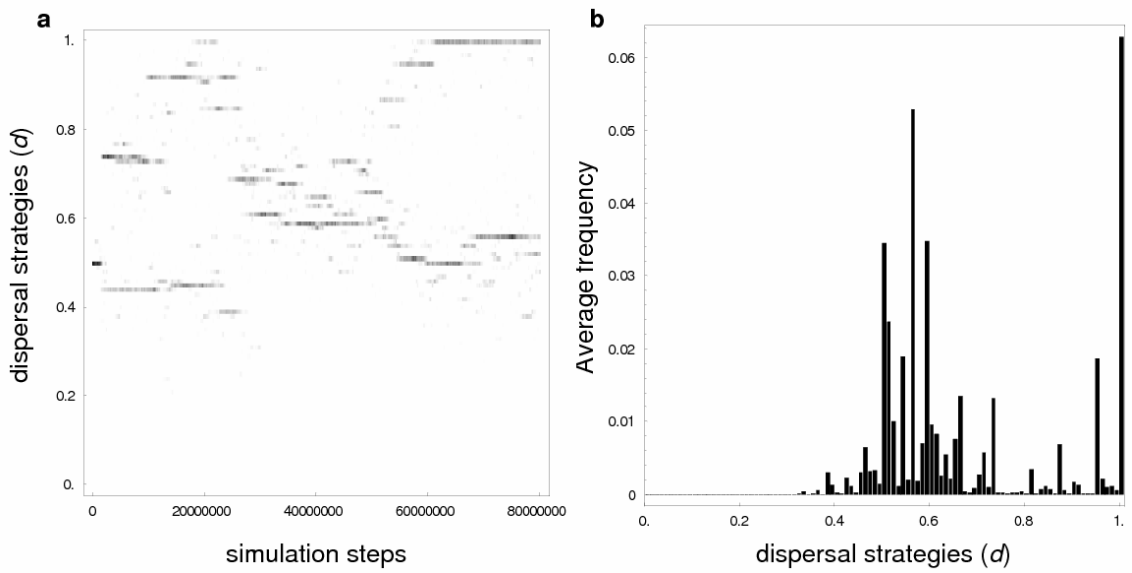

Fig. S3. Simulation results for $p = 0.5$ with demographic stochasticity. (*a*) Distribution of dispersal levels (*d*, ordinates) at different times during the simulation (steps, abscissas). Darker shades indicate a higher frequency of the corresponding dispersal level. (*b*) Frequency distribution of dispersal levels (*d*) when averaged over the last 200 records (between the 40,000,001st and 80,000,000th steps).